UNIVERSITI TEKNOLOGI MARA

MDEC™ Premier Digital Tech Institution

PEERS'23

PEERS'23

SERIES IV

THE FOURTH POSTGRADUATE RESEARCH SYMPOSIUM 2023

kppimuitmofficial kppim_uitm_official KppimUiTM

PEERS'23

SERIES IV

THE FOURTH POSTGRADUATE RESEARCH SYMPOSIUM 2023

DISCLAIMER & COPYRIGHT

PEERS'23 Series IV: The Fourth Postgraduate Research Symposium 2023

e ISBN 978-629-97817-1-4

9 786299 781714

Published by:

College of Computing, Informatics and Mathematics
Al-Khawarizmi Building
Universiti Teknologi MARA
40450 Shah Alam Selangor
Darul Ehsan, Malaysia

EDITORIAL

Editors
Ts. Dr. Norizan Mat Diah
Prof. Dr. Nursuriati Jamil
Assoc. Prof. Ts Dr. Nur Atiqah Sia Abdullah
Assoc. Prof. Dr. Shafaf Ibrahim
Dr. Tengku Zatul Hidayah Tengku Petra

PEERS'23 SERIES IV COMMITTEE

Patron Advisor 1
Prof. Ts. Dr. Haryani Haron

Patron Advisor 2
Prof. Ts. Dr. Wan Abdul Rahim Wan Mohd Isa

Advisor
Assoc. Prof. Dr. Norjansalika Janom
Assoc. Prof. Ts. Dr. Nur Atiqah Sia Abdullah

General Chair
Dr. Zainura Hj Idrus

Co-Chair
Assoc. Prof. Dr. Nasiroh Omar
Dr. Afiza Ismail

Secretary
Nur Aina Khadijah
Nur Azima Alya Narawi

Treasurer
Norasiah Mohammad
Nur Aina Khadijah Adnan

Operation
Dr. Tajul Rosli Razak
Dr. Norzila Musa
Ahmad Taufiq Hj Mohamad

Program
Hafizatul Hanin Hamzah
Hana Fakhirah Almarzuki
Dr. Shakirah Hashim

Publication
Prof. Dr. Nursuriati Jamil
Assoc. Prof. Ts. Dr. Nur Atiqah Sia Abdullah
Ts. Dr. Norizan Mat Diah
Assoc. Prof. Dr. Shafaf Ibrahim
Dr. Tengku Zatul Hidayah Tengku Petra
Nurul Syeilla Syazhween Zulkefli
Norharziana Yahaya Rashdi

Registration, Food & Beverages
Dr. Suzana Ahmad
Dr. Noor Latiffah Adam

Logistics
Muhamad Ridhwan Mohamad Razali
Hafizatul Hanin Hamzah

Protocol
Prof. Dr. Nursuriati Jamil
Assoc. Prof. Dr. Marina Ismail

Exhibition
Dr. Zainura Hj Idrus
Dr. Siti Arpah Ahmad
Dr. Sharifah Aliman
Assoc. Prof. Dr. Fariza Hanis Abdul Razak
Assoc. Prof. Dr. Nor Shahniza Kamal Bashah

Speakers & Panels Industry
Prof. Madya Zaidah Ibrahim
Syamsulhairi Yaakop
Dr. Syed Mohd Zahid Syed Zainal Ariffin

Jury
Assoc. Prof. Dr. Nurazzah Abd Rahman
Assoc. Prof. Dr. Sofianita Mutalib
Dr. Shakirah Hashim
Dr. Siti Khatijah Nor Abdul Rahim

Student Volunteers
Muhammad Idzham Mazlan
Danial Thaqif Mohamed Fikri
Nur Fatin Izati Tajuddin
Nurmi Adilah Amiludin
Atiqah Basirah Miswan
Nur Fatihah Md Nazri
Muhammad Aleef Firdaus Mohamad Fathil Zaki
Sarah Batrisya Ismandi
Muhammad Azri Sulhi Suhaidi
Ahmad Haikal Ahmad Zulfikri
Hazira Alisa Hardi
Muhammad Hafiz Naufal Hasmuni
Muhammad Nafeez Suazam
Miqdad Zaidi
Nor Tasnim Amani Che Ramli
Nurul Izzah Iddarus
Siti Jessmin Izz-Loi
Muhammad Amir Sufi Zulkifli
Nur Aisyah Zuhairah Zulkifli
Nurfathin Dina Ajah Maiden

PREFACE

In an era of relentless technological advancement, it is imperative for universities and educational institutions to equip graduates with the skills and knowledge needed to not only survive but thrive in the ever-evolving industry landscape. The Ministry of Higher Education (MOHE) in Malaysia shares this vision, dedicating itself to enhancing the quality of higher education and acknowledging its pivotal role in shaping the nation's future.

Within this context of academic excellence and industry readiness, the "Fourth Postgraduate Research Symposium 2023 (PEERS'23 Series IV)" stands as a prominent initiative. This symposium, one of several aimed at bolstering the visibility of the College of Computing, Informatics, and Mathematics, serves as a vital bridge connecting academia and industry. It opens doors for collaboration, knowledge exchange, and mutual growth.

PEERS'23 offers students an exceptional platform to enhance their networking skills by engaging with industry professionals who attend the event. This interaction provides invaluable insights that will shape their future careers. Importantly, the exchange is not one-sided; industry stakeholders attending PEERS'23 can also glean fresh perspectives and innovative ideas from the students.

This publication is a testament to the dedication and hard work of the students who participated in PEERS'23. It encompasses 27 extended abstracts, each representing a research project undertaken by a master's student who presented their work at the symposium. These abstracts have undergone rigorous review to meet the high academic standards set by PEERS'23.

We extend an invitation to researchers, students, and readers to explore the wealth of knowledge presented within these pages. The abstracts cover a wide array of topics, spanning from computer science and data analytics to information technology and human-computer interaction, showcasing the diverse range of research pursuits within the College of Computing, Informatics, and Mathematics. Furthermore, we encourage our readers to connect with the authors of these abstracts. Collaborative opportunities often arise from the exchange of ideas and perspectives.

In essence, this book is a testament to the power of education, the potential of students, and the importance of collaboration. We hope that you, our readers, will find inspiration, insights, and opportunities within these pages as we collectively strive to make a positive impact on the ever-changing landscape of higher education and industry.

Welcome to the world of PEERS'23, where knowledge knows no bounds, and the future beckons with open arms.

Dr. Zainura Hj Idrus
General Chair PEERS' 23 Series IV

TABLE OF CONTENTS

COMPARATIVE ANALYSIS OF MACHINE LEARNING TECHNIQUES ON ORGANIZATIONAL CITIZENSHIP BEHAVIOUR: UTILIZING UNIVERSITY STUDENT DATA

Abdu Ibrahim Adamu[1], Zainura Idrus[2*], Jahaya Abdul Hamid[3]

[1,2]College of Computing, Informatics and Mathematics, Universiti Technologi MARA, Shah Alam, Selangor, Malaysia

[3]Kolej Matrikulasi Kedah, Kementerian Pendidikan Malaysia, Changlun, Kedah, Malaysia

[1]abduibrahimadamu@yahoo.com, [2]zainura501@uitm.edu.my, [3]jahaya@kmk.edu.my

ABSTRACT

Student organisational citizenship behaviour (OCB) is referring to a voluntary behaviour displayed by students within university. OCB is important because it promotes a positive study environment by enhancing teamwork, cooperation, going beyond formal study requirement and overall productivity, ultimately leading to the success in their study. One of the methods used to analyse OCB is machine learning techniques, which involve using algorithms to find patterns and make predictions. However, there is lacking research in comparing these techniques. Thus, this research aims to evaluate the accuracy of various machine learning algorithms in clustering students' behaviour in exhibiting OCB. By comparing different machine learning techniques, this research seeks to identify the most reliable and efficient method for clustering students' levels of OCB. This research has utilized the Machine Learning Life Cycle as its operational framework, and K-means, Hierarchical, and DBSCAN are the chosen algorithms. The results show that the score of K-means is better than the other two techniques in terms of Silhouette, Calinski-Harabasz Index, and Davis-Bouldin. This concludes that K-means is the suitable model for creating OCB models for student data, providing significant insights.

Keywords*— Student Organisational Citizenship Behaviour (OCB), Clustering Technique, Clustering Model, Machine Learning, Performance Metrics*

1. INTRODUCTION

In recent years, machine learning techniques have gained significant attention in various fields, including social sciences and human resources management. These techniques have proven to be effective in analyzing and predicting complex behavior patterns. One such area where machine learning can be applied is in the study of Organization Citizenship Behavior (OCB) among university students.

OCB refers to discretionary behavior that contributes to the overall functioning and effectiveness of an organization. It encompasses voluntary actions that go beyond the formal job requirements and include behaviors such as helping colleagues, volunteering for additional tasks, and showing concern for the organization's well-being (Khiong et al., 2022; Ndoja & Malekar, 2020). Similarly, student OCB refers to the positive, voluntary, and discretionary actions that students take to support the institution's well-being and the learning environment. Understanding and promoting OCB among university students is crucial as they represent the future workforce and organizational leaders (Dutta & Sahney, 2022; Krimbill et al., 2019).

The concept of OCB (Organizational Citizenship Behavior) encompasses four important dimensions: altruism, civic virtue, conscientiousness, and courtesy. Altruism is demonstrated when students selflessly help other students with academic tasks, such as completing assignments or studying for exams. Civic virtue entails students voluntarily participating in and contributing to various school activities. Conscientiousness is reflected in students' punctuality, timely submission of assignments, and active engagement in classroom discussions and activities. Lastly, courtesy involves the respectful gesture of informing instructors when unable to attend a class.

Machine learning techniques provide a novel approach to analyze and predict OCB based on extensive data collected from university students. These techniques offer the potential to identify

the key factors that influence OCB and develop models that can accurately cluster students who are engaging in OCB. This knowledge can aid universities and organizations in designing strategies to encourage and foster OCB among students and improve overall institutional performance.

The aim of this paper is to compare and evaluate different machine learning techniques in clustering OCB among university students. By examining three algorithms, which are K-means, Hierarchical clustering and DBSCAN, this research seeks to determine the most effective approach for clustering OCB data. The findings of this study have the potential to improve organizational policies and practices, leading to more efficient and productive study environments.

The structure of this paper is as follows: Section 2 discusses the methodology employed in this study, including data collection, feature selection, and evaluation metrics. Section 3 presents the results and findings of the different machine learning techniques. Finally, Section 4 concludes the paper by summarizing the key findings.

2. METHODOLOGY

This research utilized the Machine Learning Life Cycle as its operational framework, with modifications made to suit the specific project requirements (Hewage & Meedeniya, 2022). The cycle starting with problem formulation and research objectives. Then followed by the second cycle which is data gathering. Data are gathered from a sample of students from one university through online questionnaire. The dataset encompasses various features, including demographics, extracurricular activities, and social interactions, among others. Each survey response denotes the extent to which students engage in citizenship behaviours within the university setting.

The third cycle focuses on preparing the data for feature extraction. Principal component analysis was employed to reduce the dataset's dimensionality to enhance the accuracy of the analysis. The fourth cycle involves designing, developing, and validating the models. Three models were developed using three clustering algorithms: K-means, Hierarchical Clustering and DBSCAN. These models were then compared based on their performance. Three performance metrics used are Silhouette, Calinski-Harabasz Index, and Davis-Bouldin. Finally, the outcomes were presented using data visualization techniques.

3. RESULTS AND DISCUSSION

This section presents the outcomes of the research study where the performance of different clustering algorithms, namely K-Means, Hierarchical Clustering and DBSCAN, were evaluated. In addition, since this research have applied Principal Component Analysis (PCA), the model performance will be evaluated using validation metrics. The validation metrics used are the Silhouette Coefficient, Calinski-Harabasz Index and Davies-Bouldin Index. These metrics provide insights into the quality of cluster formations and separation within the datasets. Table 1 presents the results of the validation metrics for the different models used in the study after applying PCA.

3.1 K-mean Clustering

For K-mean clustering model, the Silhouette Coefficient is measured at 0.34366529, indicating a moderate clustering quality. The Calinski-Harabasz Index is measured at 395.72254557, indicating that the cluster formations are well-separated and dense. The Davies-Bouldin Index is measured at 0.88234530, which suggests good clustering quality with minimal overlap between clusters.

3.2 Hierarchical Clustering

In the case of the Hierarchical Clustering model, the Silhouette Coefficient is measured at 0.28421078. This indicates a slightly lower clustering quality compared to K-mean clustering. The Calinski-Harabasz Index is measured at 293.53016838, indicating that the cluster formations are reasonably well-separated, but not as dense as in K-mean clustering. The Davies-Bouldin

Index is measured at 1.12510223, which suggests a moderate clustering quality with a slightly higher level of overlap between clusters compared to K-mean clustering.

3.3 DBSCAN

The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) model shows relatively lower performance in terms of the Silhouette Coefficient, which is measured at 0.12844345. This indicates a lower clustering quality with clusters that may not be well-separated. The Calinski-Harabasz Index is measured at 36.24213806, suggesting that the cluster formations are less dense and less well-separated compared to the previous models. The Davies-Bouldin Index is measured at 9.88737707, indicating a poor clustering quality with significant overlap between clusters.

Overall, the evaluation of validation metrics after applying PCA reveals that the K-mean clustering model performs the best, followed by hierarchical clustering. DBSCAN shows relatively weaker clustering quality. These findings provide insights into the performance of the models and can guide the selection of an appropriate clustering algorithm for future analysis.

Table 1. Validation metrics for the models after applying PCA.

	Silhouette Coefficient	**Calinski-Harabasz Index**	**Davies-Bouldin Index**
K-mean Clustering	0.34366529	395.72254557	0.88234530
Hierarchical Clustering	0.28421078	293.53016838	1.12510223
DBSCAN	0.12844345	36.24213806	9.88737707

4. CONCLUSION

In this research, three clustering algorithms have been analysed by using students' OCB data from one of the universities in Malaysia. Three clustering models have been developed with K-mean, Hierarchical and DBSCAN algorithm. In terms of performance, Silhouette, Calinski-Harabasz Index, and Davis-Bouldin have been used as the performance metrics. Based on the analysis, it can be concluded that the K-means clustering algorithm performed the best among the three evaluated methods, followed by hierarchical clustering, while DBSCAN showed relatively weaker clustering quality. These findings provide valuable insights into the performance of these models and can guide the selection of an appropriate clustering algorithm for future analysis.

REFERENCES

1. Dutta, V., & Sahney, S. (2022). Relation of principal instructional leadership, school climate, teacher job performance and student achievement. *Journal of Educational Administration*, 60(2), 148–166. https://doi.org/10.1108/JEA-01-2021-0010
2. Hewage, N., & Meedeniya, D. (2022). *Machine Learning Operations: A Survey on MLOps Tool Support*. https://doi.org/10.48550/arXiv.2202.10169
3. Khiong, K., Tinggi, S., Agama Buddha, I., & Boyolali, S. (2022). Teachers Performance and Satisfaction of Medical School: How the Role of Knowledge Sharing, and Organizational Citizenship Behaviour (OCB)? *Journal of Industrial Engineering & Management Research,* 3(6), 138-147. https://doi.org/10.7777/jiemar.v3i6.407
4. Krimbill, E. M., Goess, D. E., & Escobedo, P. V. (2019). Organizational Citizenship and Teacher Evaluation: Using the T-TESS to Promote OCB and Improve Student Outcomes, *School Leadership Review*, 14 (1), 12-25. https://scholarworks.sfasu.edu/slr/vol14/iss1/2
5. Ndoja, K., & Malekar, S. (2020). Organisational citizenship behaviour: a review. *International Journal of Work Organisation and Emotion*, 11(2), 89-104. https://doi.org/10.1504/IJWOE.2020.110629

KEYWORD SELECTION FROM DOCUMENTS USING WEB SCRAPING

Aina Shahidah Mohd Shafri[1] and Hayati Abd Rahman[2*]

[1,2] College of Computing, Informatics and Mathematics, Universiti Technologi MARA, Shah Alam, Selangor, Malaysia

[1]shahidahmdshafri@gmail.com, [2]hayatiar@tmsk.uitm.edu.my

ABSTRACT

Web Scraping is the process of collecting data by extracting data from a website that consists of two techniques: manual scraping and automated scraping. Price monitoring, market research, news monitoring, email marketing, and sentiment analysis are a few examples of web scraping that have been used. This paper will discuss the process of web scraping used to filter selected keyword mentions from a selected URL domain from a Malay online newspaper website. The process starts by identifying and analyzing the URL domain, the category of the news being published, the number of published news in each category, and selecting the keyword mentioned from the URL category. A total of 70 documents were used in this experiment varying from 11 news categories. The analysis showed 55 out of 70 documents being scraped correctly, and 15 of them were not when the keyword mentions appeared when the scraper showed the presence of the words. Meanwhile, the other 15 documents are identified as being scraped incorrectly after crosschecking manually with the list of the documents being collected at the beginning of the experiment. The data obtained by the process mentioned will be also visualized in this research.

Keywords*: web scraping, news scraping, filter keyword*

1. INTRODUCTION

The internet is becoming a source of information and information for everyone regardless of rank whether they are students, teachers, or workers in other sectors, the internet is very important at this time. Everyone can get any information on the internet with just one click, and because there is too much information and data obtained from the internet, it makes it difficult for humans to get specific information about certain things.

By using web scraping, the desired information can be filtered according to the desired category. There are two types of web scraping: manual and automated (Chapman, 2020). Manual scraping is when the copy and paste process is being done, while automated scraping is when using any module or library that matches the language usage in the website to be scrapped.

Web scraping is widely used for different industries such as retail and manufacturing, equity and financial research, data science, risk management, and product, marketing, and sales purposes (Patel, 2022). A few needs of web scraping can be used for price monitoring, market research, news monitoring, sentiment analysis, and email marketing. The effectiveness of web scraping can be seen by its performance in automatic data collection (Significance and effectiveness of web scraping services, 2019). Most of this aggregation is running crawlers and parsers (Suciu, 2022). An instance of the application of news scraping that has been developed previously is used for analyzing how leaders' statements during the period of COVID-19 have impacted the number of COVID-19 cases daily (Dewi et al., 2019). The need for news monitoring by news scraping will be emphasized in this research.

This research is focused on how keyword selection from online newspaper documents can be done using web scraping. Utusan Melayu, which is one of the online newspapers is decided to be the domain of this research as it is a legitimate source and reputable news company to prevent inaccurate information. The number of news being published in the categories is taken into high consideration as it will also be affected in this research.

2. METHODS

This research begins by selecting the URL Domain by finding a stable online newspaper website to prevent inaccurate information from being used in this research hence, the Utusan Melayu online website is being chosen.

Next, these selected URL Domains are validated whether it is valid or not. This is important as it becomes evident that the ULR Domain being selected is from a reliable URL. The system will display the status code = 200 if the URL is valid, and displaying not valid if the status of the URL is not 200.

The next process is the main page of the URL domain will undergo the crawling process which will find all the links in the HTML codes on the page. A few Python libraries being leveraged are BeautifulSoup, Selenium, Requests, and Scrapy as there are multiple programming languages that can be found on the webpages. In this process, some of the links found cannot be identified in what categories it is being published. Therefore, the web scraper decided to identify the category of the news being published in the URL domain to make it easier in the next keyword searching process.

After identifying the categories involved, the news published in each category is extracted and saved in the *.csv file. The number of news being published for each of the categories with the title is recorded in the file.

The user is required to insert the keyword selection to be scrapped through the news published in the selected category in the next process by inserting the URL of the category and the selected keyword to be searched within the document.

3. RESULTS AND DISCUSSION

All the categories of the news being published are recognized as 11 categories which are “Nasional”, “Terkini”, “Premium”, “Seri”, “Pancaindera”, “Rencana”, “Sukan”, “Gaya”, “Luar Negara”, and “Ringgit”, and all the news published in each of the categories mentioned is recorded in the *.csv file. The total number of news published in each of the categories is also being recognized in this research by collecting the data counts from the saved *.csv file. As a result of all the steps mentioned in the methods, this research will display the keyword in the selected news category after inserting the keyword (Figure 1) as in Figure 2.

Figure 1. Example of Keyword Mention Selected

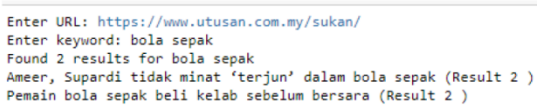

Figure 2. The Results Found After the Keyword Mention

This research also will be producing visualization such as line graphs, bar graphs, and pie charts, to summarize all the findings such as the number of news being published for each of the categories as in Figure 3, Figure 4, and Figure 5.

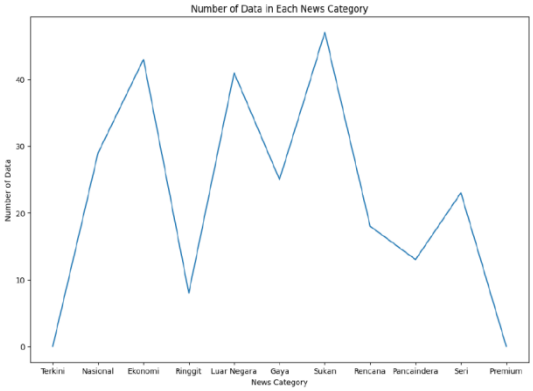

Figure 3. The Line Graph of the News Published in Each Category

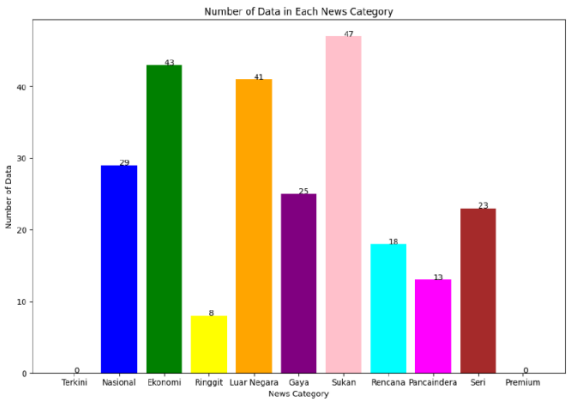

Figure 4. The Bar Graph of the News Published in Each Category

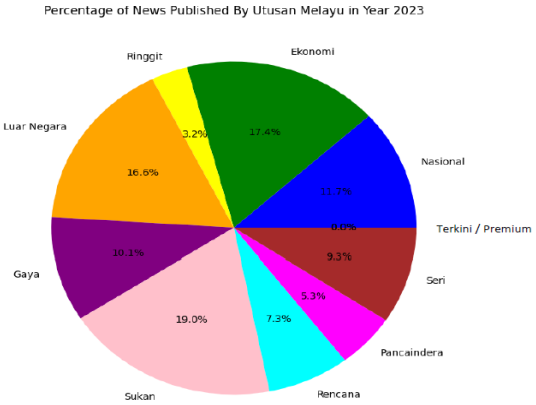

Figure 5. The Percentage of News Being Published in Each Category

4. CONCLUSIONS

This research adjudicates the news scraping of Utusan Melayu online newspaper. Several applications and fields of web scraping are being overviewed during the research. The parameter used in scraping websites is also being identified by obtaining a thorough understanding of the scraping models, and comparative analysis is being done between the similar system techniques implemented in the previous study. These parameters are the total number of files being scraped, the categories of the news, and the keyword being scraped by the webpage. This web scraper is being developed through a preliminary study including data collection, algorithm design, system design, and the system requirement. After several processes are completed, the result of this research is obtained after the system displays the outcome of the keyword being mentioned in the selected URL which has been recorded as the news category being published. A few modules and libraries have also been analyzed and identified to suit the developed system and the URL domain in this research.

REFERENCES

1. Chapman, R. (2020, February 11). Top 10 web scraping techniques: Limeproxiess. Limeproxies. From https://limeproxies.netlify.app/blog/top-10-web-scraping-techniques
2. Dewi, L. C., Meiliana, & Chandra, A. (2019). Social Media Web Scraping using Social Media Developers API and Regex. *Procedia Computer Science, 157*, 444–449. doi:10.1016/j.procs.2019.08.237
3. Patel, H. (2022, October 21). How Web Scraping is Transforming the World with its Applications. Medium. https://towardsdatascience.com/https-medium-com-hiren787-patel-web-scraping-applications-a6f370d316f4
4. Significance and Effectiveness of Web Scraping Services. (2019, May 26). BotScraper.com. https://www.botscraper.com/blog/significance-and-effectiveness-of-web-scraping-services
5. Suciu, D. (2022, January 7). When is Web Scraping the Answer? The Top 7 Uses Cases for Data Extraction | by Dan Suciu | Medium. Medium. https://dan-suciu.medium.com/when-is-web-scraping-the-answer-the-top-7-uses-cases-for-data-extraction-d95cb72d12ac

ENHANCEMENT OF RECOMMENDATION APPROACHES IN ONLINE ENTERTAINMENT

Nur Anis Afiqah Mazlan[1] and Nur Atiqah Sia Abdullah[2*]

[1, 2] School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]anisafiqa99@gmail.com, [2]atiqah684@uitm.edu.my

ABSTRACT

The rapid proliferation of technology and the internet has significantly transformed the landscape of online entertainment platforms. Notably, YouTube and Netflix have established itself as a premier on-demand broadcast service with an extensive catalogue of movies and TV shows. However, this technological advancement has inadvertently led to a pressing problem: information overload, wherein individuals are inundated with an overwhelming volume of data, making content discovery and recommendations increasingly challenging. This paper addresses the problem of information overload in the realm of online entertainment by proposing a novel approach that leverages Tags and Genres to enhance recommendation systems. The aim of this study is to alleviate the burden of information overload for users by providing more accurate and personalized content recommendations. We employ a methodology that involves data preprocessing and the incorporation of additional genres into the recommendation dataset. Data preprocessing techniques are applied to refine and optimize the dataset, ensuring that it is well-suited for recommendation algorithms. Furthermore, we augment the dataset with ten new genres, enriching the content categorization and improving the diversity of recommendations. The results of our study demonstrate that the proposed approach yields significant improvements in the quality of content recommendations within the online entertainment platforms. By effectively addressing the issue of information overload through the integration of Tags and Genres, users can benefit from more relevant and engaging content suggestions. These findings underscore the effectiveness of our method in enhancing the online entertainment experience, ultimately contributing to more satisfying and personalized user journeys. In conclusion, this research successfully tackles the challenge of information overload in online entertainment by introducing a novel recommendation approach centred around Tags and Genres.

Keywords: *online entertainment, tags and genres, information overload, recommender.*

1. INTRODUCTION

Nowadays, video online entertainment websites such as YouTube, Netflix, Spotify, etc. caught everyone's eyes in society (Duong-Trung et al., 2020). Digital streaming services and online viewing platforms are providing crucial respite from the repetitiveness and mundanity of day to day of self-isolation as the COVID-19 virus relentlessly spreads around the world since the end of 2019. Online entertainment refers to the use of applications and services on the internet to enable online users to engage in an entertaining activity. Recently digital streaming services such as YouTube, Netflix, Amazon Prime, HBO, Spotify, Pandora, and iTunes have witnessed large-scale adoption (Lamkhede & Das, 2019).

The recommendation systems have been widely used to provide recommendations to users based on user preferences. The recommendation system recommends items such as books, images, news, music, movies, etc. The information or items which interest the users will be produced using a recommendation system (Ferreira et al., 2020).

There are many types of recommendation systems such as content-based (Kvifte et al., 2022) (Fernández, 2018), collaborative filtering, knowledge-based (Fayyaz et al., 2020), and hybrid recommendations system. The most common recommendation systems approaches are

collaborative filtering, content-based, and hybrid-based recommendation systems (Fayyaz et al., 2020).

The entertainment platforms such as Netflix and YouTube are recommendation systems based on previous user activities and the contents of the visited sites to provide related content to reduce the searching time and improve the availability of data. This paper aims to review the recommender approaches, which solve the problem of data overload and constant updates. The recommendation systems aim to reduce the overload of information and provide recommendations that could assist the users in the decision recommendation systems (Ferreira et al., 2020). Online users suffer due to the information overload problem and to overcome the problem the recommender systems are widely used as a decision support tool to address this situation. The term "information overload" refers to obtaining too much information for a person to comprehend. Online entertainment websites such as YouTube and Netflix are hampered by inadequate information which consequences in less accurate recommendations by just relying on a content-based recommendations approach (Sharma & Li, 2021).

This paper suggests using content-based filtering and adding tags and genres to enrich the dataset. The main overview of this research is to enhance movie recommendations in online entertainment to solve information overload.

2. METHODS

For this research, the study design is the five steps of design science research. The steps are awareness of problems, suggestions, development, evaluation, and conclusion. One of the problems in this research is the information overload on entertainment websites. In this methodology, this research adapted Tags and Genres to overcome information overloading issues.

The research suggests taking into consideration the tags and genres in the dataset. The suggestions are to enhance the online entertainment websites in the movie dataset which added tags and genres such as comedy, crime, romantic, action, musical, fantasy, history, family, cartoon, and documentary.

In this research, the movie dataset was collected from the Kaggle website. It is a website to look at any reliable data or sources. The movie dataset consists of show id, type of show, title, director, cast, country, date added, release year, rating, duration, category, and description. There are two datasets involved in Google Colab. The first dataset is the original dataset collected from the Kaggle websites while the second dataset is the enhanced dataset after data preprocessing. By going through data preprocessing, the genre and tags were added to this dataset manually based on the description of the movie review. If the review is matched with the genre, then it is labeled with 1. Or else it is labeled as 0. With the label 0 and 1, all the movies are tagged accordingly.

After adding the tags and genres to the original dataset, we run an experiment to test the recommendation result. Based on the results, it concludes the enhanced dataset gives a better result in recommending the movies. Figure 1 shows the process of the enhancement in the dataset.

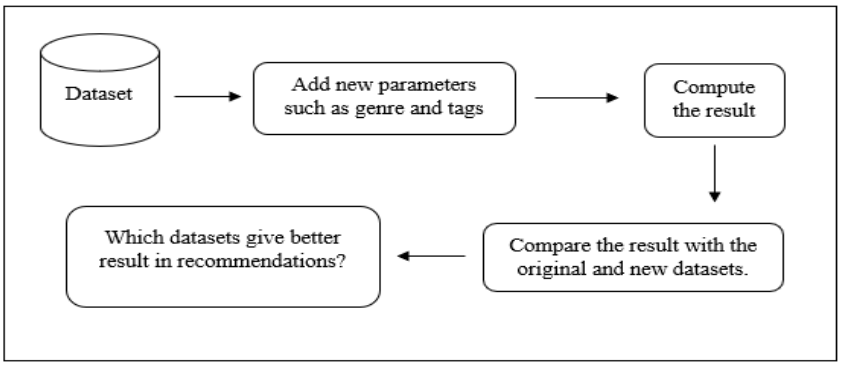

Figure 1. Illustration of the Enhancement Process

3. RESULTS AND DISCUSSION

The original dataset, which is the movie dataset consisting of 200 data was collected from the Kaggle website for this project. The movie dataset consists of show id, type of show, title, director, cast, country, date added, release year, rating, duration, category, and description. The

original datasets have many types of genres such as animation, family, comedy, musical, docuseries, action-adventure, nature, science fiction, and many more. However, these 2 or more genres are combined into one column. While running with the original dataset in Google Colab, multiple classifications of genres in one column could not give accurate results as it has shown no output when searching for 2 or more genres.

There are 10 genres added to the original dataset. The 10 added genres are funny, crime, romantic, action, musical, fantasy, history, family, cartoon, and documentary. After going through data preprocessing, the dataset is reclassified into 10 genres with binary codes 0 and 1. The binary code 1 stands for the presence of the genres in the movie list while code 0 stands for no related genres in the movie list.

This research compares the results using two datasets, which are the original dataset and reclassified dataset. The original dataset contains multiple genres in one column while the new dataset was added with genres in each column. The 10 genres in each category are classified in their columns.

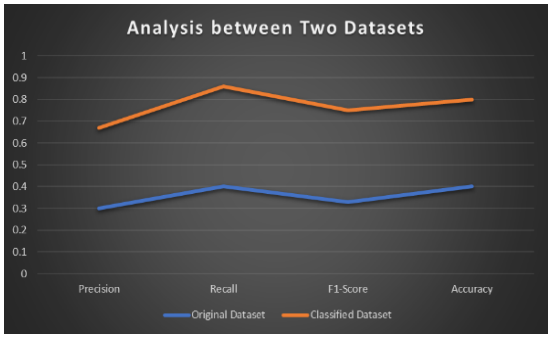

Figure 2. The Precision, Recall, F1-Score, and Accuracy

	Original Dataset	Classified Dataset
Precision	0.30	0.67
Recall	0.40	0.86
F1-Score	0.33	0.75
Accuracy	0.40	0.80

Figure 3. The Analysis of Two Datasets

Figure 3 shows the analysis between two datasets of the original dataset and the classified dataset. The original dataset was going through data preprocessing by adding genres to the new dataset and known as a classified dataset. The genres were reclassified into 10 genres which are funny, crime, romantic, action, musical, fantasy, history, family, cartoon, and documentary. The accuracy scores of two datasets were compared to determine which dataset performs better in recommending movies based on genre preferences. If one of the datasets shows higher accuracy, it indicates that the content-based filtering performs better in generating movie recommendations based on genre preferences.

As a result, the classified dataset showed better accuracy with 0.80 than the original dataset with a 0.40 value. The classified dataset shows higher accuracy than the original dataset.

4. CONCLUSIONS

This research is focusing on enhancing the recommendation in online entertainment. It enhances the movie recommendations by adding tag and genre. From the review, there are ten genres categorized from the movie dataset. By tagging binary code of 0/1, the dataset is added with different genres. As the result of experiment, it showed that the enhanced dataset with addition tag and genre, has successfully increased the accuracy in the recommendation of movie. In conclusion, this research successfully tackles the challenge of information overload in online entertainment by introducing a novel recommendation approach centered around Tags and Genres. The methodology employed and the results obtained affirm the feasibility and efficacy of this approach in enhancing content recommendations and ultimately enhancing the online entertainment experience.

REFERENCES

1. Duong-Trung, N., Nguyen, Q. N., Le Ha, D. N., Ha, X. S., Phan, T. T., & Huynh, H. X. (2020). Genres and actors/actresses as interpolated tags for improving movie

recommender systems. *International Journal of Advanced Computer Science and Applications*, 11(2), 67–74. https://doi.org/10.14569/ijacsa.2020.0110210
2. Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Applied Sciences (Switzerland)*, 10(21), 1–20. https://doi.org/10.3390/app10217748
3. Fernández, L. E. M. (2018). *Recommendation System for Netflix*. https://beta.vu.nl/nl/Images/werkstuk-fernandez_tcm235-874624.pdf
4. Ferreira, D., Silva, S., & Abelha, A. (2020). Recommendation System Using Autoencoders. *Applied Sciences.* 10(16), 5510. *https://doi.org/10.3390/app10165510*
5. Kvifte, T., Elahi, M., & Trattner, C. (2022). Hybrid Recommendation of Movies Based on Deep Content Features. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13236 LNCS*(November), 32–45. https://doi.org/10.1007/978-3-031-14135-5_3
6. Lamkhede, S., & Das, S. (2019). Challenges in search on streaming services: Netflix case study. *SIGIR 2019 - Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1371–1374. https://doi.org/10.1145/3331184.3331440
7. Sharma, R.S., Shaikh, A.A. & Li, E. (2021) Designing Recommendation or Suggestion Systems: looking to the future. *Electron Markets*, 31, 243–252. https://doi.org/10.1007/s12525-021-00478-z

GENERATING AND VALIDATING EVENT E-CERTIFICATE USING BLOCKCHAIN

Muhamad Nazrin Izat Abdul Rahman[1] and Prasanna Ramakrisnan[2*]

[1,2]School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]2022581933@student.uitm.edu.my, [2]prasanna@ prasanna@uitm.edu.my

ABSTRACT

Decentralized, distributed, secure, faster, more transparent, and unmodifiable are just some of the benefits of blockchain technology. These advantages surpass those of current technology. Academic certifications provided by educational institutions are crucial for students when applying for jobs, taking government examinations, attending college, and even obtaining a visa to study abroad. However, the process of issuing certificates is often unclear and not easily verifiable. Consequently, there are numerous opportunities for the fabrication of certifications. To address this challenge, blockchain technology, with its high data security and immutable storage in a distributed ledger, can be employed to combat document fraud and forgery. Through the utilization of blockchain technology, this study aims to enhance document verification. As a result, this study examines the benefits, risks, and associated obstacles in the implementation of application-supported blockchain technology with principles and norms for educational credential verification. In conclusion, this review study proposes the use of a blockchain technology system to generate and validate digital event certificates.

Keywords: *Blockchain, Non-Fungible Token, Smart Contract*

1. INTRODUCTION

In recent years, decentralized architecture has garnered substantial attention due to its versatile applications across various domains (Dwivedi et al., 2019). This decentralized framework also holds promise in addressing issues within the Internet of Things (IoT) sector, particularly in enhancing security (Babaei & Schiele, 2019). The inception of blockchain technology began with Bitcoin, a digital currency. Operating as a peer-to-peer network, it requires no personal information for participation, ensuring anonymity. Transactions within the blockchain are secured through consensus mechanisms and a transparent public ledger. Notably, the Proof-of-Work (PoW) algorithm, employed by prominent public blockchains like Bitcoin and Ethereum, plays a crucial role in achieving consensus (Lasla et al., 2020). Special nodes called "miners" verify transactions, and each transaction is authenticated using a pair of public and private keys shared among participants. The integrity of the public ledger, being immutable, guarantees the immutability of the transaction history (Hasin et al., 2021; Supriya Manglekar & Dinesha H.A., 2018). Any attempt to alter records is thwarted by other peer nodes, upholding transaction reliability.

Blockchain's research foundation centers on consensus algorithms that prioritize speed, security, and scalability. Although permissioned blockchain algorithms offer efficiency and security, their public scalability is limited. To address this limitation, lightweight public consensus mechanisms like the Directed Acyclic Graph (DAG) have been tailored for IoT platforms. Furthermore, innovative consensus models present opportunities to eliminate transaction fees within established cryptocurrency frameworks (Cai et al., 2018).

2. METHODS

The study focused on the comprehensive development of the Generating and Validating Event E-Certificate web system, with a particular emphasis on the decision to adopt a public

blockchain architecture. This choice emerged as a cornerstone of project success, driven by the need for transparency, decentralization, and security in managing event certificates. Public blockchains were deemed ideal, aligning with the goal of providing a transparent and tamper-proof certification process for all participants.

The decision-making process involved evaluating different blockchain architecture types: public, private, and consortium. Public blockchains stood out due to their transparency and decentralization attributes, which aligned perfectly with the aim of creating an open and trustworthy certificate issuance environment. The suitability of the public blockchain was further supported by its advantages in transparency, security, and immutability. This architecture ensured the lasting trustworthiness of certificates, making them resistant to alterations. Its distributed validation process bolstered security and enhanced system resilience.

The deliberate choice of a public blockchain architecture laid the foundation for subsequent development phases. It established the project's overarching objective: the creation of a robust Generating and Validating Event E-Certificate web system centered around blockchain technology. This system seamlessly serves the needs of participants and organizers, reflecting the core goals of the project.

3. RESULTS AND DISCUSSION

The Generating and Validating Event E-Certificate system underwent thorough functionality testing, which was meticulously conducted by 20 skilled testers. These testers diligently explored all aspects of the system's functionalities, ensuring comprehensive coverage of event registration, certificate generation, Metamask integration, and blockchain interactions. Their meticulous approach involved simulating various user scenarios to identify and report any discrepancies, aiming to discover potential bugs or issues that might hinder the system's performance.

The outcome of the functionality testing was overwhelmingly positive, as all 20 testers awarded the system a passing grade. This indicates that the system successfully met the specified requirements and functioned as intended, without encountering any critical defects or obstacles. The smooth event registration, reliable certificate generation, and secure Metamask integration contributed to an optimal user experience, assuring stakeholders that participants and event organizers will encounter a user-friendly and efficient platform.

Moreover, the testers' successful execution of blockchain interactions involving Next.js and Infura confirmed the accurate generation and transfer of blockchain-based certificates. This validation ensures the system's adherence to blockchain standards, delivering authentic and verifiable certificates to participants. With the completion of functionality testing, the Generating and Validating Event E-Certificate system has proven its robustness, security, and readiness for deployment, instilling confidence in its functionality and setting the stage for a reliable and dependable web platform.

Administrators can confidently manage event certificates, organize events, and control the certificate issuance process, as the positive testing results affirm the system's capability to effectively handle administrative tasks and streamline event management. The exhaustive functionality testing undertaken by the 20 testers and their unanimous passing grades serve as a testament to the system's efficiency, reliability, and adherence to project requirements. This successful completion of the testing phase provides a strong foundation for the Generating and Validating Event E-Certificate system, ensuring a seamless and secure experience for all users involved.

The second test conducted was usability testing. The purpose of this test was to obtain feedback from organizers and participants to evaluate the flow of the Generating and Validating E-Certificate System using Blockchain. This testing process involved distributing a Google Form questionnaire to all organizers and participants of the event.

Before the survey began, all organizers and participants were briefed. During the survey, participants provided feedback about the system, including suggestions, opinions, and

expectations. Finally, an analysis was conducted based on the feedback and input from the survey participants.

Table 1. User Acceptance Test of Generating and Validating Event E-Certificate System

	User Acceptance Test	Disagree	Neutral	Agree
1	I find it useful for this system be using as Event Management System and E-Certificate Generator System.	0	5	15
2	I will consider using this system because of the functionality of the system.	0	6	14
3	This system is effective solution in solving user problems when handling event because user can understand how to use the system easily.	0	10	10
4	The flow of this system is clear and understandable to be used as a platform managing event.	0	3	17
5	This system easy to be implement in any event.	0	6	14
6	Overall, I satisfied with the system.	0	2	18

4. CONCLUSIONS

The testing and results phase of the Generating and Validating Event E-Certificate web system has been a resounding success. Thorough functionality testing, conducted by skilled testers, showcased the system's robustness and efficiency. With meticulous bug hunting and validation of critical functionalities, the testers reported flawless performance, unanimously awarding the system a passing grade. The positive outcomes of the testing phase instilled confidence in the system's reliability, user-friendliness, and adherence to requirements. The successful completion of testing has laid a strong foundation for the system's deployment, assuring stakeholders of its readiness for real-world use. Consequently, the Generating and Validating Event E-Certificate system is poised to provide a seamless and secure experience for event organizers and participants alike, ensuring efficient event management and trustworthy certificate validation.

REFERENCES

1. Babaei, A., & Schiele, G. (2019). Physical Unclonable Functions in the Internet of Things: State of the Art and Open Challenges. *Sensors.* 19(14):3208. https://doi.org/10.3390/s19143208
2. Cai, W., Wang, Z., Ernst, J. B., Hong, Z., Feng, C., & Leung, V. C. M. (2018). Decentralized Applications: The Blockchain-Empowered Software System. *IEEE Access*, 6, 53019–53033. https://doi.org/10.1109/ACCESS.2018.2870644
3. Hasin, F., Munia, T. H., Zumu, N. N., & Taher, K. A. (2021). ADS-B Based Air Traffic Management System Using Ethereum Blockchain Technology. *2021 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD 2021- Proceedings*, (pp 346–350). https://doi.org/10.1109/ICICT4SD50815.2021.9396828
4. Lasla, N., Alsahan, L., Abdallah, M., & Younis, M. (2020). Green-PoW: An Energy-Efficient Blockchain Proof-of-Work Consensus Algorithm. *Computer Networks* 214(11):109118. http://arxiv.org/abs/2007.04086
5. Manglekar, S. & Dinesha H.A. (2018). Block Chain: An Innovative Research Area, In *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),* (pp 1-4). http://10.1109/ICCUBEA.2018.8697717

STUDENT LEARNING PERFORMANCE DASHBOARD USING VISUALIZATION

Muhammad Syahmie Hamka Zulkifle[1] and Prasanna A/P Ramakrisnan[2*]

[1,2]School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]2022965065@student.uitm.edu.my, [2]prasanna@uitm.edu.my

ABSTRACT
This research analyzes students' learning performance using data visualization techniques to comprehensively understand academic progress and identify critical factors influencing student achievement. By leveraging various visualization techniques, this study seeks to explore trends, patterns, and relationships within the student learning data to inform evidence-based decision-making and enhance educational practices. The research will collect and analyze student performance data from multiple subjects, semesters, and demographic factors to provide actionable insights for improving learning outcomes.

Keywords: *student learning performance, student performance, visualization techniques*

1. INTRODUCTION

In the realm of education, understanding and improving student learning performance is of paramount importance. Educational institutions constantly seek ways to enhance teaching strategies, allocate resources effectively, and provide targeted interventions to optimize student outcomes (Toasa et al., 2018). Data visualization techniques have emerged as powerful tools for analyzing and interpreting complex educational data, enabling educators to uncover patterns, trends, and intervention opportunities that can inform decision-making processes.

The field of education generates a vast amount of data related to student performance, including grades, attendance records, demographic information, and more. Traditionally, such data has been analyzed through traditional statistical methods, which may need to capture the nuances and complexities inherent in educational datasets effectively. However, with recent advancements in data visualization techniques, educators can gain deeper insights into student learning performance, identifying key factors influencing academic achievement and designing targeted interventions to address areas of improvement (Askinadze et al. 2019).

It's a good thing that recent developments in data visualization methods have given educators additional opportunities to gain deeper insights from educational data. Data visualization presents information through interactive dashboards, graphs, and charts so that users may more efficiently and thoroughly recognize patterns, trends, and relationships. Teachers can find essential facts that would otherwise be concealed in massive datasets using these visual representations.

The use of data visualization in teaching has various advantages. It enables teachers to pinpoint important variables affecting students' academic performance. By visualizing data, teachers can quickly find correlations between characteristics like student performance and elements like socioeconomic background, instructional techniques, or engagement levels. This information allows teachers to modify their instructional strategies and interventions to focus on areas that need development for specific students or groups (Ebert et al., 2016).

Research by Horvath et al. (2018) stated that the data visualization techniques allow educators to delve deeper into educational data, revealing valuable insights and informing evidence-based decision-making. Educators can better assess student learning performance, identify influencing factors, and create focused interventions by going beyond conventional statistical

methodologies. Data visualization's emphasis on visuals encourages better communication and teamwork, promoting a shared understanding of educational data and improving students' overall educational experiences.

2. METHODS

The procedures, activities, and tools used to move a project from beginning to completion comprise a project management framework. It covers all the core components required for project planning, administration, and governance. The framework includes every aspect of the project, from necessary resources and tools to specific processes and tasks. Project management is crucial for the successful progress of software projects (Ssegawa & Muzinda, 2018). Therefore, this framework is structured to help organize and execute the project steps methodically. Figure 1 provides an overview of the operational framework.

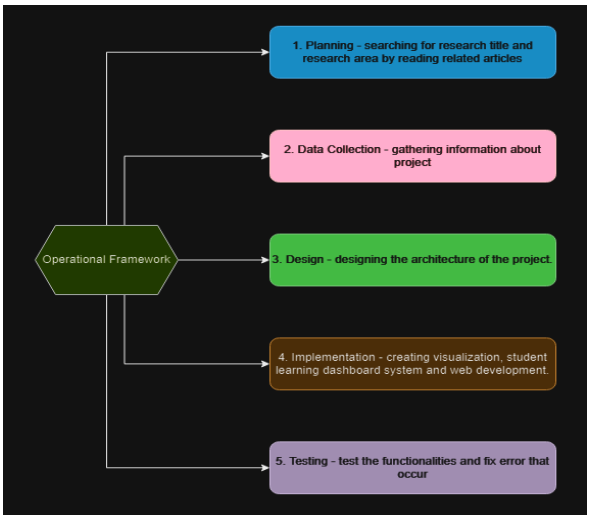

Figure 1. the overview of the operational framework

The Waterfall Model, one of the System Development Life Cycle (SDLC) models, was utilized to create the Student Learning Performance Dashboard Using Data Visualization, as stated by Yu et al., 2021. The waterfall model was highly popular during that period, as it encompassed all phases of the system development life cycle (SDLC). It earned its name because the model systematically progressed downward from one phase to another. Additionally, it was relatively easy to understand and implement. In the waterfall paradigm, each step had to be completed entirely before moving on to the next, emphasizing the sequential development of phases, as pointed out by Yu, (2018). This software development model was primarily suitable for small projects. At the end of each stage, a review was conducted to assess the progress of the thesis and determine whether it should proceed or be abandoned.

3. RESULTS AND DISCUSSION

The findings shed light on the effective utilization of these visualization techniques to present student performance data clearly and intuitively. Implementing line charts in the learning dashboard enables students to observe trends and patterns in their performance over time. This visualization technique allows for depicting longitudinal data, such as GPA or grades across different semesters. By plotting data points and connecting them with a line, students can quickly identify the trajectory of their academic progress, whether it shows an upward trend, remains stable, or exhibits fluctuations. This visualization approach facilitates a comprehensive understanding of students' performance trends, aiding in goal setting, self-assessment, and strategic decision-making.

Figure 2. Student Learning Performance Dashboard

Incorporating line and bar charts in a learning dashboard is a practical approach for visualizing student performance information. These visualization techniques give students a comprehensive overview of performance trends and enable easy comparison across subjects or categories. Using line and bar charts in the learning dashboard allows students to gain valuable insights, make informed decisions, and take proactive steps to enhance their academic achievements. Future research can explore additional visualization techniques or innovative approaches to improve the design of learning dashboards further and maximize their impact on student learning outcomes.

4. CONCLUSIONS

In conclusion, the findings of this research underscore the successful development of a web-based application dashboard that effectively summarizes students' learning performances. The study highlights the significance of incorporating clear and intuitive visualizations, such as line and bar charts, which enable students to engage with their performance data and make informed decisions. The availability of customization options further enhances the usability of the dashboard, allowing students to personalize their experience and tailor it to their specific needs. By adopting these best practices, educational institutions can provide students with a powerful tool to support their academic journey, facilitate data-driven decision-making, and ultimately improve student outcomes. Moving forward, future research should focus on continuous improvements in visualization techniques and explore innovative approaches to enhance the design and functionality of learning dashboards, ensuring their effectiveness in meeting the evolving needs of students in the educational landscape.

REFERENCES

1. Askinadze, A., Liebeck, M., & Conrad, S. (2019). Using Venn, Sankey, and UpSet Diagrams to Visualize Students' Study Progress Based on Exam Combinations. *9th International Conference on Learning Analytics and Knowledge*, 1–5.
2. Ebert, D. S., Favre, J. M., & Peikert, R. (2002). Data visualization. *Computers and Graphics (Pergamon)*, *26*(2), 207–208.
3. Horvath, D. M., Molontay, R., & Szabo, M. (2018). Visualizing student flows to track retention and graduation rates. *Information Visualisation - Biomedical Visualization, Visualisation on Built and Rural Environments and Geometric Modelling and Imaging, IV 2018*, *2017*(7), 338–343.
4. Ssegawa, J. K., & Muzinda, M. (2018). The Project domain framework (PDF): A framework for documenting a project concept into a business case. *Procedia Computer Science*, *138*, 747–755.
5. Toasa, R., Maximiano, M., Reis, C., & Guevara, D. (2018). Data visualization techniques for real-time information - A custom and dynamic dashboard for analyzing surveys' results.

Iberian Conference on Information Systems and Technologies, CISTI, *2018-June*(January 2019), 1–7.
6. Yu, J. (2018). Research Process on Software Development Model. *IOP Conference Series: Materials Science and Engineering*, *394*(3).
7. Yu, J., Zhang, J., Wu, N., Mei, Y., Zhang, D., Zhu, L., Zhu, C., & Zhu, S. (2021). A Functionality Testing Method Based on Exploratory Testing for Android Mobile Apps. *Journal of Physics: Conference Series*, *1856*(1).

COMPARATIVE STUDY OF STRING-MATCHING ALGORITHMS FOR DETECTING PLAGIARISM IN STUDENTS' ASSIGNMENTS

Puteri Azhar Azhar[1] and Siti Khatijah Nor Abdul Rahim[2*]

[1,2]College of Computing, Informatics and Mathematics, Universiti Technologi MARA, Shah Alam, Selangor, Malaysia

[1]2021253224@student.uitm.edu.my, [2]sitik781@uitm.edu.my

ABSTRACT

In today's world, copying something from other sources and claiming it as its own contribution is a crime. We have also seen it a major problem in academics where students copy some parts of original documents without obtaining proper permission from the author or developer to complete their assignments. Many software tools exist to find out and assist in the monotonous and time-consuming task of tracing plagiarism because identifying the owner of that whole text is practically difficult and impossible for markers. The main goal of this research work is to compare the performance of the string-matching algorithms in detecting plagiarism in students' assignments. Four algorithms have been compared in this research, which are the Boyer Moore Horspool, Knuth-Morris-Pratt, Rabin Karp and Naïve algorithms. We have focused on practical assignments (projects) as well as written documents that are to be submitted by students at the university. From the experimental results, we observed that these 4 algorithms can be used to perform plagiarism detection of those documents, but in terms of their effectiveness, the Naïve algorithm is much more effective and faster in the process of detecting plagiarism in the documents.

Keywords*: plagiarism; string matching; algorithms; detection, effectiveness*

1. INTRODUCTION

The continuous development of technology has brought about significant changes over the years. The increasing number of internet users has had a profound impact on scientific research and our global perspective. Search engines like Google have made it easier to find and access various documents and scientific sources (Sriyati, 2009). This convenience has greatly benefited users in their search for information and referencing scientific papers, contributing to the positive aspects of technological advancement. However, alongside these advantages, there are also disruptive issues to consider, one of which is plagiarism (Ulum, 2014). Plagiarism refers to the act of fraudulently copying an article without acknowledging the original source (Ezzikouri et al., 2016).

According to Neville (2016) in his book titled "The Complete Guide to Referencing and Avoiding Plagiarism," defines plagiarism as the act of taking someone else's ideas or writings without proper references and claiming them as one's own, which is considered a fraudulent practice by universities. Therefore, including proper citations and references is essential to ensure that work is not considered plagiarism. Within a document, there are sets of strings that are combined to form words or sentences. Various algorithms can be used for string matching, each with its own complexities. Through the use of string-matching techniques, document comparisons can be made to identify potential instances of plagiarism.

String matching algorithms can be categorized into three types based on the direction of the search they employ. These classifications are determined by the approach taken to compare a pattern or substring within a larger text or string. The three types are as follows:

1. Forward Search Algorithm - In this type of algorithm, the search begins from the start of the text and progresses towards the end. The pattern or substring being searched is until a match is found or the end of the text is reached. One example of a Forward Search Algorithm is the Naïve String- Matching Algorithm known as Brute Force.
2. Backward Search Algorithm -In contrast to the Forward Search Algorithm, the Backward Search Algorithm starts the search from the end of the text and moves towards the beginning. The pattern or substring is compared from right to left with the corresponding characters in the text. Similar to the Forward Search Algorithm, it mismatches result in shifting the starting position of the comparison until a match is found or the beginning of the text is reached.
3. Bidirectional Search Algorithm - The Bidirectional Search Algorithm combines elements of both Forward and Backward Search approaches. They involve searching from both ends of the text simultaneously, using multiple pointers or indices. The pattern or substring is compared in both directions, moving towards the middle of the text. This approach can offer advantages in terms of efficiency, as it can potentially reduce the search space and expedite the matching process.

The choice of which type of string-matching algorithm to use depends on various factors, such as the characteristics of the text and pattern, the expected size of the input, and the specific requirements of the application. Each type has its own strengths and weaknesses, and the selection should be based on the context in which the string-matching operation is performed.

Besides the above-mentioned algorithms, other algorithms can be used to detect the similarity of documents which are Boyer Moore Horspool, Knuth-Morris-Pratt, and Naïve algorithms. According to Wicaksono and Suyanto (2012), in order to create a plagiarism detection system, it is required a good algorithm for multiple types of string-matching patterns. One algorithm that is suitable for the problem of multiple string-matching patterns is the Rabin-Karp algorithm. The advantage of the Rabin-Karp algorithm compared to other string-matching algorithms is the ability to search for multiple string patterns (Cormen et al., 2022).

2. METHODS

The main goal of this research work is to compare the performance of the string-matching algorithms in detecting plagiarism in students' works. This study examines issues such as similarity answers, plagiarism detection, and algorithms' performance comparisons. This research focuses on students enrolled in the CSC776 course at Universiti Teknologi Mara (UiTM) as participants in which the contents of the documents of the student's assignment's answers are analyzed using the string-matching algorithms to determine the similarity percentage. The performance factors used are the time taken for searching the pattern, the number of iterations required, and its accuracy for single-word search, multiple-word search, and file search.

The students' submitted assignments are saved to the database to allow them to be retrieved easily later for analysis purposes. These submissions will serve as the main source of information for the study, allowing for the comparison of similarities between students' answers to determine if any instances of plagiarism are present. By developing a plagiarism detection tool, it can compare and analyze the similarity between text documents, and this tool will utilize a range of algorithms and techniques to identify potential cases of plagiarism. Figure 1 represents the flow of the plagiarism detection tool.

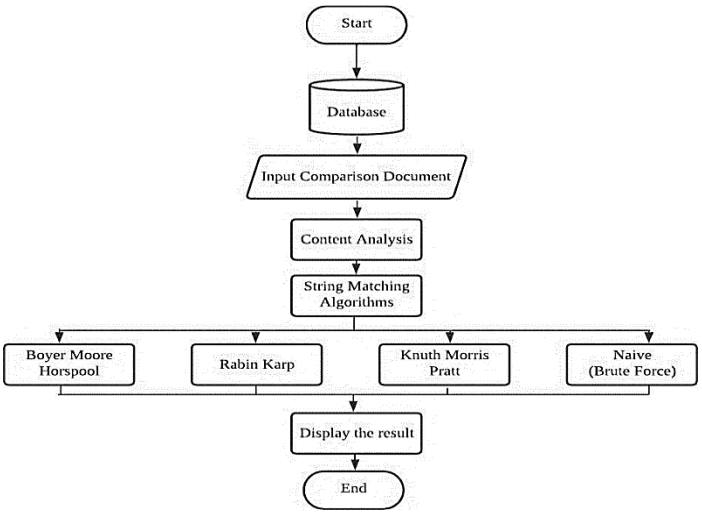

Figure 1. The flow of the plagiarism detection tool

Once the plagiarism detection tool has processed the submissions, the result will be displayed. It will highlight areas of potential similarity and indicate the sources where matches are found.

3. RESULTS AND DISCUSSION

Figure 2 shows the screenshot of the results generated from the plagiarism detection tools developed for this project in the Java Eclipse. The similarity ratio, number of the input file, number of the source file input and the amount of time taken for plagiarism detection were displayed for each algorithm used.

```
Boyer Moore Result
Similarity ratio = 0.0 Line Number of the input file= 1 Line Number of the source file input.txt = 1
---------------------------------------------------------------------------------------------------------
The amount of time Boyer Moore took for a text lenghth 67 and pattern length 53 is 53
KMP Result
Similarity ratio = 0.0 Line Number of the input file= 1 Line Number of the source file input.txt = 1
---------------------------------------------------------------------------------------------------------
The amount of time KMP took for a text lenght 67 and pattern length 53 is 67
Naive String Matching Algorithm Result
Similarity ratio = 0.0 Line Number of the input file= 1 Line Number of the source file input.txt = 1
---------------------------------------------------------------------------------------------------------
The amount of time Naive String took for a text length 67 and pattern length 53 is 15
Rabin Karp Result
Similarity ratio = 0.0 Line Number of the input file= 1 Line Number of the source file input.txt = 1
---------------------------------------------------------------------------------------------------------
The amount of time Rabin Karp took for a text lenght 67 and pattern length 53 is 172
```

Figure 2. Screenshot of the results produced in the Java Eclipse

The time taken for each algorithm, together with the time complexities are tabulated in Table 1.

Table 1: Time taken and complexities for each algorithm

Algorithm	Pre-processing	Time Complexity	Amount of time
Boyer Moore	O(M+N)	O(MN)	53s
Rabin Karp	O(N)	O(MN)	172s
Knuth Morris-Pratt	O(M)	O(M+N)	67s
Naïve (Brute Force)	None	O(MN)	15s

According to the results produced, it was observed that the Rabin Karp algorithm had the longest amount of time for a text length of 67 and pattern length of 53 which is 172s. The shortest amount of time recorded for plagiarism detection was by the Naïve String-Matching Algorithm which is 15s. Figure 3 shows the line chart for the amount of time taken by the four types of algorithms.

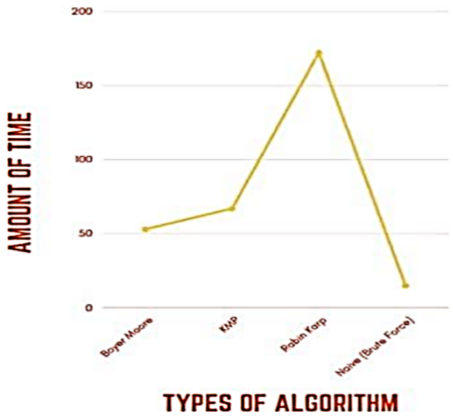

Figure 3. Line chart for the amount taken by each algorithm

There are a few findings that can be made from the analysis of the results. Firstly, both the Boyer Moore and Naïve algorithms exhibited shorter execution times compared to the other algorithms considered in the study. This suggests that these two algorithms are more efficient and capable of delivering faster results for the given text and pattern lengths. Secondly, the time complexity of both algorithms was found to be similar, indicating that they require a comparable number of computational resources to execute. This implies that the efficiency gains observed in the Boyer Moore and Naïve algorithms are not due to significant differences in their underlying computational complexity.

Furthermore, the study highlighted a particular strength of the Boyer-Moore algorithm when it comes to handling larger pattern lengths. The algorithm's design and search strategy allow it to excel in scenarios where the pattern being searched for is relatively long. This characteristic makes the Boyer-Moore algorithm a favorable choice when dealing with complex pattern-matching tasks. On the other hand, the Naïve (Brute Force) algorithm was noted for its straightforward approach. It simply compares the first character of the pattern with each character in the searchable text, making it relatively easy to implement. While it may not offer the same level of performance as the Boyer-Moore algorithm for larger pattern lengths, it can still provide satisfactory results for simpler pattern-matching requirements.

From the result and analysis, the conclusion is that Boyer Moore and Naïve (Brute Force) algorithms have less amount of time for a text length of 67 and a pattern length of 53. Both algorithms have similar time complexity. Boyer-Moore algorithm works better if the pattern

length is large. Whereas the Naïve (Brute Force) algorithm is a straightforward approach to comparing the first character of a pattern with searchable text.

4. CONCLUSIONS

In this research, the Naïve algorithm is the fast and easy-to-implement algorithm for string matching. Among the four algorithms, the Naïve algorithm is the fastest algorithm without considering the pattern length and pattern placement. Rabin Karp is the slowest algorithm when increasing the pattern length and the pattern placement. The four algorithms have significant differences in the mean time of algorithm execution. In conclusion, we observe the performances of pattern matching algorithms on txt files which the Naïve algorithm has the shortest amount of time to do the plagiarism detection. A lot of trials have been done to have meaningful results. Experimental results show that different pattern-matching algorithms give different performance factors such as the amount of time, number of pattern lengths, and similarity ratio.

REFERENCES

1. Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms*. MIT press.
2. Ezzikouri, H., Erritali, M., & Oukessou, M. (2016). Semantic similarity/relatedness for cross language plagiarism detection. In *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)* (pp. 372-374). IEEE.
3. Neville, C. (2016). EBOOK: *The Complete Guide to Referencing and Avoiding Plagiarism*. McGraw-Hill Education (UK).
4. Sriyati, T. (2009). Perkembangan internet dan strategi pemanfaatannya di perpustakaan Badan Penelitian dan Pengembangan Kehutanan. *Visi Pustaka, 12*(2), 20-24.
5. Ulum, S. (2014). Analisis plagiarisme penulisan skripsi mahasiswa lulusan tahun 2010 Jurusan Akuntansi Perguruan Tinggi X di Kota Malang/Sayidatul'Ulum (Doctoral dissertation, Universitas Negeri Malang).
6. Wicaksono, Y. A., & Suyanto, S. (2012). Analisis Dan Implementasi Algoritma Rabin-Karp Dan
   Algoritma Stemming Nazief-Adriani Pada Sistem Pendeteksi Plagiat, https://openlibrary.telkomuniversity.ac.id/pustaka/95945/analisis-dan-implementasi-algoritma-rabin-karp-dan-algoritma-stemming-nazief-adriani-pada-sistem-pendeteksi-plagiat-dokumen-teks-berbahasa-indonesia.html

UNRAVELING USER EXPERIENCE IN MOBILE BANKING: COMPARATIVE INVESTIGATION OF DATA SCALING AND DIMENSION REDUCTION TECHNIQUES

Muhammad Rafiuddin Razlat[1] and Zainura Idrus[2*]

[1,2]College of Computing, Informatics and Mathematics, Universiti Technologi MARA, Shah Alam, Selangor, Malaysia

[1]2022969601@student.uitm.edu.my, [2]zainura@fskm.uitm.edu.my

ABSTRACT

Mobile Banking (MB) offers consumers a wide range of value-added services, such as short-distance remittances, micropayments to merchants, utility bill payments, person-to-person (P2P) transfers, business-to-business (B2B) transfers, and business-to-person (B2P) transfers. Users anticipate that the MB applications are easy to use. However, this is not the case as most of the reviews on the Google Play Store and App Store express user's frustration. To delve deeper into the topic, this study explores the Malaysian experience of MB applications using the User Experience Questionnaire and machine learning (ML) technique. The scope of the study focused on the 19 banks listed by Bank Negara Malaysia (BNM) that offer MB services. User Experience (UX) examines the user's emotions triggered by the system and it is important to increase customer satisfaction and motivate user interaction. UX questionnaire was designed and unsupervised ML techniques were adopted. Before diving into machine learning, a series of investigations were carried out to examine the effectiveness of data scaling and dimension reduction. The outcomes of these experiments indicated that data scaling was not influential due to the specific characteristics of the collected dataset. On the other hand, the results suggested that only four-dimension reduction techniques warranted further investigation for the chosen clustering model. Subsequently, by completing the preliminary study on scaling techniques and dimension reduction techniques, an analysis of the UX could be conducted. This further investigation will help determine the UX evaluation of users towards the MB application.

Keywords: *mobile banking, user experience, machine learning, data scaling, dimension reduction*

1. INTRODUCTION

The convergence of fast-growing economies, access to high-speed internet, and advances in the use of mobile services have increased the popularity of mobile commerce (m-commerce) as restrictions of space and time have been removed (Huang & Tian, 2018). MB, one of the m-commerce technology offers a wide variety of services for customers such as account checking, bill payment, transferences, or mobile phone text message notifications (Mostafa, 2020). Various studies have shown that consumer interest in MB adoption is growing (Al Tarawneh et al., 2023; Shuhidan et al., 2017). Al Tarawneh et al. (2023) designed seven hypotheses to investigate antecedents of the MB. On the other hand, Shuhidan et al. (2017) examined the perceived risk of MB. Although there have been various studies on MB, the main focus has been on user adoption rather than UX. Once users have decided to use MB, the focus should be on the UX.

MB was expected to work effortlessly, but many banks offered frustrating apps. User reviews of such apps can be found on Google Play Store and App Store. There have been mixed reviews showing user frustration with the application. This frustration not only hindered the usage of the technology but also affected the emotions of the users during and after using the mobile application. Various instruments and methods are available for evaluating pragmatic and hedonic software properties (Schrepp & Thomaschewski, 2019). However, UX questionnaires like UEQ+ are a simple and inexpensive way to collect data from larger groups of users of a product. To identify hidden patterns or data clusters based on the UX questionnaire, unsupervised ML is the perfect answer for exploratory data analysis and user segmentation due to its ability to find similarities and contrasts in data (Sarker, 2021).

Based on these shortcomings, this study accessing Malaysian experience of mobile banking applications with UEQ+ pair theory and machine learning techniques. However, experiments on data scaling techniques and dimension reduction techniques are required before further data analysis using machine learning. The findings of this study provide insights into the most appropriate data scaling and dimension reduction techniques that can be used to enhance the performance of machine learning algorithms.

2. METHODS

The methodology employed in this study focused on two key areas: data scaling and dimension reduction experiments. To guide our approach, we adopted the Machine Learning Life Cycle proposed by Spjuth et al. (2021) and customized it to fit the specific requirements of our project.

Questionnaires were designed to gauge insight from users based on their demographics, function in mobile ranking that important and less important followed by UX with chosen scales such as Attractiveness, Efficiency, Dependability, Trust, Perspicuity, Value, Clarity, and Visual Aesthetics. Data were collected through an online survey from June 2022 to April 2023 with sampling method of opportunistic sampling. Over the survey period, 465 respondents participated in the study.

The analysis involved the use of ML techniques. Prior to the application of these techniques, data preprocessing and feature engineering were carried out to ensure the data was in the appropriate format and contained relevant features. This research conducted experiments on data scaling and dimension reduction techniques in order to identify the most suitable techniques for the given dataset.

3. RESULTS AND DISCUSSION

Selection of suitable scaling techniques, dimension reduction techniques, and hyperparameter tuned were investigated and evaluated based on the result performance validation of Silhouette Score (SC), Calinski-Harabasz Score (CH), Davies-Bouldin Score (DB) and computation time (Time). Linear scaling techniques (Standard Scaler, Min Max Scaler, and Robust Scaler) and non-linear scaling techniques (Logarithmic, Power Transformer, Quantile Transformer) were compared to determine suitable data scaling technique for collected dataset. Each clustering model has been trained with each data scaling technique and the computations were recorded. However, based on the computation, adoption of data scaling was found to be insignificant as dataset collected have less numerical data as compared to categorical data. This can be seen from tabulated Table 1 which shows the results of K Means algorithm with different data scaling techniques. A similar pattern was observed on other clustering models. Due to that, Standard Scaler technique will be adopted as a common approach.

Table 1. Data Scaling Performance for K Means

Scaler	SC	CH	DB	Time(s)
None	0.1602	136.28	1.72	0.37
Standard	0.1746	138.05	1.66	0.30
Min Max	0.1735	138.00	1.66	0.29
Robust	0.1735	138.00	1.66	0.28
Log Transform	0.1749	137.92	1.66	0.3
PowerTransformer(method='yeo-johnson')	0.1741	137.81	1.64	0.29
PowerTransformer(method='box-cox')	0.1745	138.04	1.67	0.29
Quantile Transformer	0.1745	138.07	1.67	0.29

In terms of dimensionality reduction, techniques such as Principal Component Analysis (PCA), Fast Independent Component Analysis (FastICA), Incremental Principal Component Analysis (IncrementalPCA), Kernel Principal Component Analysis with Radial Basis Function (KernelPCA – rbf), Kernel Principal Component Analysis with Polynomial Kernel (KernelPCA – poly), Locally Linear Embedding (LLE), Truncated Singular Value Decomposition (Truncated

SVD), Isomap, and Non-Negative Matrix Factorization (NMF) were investigated to escape the curse of dimensionality. Similar to data scaling comparison, each clustering model will be paired with each dimension reduction technique and the results will be recorded. The findings show that out of nine-dimension reduction techniques, only four were deemed to be ideal for each dedicated model as tabulated in Table 2.

Table 2. Ideal Dimension Reduction Technique for each of model

Dimension Reduction	Model
LLE	K Means, K Medoids, Agglomerative Clustering, Spectral Clustering
Isomap	Affinity Propagation, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH)
TruncatedSVD	Ordering Points To Identify the Clustering Structure (OPTICS)
FastICA	Density-based spatial clustering of applications with noise (DBSCAN)

4. CONCLUSIONS

This research examines the UX of MB in Malaysia. This research has explored different viewpoints that focus on UX, following phase after adoption. Additionally, research has adopted UEQ+, a modular extension of UEQ consisting of a larger list of UX scales that can be combined to create a concrete UX questionnaire (Schrepp & Thomaschewski, 2019). This research has examined the contribution of UEQ+ in the MB area particularly in Malaysia. Several data scaling techniques and dimensionality reduction techniques were adopted and compared using assessment metrics. The results of these experiments showed that data scaling was not a significant factor in improving the performance of the clustering model, due to the nature of the collected dataset. On the other hand, the findings show that only four dimension reduction techniques warrant further investigation. Since preliminary study on data scaling, dimension reduction technique has been done, appropriate data scaling and dimension reduction technique can be used in further study on data analysis needs to be done in order to determine current UX evaluation of user towards

REFERENCES

1. Al Tarawneh, M. m. A., Nguyen, T. P., Yong, D. G., & Dorasamy, M. A. (2023). Determinant of M-Banking Usage and Adoption among Millennials. *Sustainability*, *15*(10).
2. Huang, Z., & Tian, Z. (2018). Analysis and Design for Mobile Applications: A User Experience Approach. Design, User Experience, and Usability: Theory and Practice, Cham.
3. Mostafa, R. (2020). Mobile banking service quality: a new avenue for customer value co-creation. *International Journal of Bank Marketing*, *ahead-of-print*. https://doi.org/10.1108/IJBM-11-2019-0421
4. Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x
5. Schrepp, M., & Thomaschewski, J. (2019). Design and Validation of a Framework for the Creation of User Experience Questionnaires. *International Journal of Interactive Multimedia and Artificial Intelligence*, *InPress*, 1. https://doi.org/10.9781/ijimai.2019.06.006
6. Shuhidan, S. M., Hamidi, S. R., & Saleh, I. S. (2017). Perceived Risk towards Mobile Banking: A case study of Malaysia Young Adulthood. *IOP Conference Series: Materials Science and Engineering*, *226*(1), 012115. https://doi.org/10.1088/1757-899X/226/1/012115

SENTIMENT ANALYSIS ON TWITTER: COMPARISON OF ENSEMBLED MACHINE LEARNING MODELS

Nur Amanina Ahmad Fazli[1] and Nursuriati Jamil[2*]

[1,2]School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]2020411418@siswa.uitm.edu.my, [2]lizajamil@computer.org

ABSTRACT

This research study focuses on Twitter sentiment analysis using the Python programming language and is conducted on the Google Colab platform. The study utilizes a labeled dataset for training and testing the sentiment analysis models. Data pre-processing techniques, including stemming and lemmatization are employed to enhance the data quality. Feature extraction is performed using CountVectorizer(), enabling the conversion of textual data into numerical features. Machine learning algorithms, specifically Naive Bayes and Random Forest are utilized for sentiment analysis and their accuracy scores are compared. We also provided insights into the recall, precision, and F1-score for sentiment classification. Additionally, fine-tuning is implemented through model combination techniques, comparing multiple combinations of Naive Bayes models and combinations of Naive Bayes and Random Forest using the voting technique. This research study aims to enhance sentiment analysis accuracy on Twitter data and provides a valuable evaluation of the models' performance and fine-tuning approaches.

Keywords: *Machine Learning, Models Combination, Naïve Bayes, Performance Measure, Random Forest, Sentiment Analysis*

1. INTRODUCTION

This research focuses on sentiment analysis on Twitter, using machine learning algorithms Naïve Bayes and Random Forest. The study begins by collecting a dataset of labeled tweets and pre-processing the data through techniques like cleaning, tokenization, and normalization. Text present on these media is unstructured in nature, so to process them first we need to pre-process, and then features are extracted from the pre-processed data (Ahuja et al., 2019). Feature extraction is then performed using methods like bag-of-words and n-grams. The performance of the machine learning models is evaluated using metrics such as accuracy, precision, recall, and F1-score.

The motivation for this research lies in the significant role of social media platforms like Twitter in shaping public opinion and sentiment. Sentiment analysis can provide valuable insights into user emotions and sentiments, which has applications in various domains (Liu, 2015). However, sentiment analysis on Twitter poses unique challenges due to the informal language, abbreviations, and limited tweet length (Barbosa & Feng, 2010). Hence, the aim is to enhance the accuracy of sentiment analysis on Twitter and explore performance evaluation and fine-tuning approaches.

The objectives include comparing Naïve Bayes and Random Forest models in terms of accuracy and efficiency, exploring fine-tuning techniques and evaluating the performance of the fine-tuned models against baseline models. The scope of the study focuses on the comparison of these two models using a selected dataset and applying pre-processing techniques. The study also emphasizes the significance of understanding the strengths and weaknesses of different machine learning algorithms for sentiment analysis, not only for Twitter but also for other text classification tasks. By comparing the performance of Naïve Bayes and Random Forest models

on a common dataset, researchers can identify which algorithm performs better under different conditions. Additionally, there is a need to address the challenge of analyzing sentiment on Twitter by developing novel techniques and algorithms. The research questions revolve around comparing ensemble methods, fine-tuning machine learning models, and evaluating the usefulness of different evaluation metrics.

The significance of this research lies in its potential to develop more effective sentiment analysis systems for Twitter, which can have applications in marketing, politics, and social media monitoring. The comparison of algorithms contributes to the broader understanding of text classification while addressing the challenges specific to sentiment analysis on Twitter.

2. METHODS

This study conducted a Twitter sentiment analysis using Python language and perform in the Google Colab platform. It utilizes a dataset obtained from Kaggle, which consists of tweets categorized into positive, negative, and neutral sentiments. The dataset contains 27,481 tweets. The dataset is read, and data cleaning techniques are applied to remove noise and unnecessary words from the tweets. The dataset utilized for the study is pre-processed and divided into training and testing subsets. For the data cleaning process, it involves removing stop words, punctuation, and other unwanted characters (Muriel Kosaka, 2020).

Next, the sentiment classes are transformed into numerical values, and feature extraction is performed using the *CountVectorizer* method. The maximum number of features is set to 3000, capturing the most frequent words. After the cleaned data is converted into the sparse matrix, the two machine learning algorithms, Naïve Bayes, and Random Forest are employed for sentiment analysis. The performance of the models is evaluated using metrics such as accuracy, precision, recall, and F1-score. A classification report shows the results, and they are compared based on accuracy scores.

Further analysis involves fine-tuning of the less-performing model through model combination techniques. A voting technique is employed to combine the predictions of multiple models and improve sentiment analysis accuracy (Cui et al., 2023). By following this methodology, the research study aims to compare the performance of Naïve Bayes and Random Forest models, explore model combination techniques, and evaluate the effectiveness of sentiment analysis on Twitter data.

3. RESULTS AND DISCUSSION

The analysis and results of the study indicate that the Random Forest model outperformed the Naïve Bayes model in terms of accuracy, with an improvement of 3%. The classification reports for both models show precision, recall, and F1-score for each sentiment class. For the negative sentiment, the Random Forest model achieved higher precision and recall compared to Naïve Bayes. For the neutral sentiment, both models had similar performances. However, for positive sentiment, the Random Forest model showed higher precision and lower recall compared to Naïve Bayes.

Table 1. Overall results for Naïve Bayes and Random Forest according to Classification Report

Algorithm	Sentiment Class	Precision	Recall	F1-score	Support	Accuracy (F1-score)
Naïve Bayes	0	0.66	0.61	0.63	2,296	0.65
	1	0.60	0.67	0.63	3,323	
	2	0.72	0.67	0.69	2,626	
Random Forest	0	0.69	0.64	0.66	2,250	0.70
	1	0.69	0.68	0.69	3,410	
	2	0.72	0.77	0.74	2,585	

The fine-tuning process involved combining multiple Naïve Bayes models and combining Naïve Bayes with Random Forest. However, combining multiple Naïve Bayes models did not significantly improve the accuracy. On the other hand, combining Naïve Bayes with Random Forest led to an accuracy increase of 3%. The classification reports for the combined model showed improvements in precision, recall, and F1-score for certain sentiment classes.

The comparison of precision, recall, and F1-score between the original Naïve Bayes model and the combined Naïve Bayes and Random Forest model revealed improvements in some cases. The analysis also included visualizations of the precision, recall, and F1 score for each sentiment class. In conclusion, the study showed that the Random Forest model had better performance than Naïve Bayes in terms of accuracy. The combination of Naïve Bayes and Random Forest led to further improvements in accuracy and certain performance metrics.

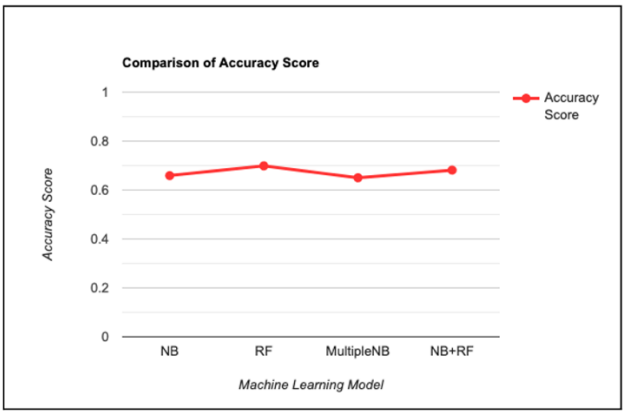

Figure 1. Line graph of accuracy score for four different models

4. CONCLUSIONS

In conclusion, fine-tuning a supervised machine-learning model has successfully improved Twitter sentiment analysis. Combining Naive Bayes and Random Forest models increased the accuracy of the less-performing model. Precision, recall, and F1-score metrics also improved for specific sentiment classes. The combination model technique showed better performance, increasing the overall accuracy score from 0.65 to 0.68. However, combining multiple models of the same type did not yield significant differences.

One limitation of this research is the reliance on a specific dataset obtained from Kaggle. The dataset used may have certain biases or limitations, such as the inclusion of specific types of tweets or a limited range of sentiments. This restricts the generalizability of the findings to a broader range of Twitter data. Using a more diverse and representative dataset could enhance the robustness and applicability of the research results. Future enhancements can focus on improving data preprocessing techniques to ensure reliable data. Additionally, exploring

advanced deep learning models for sentiment analysis can be beneficial, as they have shown promising results in various natural language tasks.

REFERENCES

1. Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, *152*, 341–348. https://doi.org/10.1016/j.procs.2019.05.008
2. Barbosa, L., & Feng, J. (2010). *Robust Sentiment Detection on Twitter from Biased and Noisy Data*. http://twittersentiment.appspot.com/
3. Cui, S., Han, Y., Duan, Y., Li, Y., Zhu, S., & Song, C. (2023). A Two-Stage Voting-Boosting Technique for Ensemble Learning in Social Network Sentiment Classification. *Entropy*, *25*(4), 555. https://doi.org/10.3390/e25040555
4. Liu, B. (2015). *Sentiment Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9781139084789
5. Muriel Kosaka. (2020). *Cleaning & Preprocessing Text Data for Sentiment Analysis*. https://towardsdatascience.com/cleaning-preprocessing-text-data-for-sentiment-analysis-382a41f15 0d6

EXTENDED ABSTRACT OF ENHANCING VPN LOAD BALANCING WITH EIGRP ROUTING PROTOCOL

Ainin Shuhada Alwi[1] and Zolidah Kasiran[2*]

[1,2] *School of Computing Sciences, College of Computing, Informatics and Mathematics*
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]shuhadaainin@gmail.com, [2]zolidah@tmsk.uitm.edu.my

ABSTRACT

A virtual private network (VPN) is a secure and encrypted internet connection that connects two networks or devices. VPN is widely used nowadays however there are some critical issues such as additional overhead, security vulnerabilities and scalability. Due to added overhead caused by encryption and encapsulation, VPNs may result in decreased network throughput and performance. There are many ways of load balance such as Round Robin, Least Connection and IP Hash. However, those techniques often have issues that need to be improvised over time such as inefficient load distribution, single point failure and configuration complexity. This research aims to examine the network performance of VPN load balancing with the Enhanced Interior Gateway Routing Protocol (EIGRP). This includes examine the network performance in VPN load balancing, implement EIGRP in VPN load balancing for improving the network performance and evaluate the performance of VPN load balancing with the EIGRP routing protocol in terms of throughput, jitter, packet loss, speed, and latency. This research will use two scenarios which are with and without EIGRP to compare the network performance. This way we can see the difference whether EIGRP can improve VPN load balancing network performance. A few testing such as web, email, video conferencing, video streaming and file transfer will be done to measure the performance metric mentioned above. It is expected that EIGRP can improve the network performance and can be used as another method to improve VPN load balancing.

Keywords: *EIGRP, Load Balancing, Routing Protocol, VPN*

1. INTRODUCTION

A virtual private network (VPN) is a secure and encrypted internet connection that connects two networks or devices. Abdulazeez et al. (2020) in his article mentioned that VPNs provide scalable and flexible functionality, competent bandwidth management, secure connections, and remote access. Load balancing in a virtual private network (VPN) is the process of distributing a VPN's workload or traffic among multiple servers or nodes to improve performance and reduce server overloading. Load balancing guarantees that each server only handles a fraction of the overall workload, lowering the chance of server breakdowns or unavailability.

VPN is used nowadays however there are some critical issues such as additional overhead. Due to added overhead caused by encryption and encapsulation, VPNs may result in decreased network throughput and performance. Further affecting VPN performance are bandwidth restrictions and network congestion, which can worsen latency and user experience. There are many ways of load balance such as Round Robin, Least Connection, IP Hash and Weighted Least Connection. However, those techniques often have issues that need to be improvised over time. This research aims to examine the network performance of VPN load balancing with the Enhanced Interior Gateway Routing Protocol (EIGRP) routing protocol. The findings will provide insights into the most effective load balancing methods as well as the advantages of using EIGRP as a routing protocol, allowing organisations to optimise their VPN networks for increased throughput, reduced latency, and reduced packet loss.

Norazlan Nur Fatin Nadhirah et al. (2020) and Saputra et al. (2023) uses OSPF as their technique, Inderjeet Singh & Damandeep Kaur (2022) uses EIGRP but in different scenarios. J

Joys Nancy et al. (2020) uses the current from Arduino microcontroller to balance load balancing in unbalanced problem. Meanwhile Sherlin Solomi et al. (2021) implement hub and spoke topology in a VPN utilising EIGRP and Said et al. (2022) compares the performance of IPsec and without IPsec on DMVPN using BGP.

2. METHODS

The research follows a systematic approach, comprising multiple phases to evaluate the effectiveness of EIGRP implementation in VPN load balancing. Phase I which is the conceptual study phase involves an in-depth review of existing load balancing techniques and their limitations, highlighting the need for more comprehensive solutions. Phase II involves the environment setup without EIGRP, and relevant network performance metrics are identified, forming the basis for subsequent analysis. The network performance of VPN load balancing is tested in this phase. Phase III is the protocol implementation phase which EIGRP is used as routing protocol. Phase IV involves the performance evaluation of VPN load balancing with EIGRP, and the performances is analysed. Phase V is the results and findings. The results and their implications is presented.

3. RESULTS AND DISCUSSION

This study uses relevant network performance metrics such as throughput, jitter, packet loss, speed, and latency. The results are divided into two which are without and with implementation of EIGRP. Based on the results will be seen which scenarios are better in improving VPN load balancing.

The use of EIGRP in a VPN load balancing is expected to enable more intelligent and adaptable routing decisions. EIGRP continuously monitors network circumstances, allowing it to respond swiftly to changes such as server failures or network congestion and divert traffic to available resources. This improves load distribution and guarantees that the VPN service runs properly even when demand levels vary.

The study also emphasizes the advantages of EIGRP's hierarchical design and route summarization. This allows the VPN load balancing system to manage routing information more efficiently, lowering the size of routing tables and increasing system scalability. As the VPN service expands to accommodate additional users and resources, the implementation of EIGRP ensures that the system can handle rising traffic needs effectively and efficiently.

4. CONCLUSIONS

Finally, the study on improving VPN load balancing with EIGRP is expected to give a persuasive approach to improving the performance, reliability, and security of VPN services. The study demonstrates how dynamic path selection, fast convergence, and effective resource utilization may dramatically improve the overall VPN experience by integrating EIGRP's advanced routing capabilities into the load balancing system.

The research's projected outcomes indicate a number of good outcomes. EIGRP adoption is expected to result in enhanced load distribution among VPN servers, ensuring that resources are used effectively, and network traffic is dispersed equally. EIGRP's faster convergence helps the load balancing system to swiftly respond to network changes, decreasing service disruptions and improving user experience.

REFERENCES

1. Abdulazeez, A. M., Salim, B. W., Zeebaree, D. Q., & Doghramachi, D. (2020). Comparison of VPN Protocols at Network Layer Focusing on Wire Guard Protocol. *International Journal of Interactive Mobile Technologies*, 14(18), 157–177.
2. Inderjeet Singh, & Damandeep Kaur. (2021). Virtual Private Network Load Balancing Using EIGRP Routing. In *iJournals: International Journal of Software & Hardware Research in Engineering (IJSHRE)* 4(11) 61-65

3. Nancy, J.J, Mani, T., Rohith, S., Saranraj, S. & Vigneswaran,, T. "Load Balancing using Load Sharing Technique in Distribution System," *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2020, pp. 791-794
4. Norazlan, N. F. N., Rahman, R. A., & Kassim, M. (2020, April). Virtual Private Network Load Balancing Using OSPF Routing. In *2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (pp. 164-169). IEEE.
5. Said, M. A., & Jadied, E. M. (2022). Analysis of IPsec Implementation on Dynamic Multipoint VPN Protocol Using Routing Border Gateway Protocol. *Building of Informatics, Technology and Science (BITS)*, *4*(2), 595-605.
6. Saputra, O. T., Iskandar Mulyana, D., & Akbar, Y. (2023). *Optimasi Pembagian Beban Dan Keamanan Network Menggunakan OpenVPN Virtual Private Network Menggunakan OSPF Routing Protocol*. Progresif: Jurnal Ilmiah Komputer, 19 (1) 61-70,
7. Solomi, V. S., Stela, D., & Sandhiya, D. (2021, March). Implementation of HUB and Spoke Topology in VPN Using EIGRP. In *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 135-142). IEEE.

A SYSTEMATIC LITERATURE REVIEW ON HYBRID ENCRYPTION KEY FOR SECURING DATABASE

Nur Asyiqin Mohd Daresani[1] and Saidatul Izyanie Kamarudin[2*]

[1,2] School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]nurasyiqin0194@gmail.com, [2]saidatulizyanie@uitm.edu.my

ABSTRACT

Securing employee databases is vital for businesses, but current strategies face vulnerabilities to breaches and insider threats. Encryption offers a reliable solution. According to Kerckhoffs' principle, the encryption strategy remains secure even if the attacker knows everything except the key. Thus, key design is crucial in cryptography. Hybrid key, combining Fernet and Diffie-Hellman, is proposed to protect employee databases. The study aims to analyze single cryptographic key protection, implement the hybrid key, and evaluate its performance using Python. Findings show that the hybrid key addresses single key limitations by combining symmetric and asymmetric strengths, enhancing security and resilience against attacks. It simplifies key management, adapts to different scenarios, and provides a robust security solution for employee databases. The research contributes to understanding the advantages of hybrid keys and implementing it in database security.

Keywords: *Symmetric encryption, Asymmetric encryption, Hybrid key, Database security*

1. INTRODUCTION

Employee databases have valuable information and confidential data of employees. It is targeted by cybercriminals or attackers for their own benefit, leading to data breaches and various risks. One of the best ways to protect is by using cryptography. Traditional cryptographic systems often employ a single key for both encryption and decryption, leaving past and future communications vulnerable if the key is compromised. Furthermore, managing a single key across multiple devices and users can prove challenging, raising concerns about accidental sharing or mishandling. Organizations opt for hybrid key by combining the strength of symmetric and asymmetric key to address these limitations. The objective of this research is to study on hybrid key approach to securing data. The vulnerabilities of a single key and the advantages of hybrid key will be analyzed in terms of security and performance efficiency.

Several recent studies have explored encryption techniques for securing databases. Hota (2020) applied the AES algorithm but faced challenges in managing the secret key. Boicea et al., (2017) studied RSA, ElGamal, and ECIES for database encryption. Ermatita et al. (2020), Somasundara Rao et al., (2021), and Arif et al.(2022) proposed hybrid encryption approaches using combinations of AES, Diffie-Hellman, Triangle Chain Cipher, and Vigenere Cipher. However, these studies lacked performance evaluations. Zaw et al. (2019) utilized AES and ECC for database encryption but experienced longer encryption times and decreased performance due to managing multiple keys.

Vidhya et al. (2019) researched hybrid key generation for RSA and ECC. By utilizing an Exclusive OR (XOR), the RSA private and public keys can be coupled with an ECC private and public key to generate the new hybrid private and public keys. Yusfrizal et al. (2019) researched on key management which combines Diffie-Hellman key exchange with AES encryption. Kurt et al. (2020) proposed a hybrid key creation and key verification technique in which the exposed information during the key verification procedure is negligible and the confirmed keys are equal by combining physical layer (PHY) key generation with an embedded key.

2. METHODOLOGY

The study begins with a comprehensive literature review as the foundation for the research. By fully utilizing UiTM library E-Resources, data can be easily collected from various online databases such as Scopus, IEEE Explore, Science Direct, and ResearchGate. The keyword chosen is “symmetric encryption”, asymmetric encryption”, “hybrid key”, and “database”. It involves an in-depth exploration of existing journals or studies related to symmetric, asymmetric, and hybrid key in securing databases or data storage.

The result of data collection was stored in Mendeley, a reference management tool. This tool helps to manage a large number of references and removes duplicate references. The inclusion and exclusion criteria as below:

- Inclusion Criteria: The manuscript must be oriented to the application of these three types of encryptions to database or data storage. Other than that, the manuscript must be a full paper.
- Exclusion Criteria: Manuscripts were excluded if published before 2017. The articles published on some public platforms such as GitHub are also excluded.

3. RESULTS AND DISCUSSION

A first screening of excluded articles was conducted based on titles, abstracts, and keywords. The second screening was conducted by reading the full text and a quality assessment (in terms of domain, tools and algorithm). 14 articles met the inclusion criteria from a total of 50 studies selected.

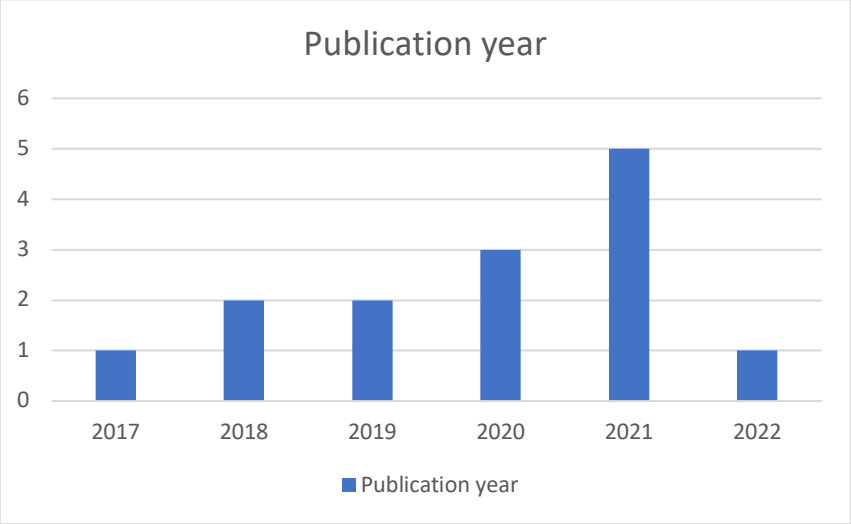

Figure 1. Publication year of 14 selected articles

4. CONCLUSIONS

A hybrid key really helps overcome the disadvantage of a single key. It provides a higher level of security and makes it more resilient against attack by combining symmetric and asymmetric key. By leveraging the strengths of both symmetric and asymmetric encryption, the hybrid key offers enhanced security and resilience against sophisticated attacks, making it a robust solution for data protection. The research underlines that the hybrid encryption key holds promise as a robust security solution, contributing to database security and cryptography. Its potential application in protecting employee databases and overcoming the vulnerabilities of a single key system makes it an important area of study for future advancements in data security.

As technology evolves, the hybrid key approach can serve as a foundation for further innovations in securing sensitive information and mitigating risks in an ever-changing threat landscape.

REFERENCES

1. Arif, A., Perdana, A., & Budiman, A. (2022). Implementation Combination Cryptographic Algorithm Triangle Chain Cipher and Vigenere Cipher In Securing Data in Database. *Sinkron*, *7*(1), 100–110. https://doi.org/10.33395/sinkron.v7i1.11236
2. Boicea, A., Radulescu, F., Truica, C. O., & Costea, C. (2017). Database Encryption Using Asymmetric Keys: A Case Study. *Proceedings - 2017 21st International Conference on Control Systems and Computer, CSCS 2017*, *May*, 317–323. https://doi.org/10.1109/CSCS.2017.50
3. Ermatita, Bayu Prastyo, Y., Pradnyana, I. W. W., & Adrezo, M. (2020). Diffie-Hellman Algorithm for Securing Medical Record Data Encryption keys. *Proceedings - 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020*, 296–300. https://doi.org/10.1109/ICIMCIS51567.2020.9354297
4. Hota, S. (2020). *Database Security By Applying Aes Algorithm : a Review*. *11*(11), 1809–1814. https://doi.org/10.34218/IJARET.11.11.2020.
5. Kurt, G. K., Khosroshahi, Y., Ozdemir, E., Tavakkoli, N., & Topal, O. A. (2020). A hybrid key generation and a verification scheme. *IEEE Transactions on Industrial Informatics*, *16*(1), 703–714. https://doi.org/10.1109/TII.2019.2950465
6. Somasundara Rao, M., Venkata Rao, K., & Krishna Prasad, M. H. M. (2021). Hybrid Security Approach for Database Security using Diffusion based cryptography and Diffie-Hellman key exchange Algorithm. *Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2021*, 1608–1612. https://doi.org/10.1109/I-SMAC52330.2021.9640762
7. Vidhya, E., Sivabalan, S., & Rathipriya, R. (2019). Hybrid Key Generation for RSA and ECC. *Proceedings of the 4th International Conference on Communication and Electronics Systems, ICCES 2019*, *2*, 35–40. https://doi.org/10.1109/ICCES45898.2019.9002197
8. Yusfrizal, Y., Meizar, A., Kurniawan, H., & Agustin, F. (2019). Key Management Using Combination of Diffie-Hellman Key Exchange with AES Encryption. *2018 6th International Conference on Cyber and IT Service Management, CITSM 2018*, *Citsm*, 1–6. https://doi.org/10.1109/CITSM.2018.8674278
9. Zaw, T. M., Thant, M., & Bezzateev, S. V. (2019). Database Security with AES Encryption, Elliptic Curve Encryption and Signature. *Wave Electronics and Its Application in Information and Telecommunication Systems, WECONF - Conference Proceedings*, *978*, 1–6. https://doi.org/10.1109/WECONF.2019.8840125

INTEGRATING UNDERWATER SIMULATION (UWSIM) AND ROBOTIC OPERATING SYSTEM (ROS) FRAMEWORK FOR UNDERWATER WIRELESS COMMUNICATION

Laila Hanim Mohd Khairuzin [1] and Nor Azimah Khalid[2*]

[1,2] School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[1] 2022605602@student.uitm.edu.my, [2] azimahkhalid@uitm.edu.my

ABSTRACT

Underwater wireless networks communication refers to the exchange of data and information between network devices that is submerged in underwater environments. Underwater performances impose significant challenges that are affected by various underwater factors such as waves, water temperature and water pressure. This will result in high attenuation, limited bandwidth, long propagation delays, and signal scattering of the wireless connections between the devices. To mitigate these challenges, innovative research studies are conducted to understand the concept of underwater wireless network issues. New strategies and techniques in improving the reliability and efficiency of data transmission in the water are constructed using simulation using frameworks integration of UWSim and ROS. This research aims to analyse the factors contributing to low transfer rates, high latency, and packets loss in underwater wireless networks, to integrate UWSim and ROS for improving key performance metrics such as transfer rates, latency, and packets loss, to evaluate the algorithms based on the performance metrics for enhancing overall communication reliability. This can provide an overview of how to correlate parameters such as waves, water temperature and water pressure with data rate transfer, latency, and packet loss. Extensive experiments will be conducted to assess and validate performances through algorithm enhancement. Experimental data will be collected from two scenarios which are before the algorithm is enhanced and after enhancing the algorithm. Then analysis will be performed based on the results simulated from the scenarios and conclusion will be discussed to find if the enhanced algorithm will improve the underwater wireless network performance. The insights and knowledge presented in this review can assist researchers, practitioners, and policy makers in understanding the challenges faced by underwater wireless networks and identifying promising pathways for future research.

Keywords: *ROS, underwater, UWSim, underwater wireless network communication*

1. INTRODUCTION

Underwater wireless networks are gaining importance due to their potential in various applications such as oceanographic research, underwater exploration, environmental monitoring, and tourism. Unlike terrestrial wireless networks, underwater wireless networks face significant challenges in achieving reliable communication. These challenges are primarily contributed by the characteristics of the underwater environment such as waves, water temperature and water pressure. Hence causes impact on the transfer rates, latency, and packet loss.

Transfer rates, refer to the speed at which data can be transmitted through a network or communication channel. It measures how much data can be transmitted within a given period. Transfer rates are typically expressed in bits per second (bps), kilobits per second (kbps), megabits per second (Mbps), or gigabits per second (Gbps). Low transfer rates in underwater wireless networks resulting in slow data transmission, impacting real-time communication, and delaying critical information dissemination.

Latency is the time taken for a data packet to travel from the source to the destination in a network. It also represents the delay or time lag experienced in transmitting data across a network

but is typically measured in milliseconds (ms). High latency in communication, further expands the problem by introducing significant lags in sending and receiving data packets. This delay hampers the responsiveness and efficiency of communication systems operating in underwater environments.

Packet loss refers to the loss of data packets during transmission across a network. Packets loss happens when data packets fail to reach their intended destinations, poses a severe challenge to reliable communication. Packet loss is typically expressed as a percentage, representing the ratio of lost packets to the total number of packets sent. Lost packets can result in missing or incomplete data at the receiving end, leading to degraded performance or disruptions in real-time applications.

2. METHODS

Research framework is used to help in conveyed this research meaningfully. There are four phases of critical components of this framework which are conceptual study, factor analysis, framework integration and performance evaluation. In understanding the basis of the ideas of underwater wireless networks, conceptual study is being done. The outcomes of these conceptual studies are expected to assist in identifying the research problems and research gaps. From the preliminary studies, the factor analysis of underwater challenges is identified. Transfer rates, latency, and packet loss are the aspects that can measure the underwater challenges. In this research, it will be determined by using DTR, RTT and Packet Loss. Framework integration of UWSim and ROS phase is constructed to study the impact of waves, water temperature and water pressure towards the transfer rates, latency, and packet loss. This framework integration will help in creating realistic underwater environment and wireless network simulation that can be modified using algorithms, control mechanisms and network protocols. Performance evaluation is the last component that is performed in proving these research problems and answering the research questions. This is where the wireless underwater reliabilities are determined by running the experiment using a selected algorithm. Data is then collected, and analysis will be performed based on the result. Inclusion criteria of this research involve peer-reviewed articles and conference papers focusing on underwater network challenges and UWSim-ROS related works that address key performance metrics. Exclusion criteria exclude duplicates, non-English publications, and studies lacking methodological precision or those not directly related to the topic.

3. RESULTS AND DISCUSSION

Transfer rates, latency, and packet loss are the basis factors of network communication. These factors are determined by using DTR, RTT and Packet Loss. DTR can be viewed as the speed of a given amount of data travelling from one place to another. Latency refers to the time delay between sending a signal and receiving a response. Whereas packet loss occurs when data packets travelling through a network fail to reach their destination. Figure 1 below shows how the UWSim and ROS integration will be deployed for this study. ROS libraries packages are used to interact with the network components in the UWSim simulation environment. In the ROS workspace, ROS nodes are created to interact with UWSim. UWSim provides ROS interfaces to publish specific simulation underwater environments. Data is collected by running the algorithm and creating scenarios into two which are before the algorithm is enhanced and after the algorithm is enhanced. This experiment will be carried out repetitively to see the pattern and trends in the DTR, RTT, and packet loss.

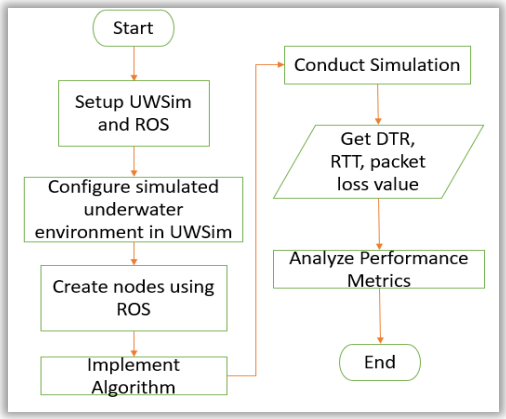

Figure 1. Flow Chart Diagram of Analysing Underwater Network Performances using UWSim with ROS.

4. CONCLUSIONS

Analyzing the factors that contribute to the communication challenges that cause low transfer rates, high latency, and packet loss in underwater wireless networks is important in having reliable underwater communication. Integration of the UWSim simulation tool with the Robotic Operating System (ROS) framework is used to evaluate the performance of underwater wireless networks. It involves creating realistic simulation environments that replicate the underwater conditions and deploying ROS-based systems for experimental studies. These extensive simulations and experiments using the selected algorithms can measure key performance metrics such as transfer rates, latency, and packet loss.

REFERENCES

1. Vieira, M. A., Vieira, A. B., & Nacif, J. A. M. (2019). Water ping: ICMP for the internet of underwater things. Computer Networks. *Computer Networks*, *152*, 54-63.
2. Centelles, D., Soriano, A., Martí, J. V., Marin, R., & Sanz, P. J. (2019, June). UWSim-NET: An open-source framework for experimentation in communications for underwater robotics. *In OCEANS 2019-Marseille* (pp. 1-8). IEEE.
3. Zedini, E., Oubei, H. M., Kammoun, A., Hamdi, M., Ooi, B. S., & Alouini, M. S. (2019). Unified statistical channel model for turbulence-induced fading in underwater wireless optical communication systems. *IEEE Transactions on Communications*, 67(4), 2893-2907.
4. Jouhari, M., Ibrahimi, K., Tembine, H., & Ben-Othman, J. (2019). Underwater wireless sensor networks: A survey on enabling technologies, localization protocols, and internet of underwater things. *IEEE Access*, 7, 96879-96899.
5. Kumar, V., & Sinha, V. K. (2020, December). Underwater wireless sensor network routing protocols: The survey. *In 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)* (pp. 359-362). IEEE.
6. Jung, L. T., & Abdullah, A. B. (2011, June). Underwater wireless network energy efficiency and optimal data packet size. *In International Conference on Electrical, Control and Computer Engineering 2011 (InECCE)* (pp. 178-182). IEEE.
7. Teixeira, F. B., Moreira, N., Abreu, N., Ferreira, B., Ricardo, M., & Campos, R. (2020, October). Udmsim: A simulation platform for underwater data muling communications. *In 2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* (pp. 1-6). IEEE.
8. Tang, Q., Chen, D., Li, Y., & Tang, L. (2019, July). Three-dimensional Visual Simulation for Trajectory Tracking of Autonomous Underwater Vehicle Based on UWSim. *In 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)* (pp. 1175-1180). IEEE.

IMPROVING DATA SECURITY THROUGH SITE-TO-SITE VPN IN HYBRID ENCRYPTION

[1]**Nurul Izzati Ramli[1] and Shapina Abdullah[2*]**

[1,2]*College of Computing, Informatics and Mathematics, Universiti Technologi MARA, Shah Alam, Selangor, Malaysia*

[1]nirkizzati@gmail.com, [2]piena@tmsk.uitm.edu.my

ABSTRACT

Cloud computing has become essential for modern businesses due to its vast data storage and computing capabilities over the Internet. Data security is a significant problem that may be improved by a variety of encryption techniques. The goal of this study is to enhance data security during transmission by examining symmetric and asymmetric key encryption techniques. In comparison to asymmetric encryption, symmetric encryption is less secure even though it can encrypt huge amounts of data quickly. The implementation of hybrid ECC and Aes-128 encryption is examined in this study, and their performance in data encryption and decryption is evaluated depending on variables such as data type, size, and speed. The hybrid approach combines the strengths of both algorithms, utilizing AES-128 for fast symmetric encryption and ECC for efficient key exchange and asymmetric encryption. This combination ensures compatibility with existing systems while benefiting from ECC's advantages. The study employs a quantitative approach, including a literature review to identify key theoretical frameworks and related studies in data security and encryption.

Keywords: *encryption algorithm, hybrid encryption, AES-128, ECC*

1. INTRODUCTION

Site-to-site VPNs are crucial for secure communication and connectivity between multiple sites, extending private networks over the public internet for seamless data exchange across dispersed locations. They are closely related to cloud computing as they facilitate secure connections between various locations and cloud resources, ensuring the integration of on-premises networks with cloud-based services. In a VPN setup, user data is encrypted on their device and securely transmitted to the VPN server through an initial connection negotiation, using encryption protocols like AES and 3DES. Once a secure tunnel is established, data is encrypted on the user's device, sent to the server, and decrypted on the server side, allowing secure processing and communication, ensuring the confidentiality and protection of sensitive information during transmission. This encryption safeguards data from eavesdropping or unauthorized access, making it inaccessible to potential attackers or malicious actors, preserving the privacy of activities like browsing, file transfers, and password exchanges.

According to Jaspin et al. (2021), a double encryption technique utilizing AES and RSA algorithms is proposed for data protection, demonstrating high security and resistance against propagation errors. In Lee et al.'s paper (2018), they focus on data security in cloud computing using Heraku cloud, concluding that AES cryptography is effective for data security, although larger data sizes result in increased encryption delay. Daud et al. (2018) conduct a performance analysis of network technology, specifically Dynamic Multipoint Virtual Private Network (DMVPN), utilizing both hardware (Cisco routers) and software (GNS3) in their experiments. They also compare encryption techniques (AES, 3DES, DES) within the DMVPN context, assessing parameters like throughput, jitter, and packet loss for data and video streaming.

Cryptography relies on cryptographic algorithms, mathematical functions for data encryption and decryption. There are two main types of cryptography: symmetric, which uses a single key for both encryption and decryption, offering speed but posing security risks if the key

is compromised, and asymmetric, which employs separate key pairs for encoding and decoding, enhancing security when managed correctly. In Hodowu et al.'s (2020) work, a two-level cryptographic approach and model are introduced to enhance data security in cloud computing. This model employs both symmetric and asymmetric encryption algorithms, specifically AES and ECC, to bolster data security by preventing unauthorized access and ensuring data privacy and integrity. This approach not only speeds up cryptographic tasks but also increases user trust in cloud computing. The use of ECC results in faster encryption with smaller key sizes while maintaining security, and the implementation of ECDSA further contributes to data integrity checks. In a separate work by Sasipraba et al. (2022), a prime crossover technique was introduced to encrypt files and store them in multiple cloud environments to bolster data security. The encrypted files are distributed across multiple cloud locations, and data retrieval is only decrypted by the end user or data owner to ensure enhanced security.

In conclusion, this study provides relevant knowledge regarding improving virtual private network (VPN) data security and resilience. In client-server systems, the study proposal compares hybrid encryption which combines symmetric and asymmetric techniques encryption. The study describes examines that evaluate the reliability of hybrid ECC and AES-128 encryption in controlled settings. Metrics including data size, type, and encryption/decryption time are used, along with the possibility of adjusting key sizes or evaluating performance in a range of workloads.

2. SYSTEM AND MODELLING

Several studies have been conducted on encryption, examining various cryptographic algorithm and their impact on security and performance. Figure 1 presents the suggested task using a research framework. To better organize and comprehend each action that must be performed to finish the intended job, it is separated into stages. This study compares and evaluates performance and security properties using hybrid encryption techniques. The performance evaluation aims to measure the computational efficiency and speed of encryption and decryption processes in a network. There are 5 phases involved in the research framework, which is Conceptual Study, Data Analysis, Design and Modelling, Implementation and Performance Evaluation. Diagram 1 shows the activities involved during each of the phase of research framework:

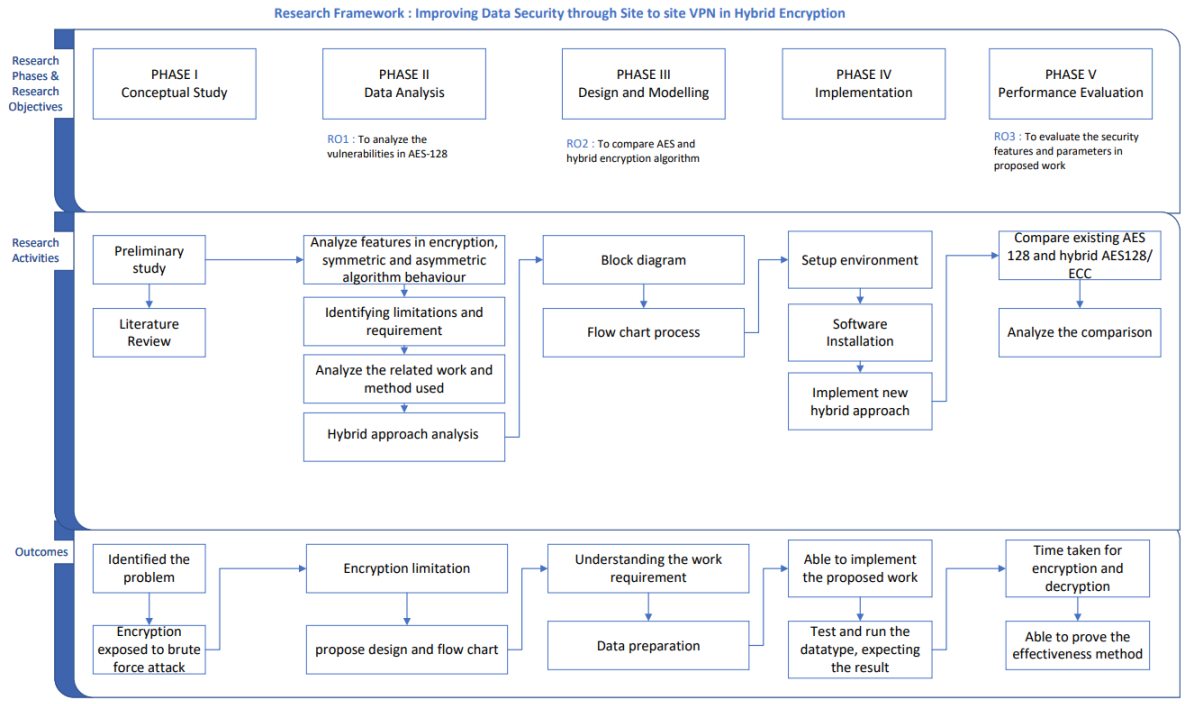

Figure 1. Research Framework

This work is a conceptual study that lays the foundation for a research project by exploring the theoretical and conceptual aspects of the research. The preliminary stage involves a thorough literature review to gather relevant ideas, concepts, and frameworks. In Phase II, data analysis will focus on comparing the behaviors of symmetric and asymmetric encryption to understand their strengths and weaknesses, aligning with Research Objective 1. The Design and Modeling phase will outline the specifics of the proposed AES-128 and ECC hybrid Model and data preparation, with a particular emphasis on the integration of ECC with AES. The implementation phase will involve installation and configuration, utilizing Python for development and system evaluation. The expected results, based on experiments conducted in Phase IV, will highlight the advantages of hybrid encryption in site-to-site VPNs, combining asymmetric encryption for secure key exchange and symmetric encryption for bulk data transmission.

The design details for the AES-128 and ECC hybrid model, as well as data preparation, are included in this section. In order to meet comprehension, following discussion of the flow process and presentation layer will focus on the algorithm used in this technique and emphasize the significance of combining ECC with AES. A high-level overview of the AES-ECC encryption and decryption procedure is shown in Figure 2.

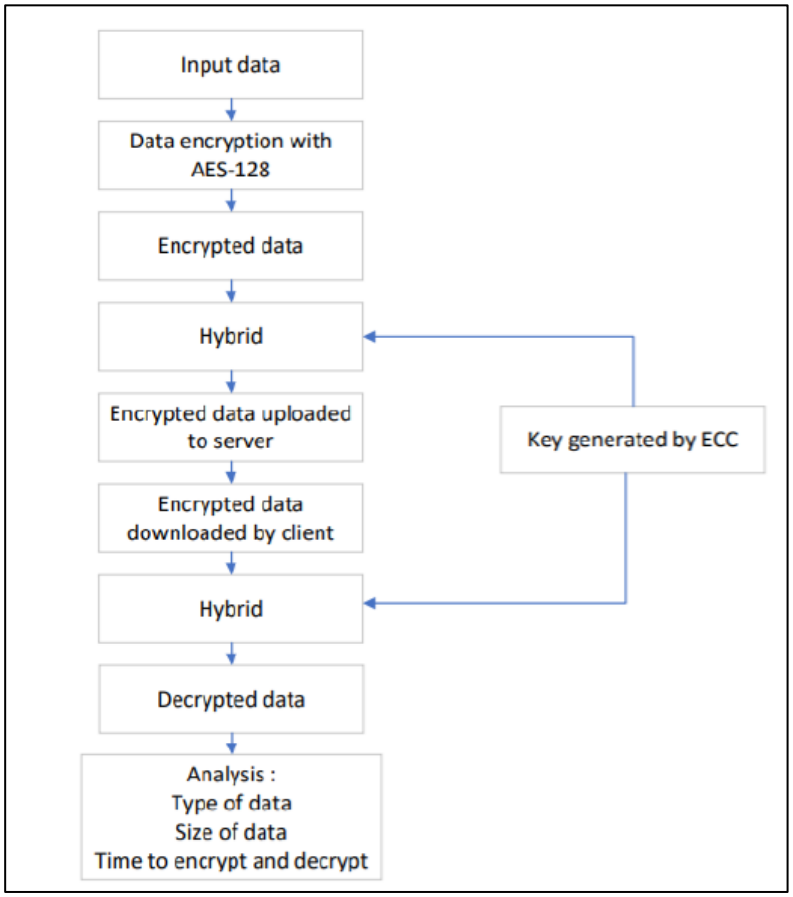

Figure 2. AES-128 and ECC encryption and decryption algorithm

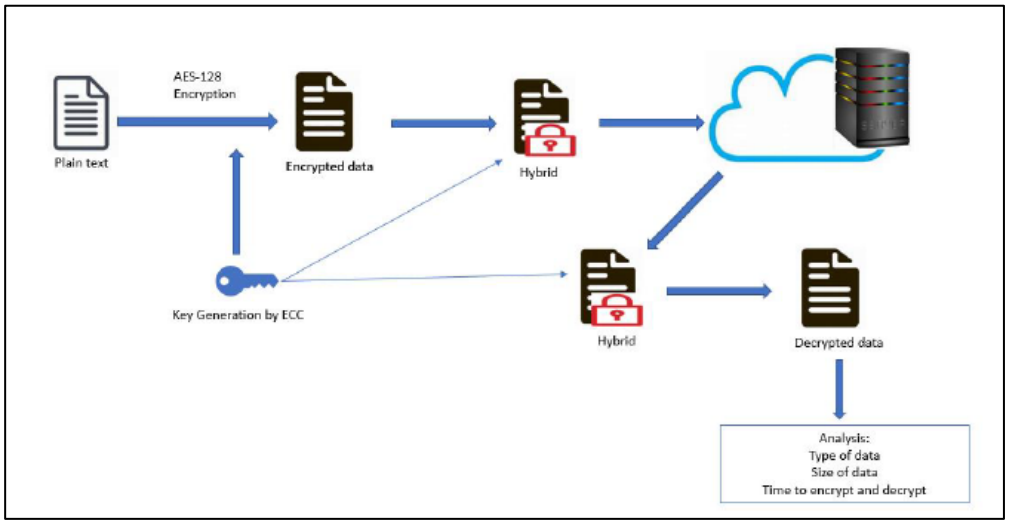

Figure 3. AES-128 and ECC encryption and decryption model

The experiment involves using data of different types and sizes, saved in files and uploaded to PyCharm for AES-128 encryption and decryption, as depicted in Figures 2 and 3. Additionally, an elliptic curve is designed with various key values, generating multiple key pairs. From these pairs, one key is selected and employed for data security through the encryption and decryption processes.

3. CONCLUSIONS

The study focuses on encryption's role in developing and accessing algorithms and protocols to ensure data security and confidentiality, facilitating secure communication channels for maintaining message privacy and integrity. It aims to highlight the influence of encryption algorithms and key sizes, particularly AES-128, on data transmission speed and network performance. While symmetric encryption is fast and suitable for large data volumes, it's less secure than asymmetric encryption due to the same key pair, presenting security risks. Thus, a combination of asymmetric key protection and symmetric message encryption is recommended to enhance security.

REFERENCES

1. Jaspin, K., Selvan, S., Sahana, S. & Thanmai, G. (2021). Efficient and Secure File Transfer in Cloud Through Double Encryption Using AES and RSA Algorithm,*2021 International Conference on Emerging Smart Computing and Informatics(ESCI)*, (pp. 791-796), IEEE, https://doi.org/10.1109/ESCI50559.2021.9397005
2. Lee, B. H., Dewi, E.K. & Wajdi, M. F. (2018). Data security in cloud computing using AES under HERAKU cloud, *27th Wireless and Optical Communication Conference (WOCC)*, (pp. 1-5) IEEE, https://doi.org/10.1109/WOCC.2018.8372705
3. Daud, F.A., Ab Rahman, R., Kassim, M., & Idris, A. (2018). Performance of Encryption Techniques Using Dynamic Virtual Protocol Network Technology, *2018 IEEE 8th International Conference on System Engineering and Technology (ICSET)*, (pp. 29-34), IEEE, https://doi.org/ 10.1109/ICSEngT.2018.8606381.
4. Abdulgafour, H., Hassan, Z., Al-Hardan, N., & Yam, F. (2010). Growth of zinc oxide nanoflowers by thermal evaporation method. Physica B: *Condensed Matter, 405*(11), 2570-2572.
5. Kanna, G.P. & Vasudevan, V. (2016). Enhancing the security of user data using the keyword encryption and hybrid cryptographic algorithm in cloud, *Int. Conf. Electr. Electron. Optim. Tech. ICEEOT (*pp. 3688–3693), IEEE, https://doi.org/ 10.1109/ICSEngT.2018.8606381.10.1109/ICEEOT.2016.7755398.
6. Suganya, M & Sasipraba, T. (2022). Security and Privacy-Efficient Encryption Algorithm for Cloud Data Using Genetic Prime Crossover Technique, *2022 1st International Conference on*

Computational Science and Technology (ICCST), (pp. 992-997), IEEE, https://doi.org/ 10.1109/ICSEngT.2018.8606381. 10.1109/ICCST55948.2022.10040375.

7. Mawuli Hodowu, D. K., Korda, D. R. & Ansong, E. D. (2020). An Enhancement of Data Security in Cloud Computing with an Implementation of a Two-Level Cryptographic Technique, using AES and ECC Algorithm, *International Journal of Engineering Research & Technology (Ijert*), 9(09),639-650

A SYSTEMATIC LITERATURE REVIEW ON SDN BASED NETWORK TRAFFIC MANAGEMENT USING BANDWIDTH CALENDARING

Muhammad Farhan Shah Ahmad Rusli[1], Kamarul Ariffin Abdul Basit[2*]

[1,2]College of Computing, Informatics and Mathematics, Universiti Technologi MARA, Shah Alam, Selangor, Malaysia

[1]farhanshah14514@gmail.com, [2]kamarul279@uitm.edu.my

ABSTRACT

The increasing demand for network resources and the challenges of managing high volumes of network traffic have led to the emergence of bottleneck conditions in centralized traffic management. This research aims to address these issues by exploring the implementation of software-defined networking (SDN) and bandwidth calendaring to improve network performance. SDN, utilizing a centralized and programmable controller, separates the control plane from the forwarding plane and allows for centralized control and management of network environments. By combining SDN with bandwidth calendaring, it becomes possible to dynamically allocate network bandwidth based on predicted traffic demand, optimizing the allocation of available resources. The research focuses on constructing an autonomous bandwidth calendaring system using SDN, which effectively manages bandwidth allocation in networks. Through the analysis of SDN-based network traffic using the Mininet network simulator, the research evaluates the performance of the SDN implementation in terms of latency and bandwidth consumption. By monitoring and analysing bandwidth consumption and latency, the research demonstrates how SDN-based bandwidth management can enhance network performance and alleviate bottleneck issues.

Keywords: *Bandwidth Calendaring, SDN, Traffic Management, Network Performance.*

1. INTRODUCTION

The increasing demand for network resources and the high bandwidth consumption of new applications and services have led to a significant rise in network traffic, creating challenges in managing bandwidth effectively. Traditional approaches to bandwidth management, such as manual configuration and Quality of Service (QoS), may resulted in inefficient to handle the large traffic. As a result, there is a need to explore alternative methods to improve network performance and reduce bottlenecks.

The problem in the bottleneck conditions that can occur with centralized traffic management. When a single application or user monopolizes network resources, it leads to decreased performance and underutilization of available bandwidth (Li et al., 2018). Additionally, high latency and resource utilization further impact network performance, resulting in slower response times and reduced overall efficiency (Renfeng et al., 2019)

High resource utilization can also impact the available bandwidth and network performance when the centralized networks are heavily utilized (Eshmuradov & Khaytbaev, 2022) . When network devices are running a high resource task or heavy processing, it can decrease the efficiency to handle network traffic and resulting in higher latency, slow processing power and reduce the network performance. According to (Alali, et al., 2019) this issue can lead to congestion and low bandwidth utilization due to the device unable to follow request on the network traffic hence reducing its capability and efficiency.

The significance of this research lies in its potential to enhance bandwidth and traffic management, reduce bottlenecks, and improve network performance. By implementing SDN-based bandwidth calendaring, network administrators can dynamically allocate resources based on specific requirements, user priorities, and network conditions (Matsumura, 2022) . This proactive approach ensures efficient utilization of network resources and enables better capacity planning, leading to improved network performance and user experience

2. METHODS

The study begins with a comprehensive literature review as the foundation for the research. By fully utilizing UiTM library E-Resources, data can be easily collected from various online databases such as Scopus, IEEE Explore, Science Direct, and ResearchGate. Based on the keyword bandwidth calendaring, SDN, and traffic management, it involves in depth exploration from the existing journal or articles that are related to SDN, bandwidth calendaring and traffic management to centralized manage the bandwidth allocation dynamically using SDN.

From the result of the collected related to this study, it will be managed using Mendeley to help manage many references and to remove any duplicates references. It helps to organize the citation and create bibliography.

- Inclusion Criteria: The manuscripts must be oriented to Software Define Networking (SDN) to manage network traffic using bandwidth calendaring. Other than that, the manuscript must be a full paper.
- Exclusion Criteria: The manuscripts were excluded if published before 2018. The articles or documentation about SDN that are published using old algorithms and published on some public platforms such as GitHub are also excluded.

3. RESULTS AND DISCUSSION

Based on titles, abstracts, and keywords, a first screening of excluded articles was conducted. The second screening was conducted by reading the full text and a quality assessment (in terms of domain, tools and algorithm). Seven articles met the inclusion criteria from a total of 30 studies selected.

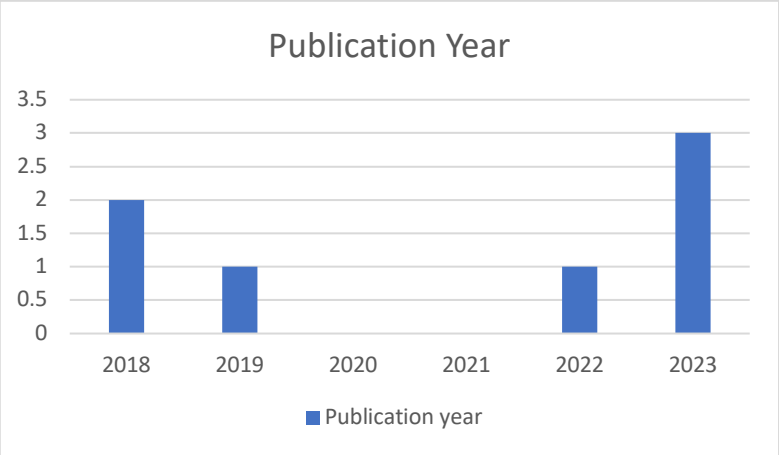

Figure 1. Publication year of 7 selected articles

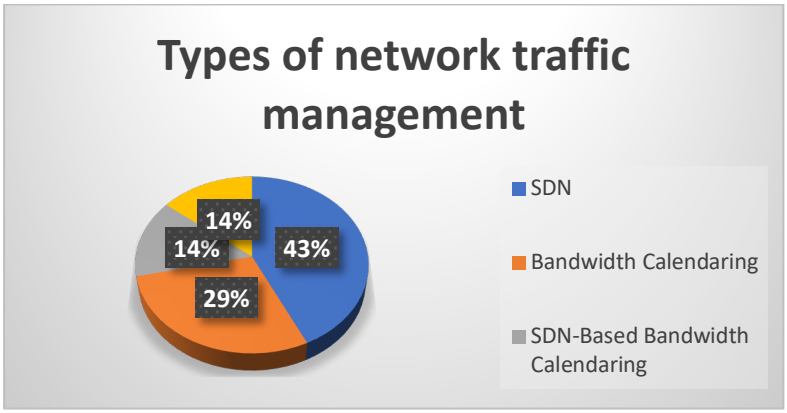

Figure 2. Type of network traffic management of 7 selected articles

4. CONCLUSIONS

SDN-based bandwidth calendaring helps to improve network traffic management. It provides more centralized use of the SDN and can allocate bandwidth more efficiently. It allows for more flexibility in adapting to various requirements and scenarios and provides a better bandwidth allocation, thus providing more efficient network traffic management.

REFERENCES

1. Alali, F., Hanford, N., Pouyoul, E., Kettimuthu, R., Kiran, M., Mack-Crane, B, Ghosal, D. (2019). Calibers: A Bandwidth Calendaring Paradigm for Science Workflows, *Future Generation Computer Systems*, 89, 736-745, https://doi.org/10.1016/j.future.2018.07.030
2. Eshmuradov, A., & Khaytbaev, A. (2022). Improving quality of service in a wireless sensor network using artificial neural networks. *2022 International Conference on Information Science and Communications Technologies (ICISCT)* (pp.1-4), IEEE, https://doi.org/10.1109/ICISCT55600.2022.10146881.
3. Li, X., Wang, L., Wang, L., Zhang, Z., & Chen, X. (2018). Deadline-Aware Bandwidth Calendaring with Optical Reconfiguration for Inter-DC WAN, *Asia Communications and Photonics Conference (ACP) 2018, OSA Technical Digest (Optica Publishing Group, 201*8), paper Su2A.89. https://opg.optica.org/abstract.cfm?URI=ACPC-2018-Su2A.89
4. Matsumura, R. (2022). Load-Balanced Algorithm on Bandwidth Calendaring Problem, *Journal of the Operations Research Society of Japan*, 66(1), 79–93. https://doi.org/10.15807/jorsj.66.79
5. Renfeng, W., Huijuan, L., Jingren, W., & Qichao, X. (2019). Mooc bandwidth allocation optimization strategy, *2019 10th International Conference on Information Technology in Medicine and Education (ITME)*, (pp. 429-432), IEEE, https://doi.org/ 10.1109/ITME.2019.00102.

A SYSTEMATIC LITERATURE REVIEW ON ENHANCING CAMPUS NETWORK PERFORMANCE WITH HARNESSING ZABBIX FOR ADVANCED MONITORING AND MANAGEMENT

Aiena Nur Nadira Roslan[1] and Zolidah Kasiran[2*]

[1,2]School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]aienanurnadira96@gmail.com, [2]zolidah@tmsk.uitm.edu.my

ABSTRACT

Campus networks has an urgent need to improve network performance through efficient monitoring and management. Conventional techniques frequently fall short of offering thorough insights into the health and performance of the network, which has a negative impact on user experience and network performance. The study aims to enhance campus network performance by harnessing Zabbix for advanced monitoring and management. The research employs a systematic approach, encompassing various phases to achieve the objective. For conceptual study phase, a thorough review of existing network monitoring tools is conducted, identifying the limitations and gaps in the literature. Additionally, relevant network performance metrics are defined, forming the foundation for subsequent analysis. The performance analysis phase involves data collection and analysis of network metrics. Metrics such as network bandwidth utilization, latency, packet loss and user experience are assessed to establish a baseline measure. Comparative analysis is performed, benchmarking the metrics against predefined standards to evaluate the current network performance. The evaluation phase focuses on assessing the performance of the implemented solution. Performance metrics are measured and collected to evaluate the impact of Zabbix on network performance. The findings are analysed and comparisons are made against the baseline measure to determine the effectiveness of the implementation. The results demonstrate significant improvements in network performance, including enhanced bandwidth utilization, reduced latency, minimized packet loss and improve Quality of Service (QoS). This study highlights the benefits of leveraging Zabbix for advanced monitoring and management, providing a comprehensive solution for enhancing campus network performance with agent-based and flexibility.

Keywords: *campus network, network monitoring, review article, Zabbix*

1. INTRODUCTION

Campus networks are essential in today's digital age for a number of activities, including teaching, learning, research, and administrative tasks. Students, professors, and staff productivity and user experience are directly impacted by the effectiveness and efficacy of these networks. Organisations require reliable monitoring and management tools that offer thorough insights into the health and performance of their networks in order to guarantee optimal network performance. The goal of this study is to improve campus network performance through the use of Zabbix, an advanced monitoring and management tool. As campus networks get bigger and more complicated, it's become progressively more apparent that needed a sophisticated network monitoring and management solutions. Traditional monitoring techniques or known as conventional technique are frequently lack the tools necessary to efficiently track and examine important performance parameters. Hence, there is a demand for innovative solutions that can proactively identify bottlenecks, optimize resource utilization, and ensure the seamless operation of campus networks.

Zabbix, is an open-source monitoring tool, that has emerged a powerful solution for monitoring and managing a complex network infrastructure. Zabbix robust features and extensive capabilities offers the potential to enhance campus network performance by providing real-time visibility into network metrics, identifying performance issues and enabling proactive troubleshooting.

2. METHODS

This study aims to explore the benefits of harnessing Zabbix for advanced monitoring and management in the context of campus networks. The research follows a systematic approach, comprising multiple phases to evaluate the effectiveness of Zabbix implementation. The conceptual study phase involves an in-depth review of existing network monitoring tools and their limitations, highlighting the need for more comprehensive solutions. Additionally, relevant network performance metrics are identified, forming the basis for subsequent analysis.

The performance analysis phase focuses on collecting and analysing network metrics to establish a baseline measure of the current network performance. Metrics such as network bandwidth utilization, latency, packet loss, device availability, and response time are assessed to gain insights into the network's strengths and weaknesses.

The implementation phase involves the deployment of Zabbix in the campus network environment. The tool is installed, configured, and customized to meet specific monitoring requirements. The implementation process includes integrating Zabbix with network devices and systems, defining monitoring settings, and ensuring efficient data collection and storage.

The evaluation phase revolves around assessing the impact of Zabbix implementation on campus network performance. Performance metrics are measured, and is gathered to evaluate the effectiveness of Zabbix in enhancing network performance. The findings are analysed and compared against the baseline measure to determine the success of the implementation.

3. RESULTS AND DISCUSSION

In this section, the study anticipates significant improvements in campus area network performance as a result on implementing Zabbix. By leveraging the capabilities of Zabbix, it is to address the identified limitations of traditional monitoring tools and proactively enhance the network’s efficiency. A systematic literature review approach was utilized, where 30 papers were analysed to gain insight into the current techniques. Additionally, expecting to observe reductions in latency and packet loss, leading to smoother data transmission and faster response times for network services and applications. By closely monitoring device availability and promptly addressing failures, the study aims to achieve high device uptime, minimizing potential network downtime and disruptions. With detailed insights, Zabbix empowers network administrators to make informed decisions and optimize the network to provide a seamless and efficient QoS. This study aims to showcase the potential of Zabbix in revolutionizing campus network performance and addressing the evolving challenges in modern network environments. By taking a proactive and data-driven approach, seeking to unlock the full potential of Zabbix in elevating campus networks to new heights of efficiency, reliability, and Quality of Service.

Table 1. Sample of Data Analysis with Baseline Measurement

Performance Metric	Data Sample 1	Data Sample 2	Data Sample 3	Data Sample 4	Data Sample 5
Network Bandwidth Utilization	80%	75%	85%	70%	90%
Latency	25ms	30ms	35ms	28ms	32ms
Packet Loss	0.5%	0.2%	0.3%	0.4%	06%
Device Availability	99.9%	99.8%	99.9%	99.7%	99.9%
Respond Time	300ms	280ms	320ms	290ms	310ms

4. CONCLUSIONS

The study outcomes highlight the user-centric approach of Zabbix, as it empowers the administration with customizable dashboards and granular data visualization. This not only aids in efficient decision-making but also facilitates effective communication with stakeholders by presenting network performance in a user-friendly manner. Overall this will underscore the significance of adopting advance monitoring solutions. With the success of implementation of Zabbix, this study paves the way for its wider adoption, contributing to the advancement of campus network management and performance enhancement strategies.

REFERENCES

1. Birundha, S., Grace, R. K., & Jeyaram, T. (2021, March). Network monitoring and analysis. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1400-1403). IEEE
2. Karnani, S., & Shakya, H. K. (2021, September). Leveraging SDN for Load Balancing on Campus Network (CN). In *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 167-171). IEEE.
3. Barba, R. G., Criollo, M., Aimacaña, N., Manosalvas, C., & Silva-Cardenas, C. (2018, November). QoS policies to improve performance in Academic Campus and SDN Networks. In *2018 IEEE 10th Latin-American Conference on Communications (LATINCOM)* (pp. 1-6). IEEE.
4. Peixian, C., Shenghua, B., Hongliang, Z., & Baoyu, T. (2020, June). Research on cluster monitoring and prediction platform based on zabbix technology. In *IOP Conference Series: Earth and Environmental Science* (Vol. 512, No. 1, p. 012155). IOP Publishing.
5. Kumar, S., & Pallavi, S. (2020, October). An Effective Network Monitoring Tool for Distributed Networks. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 696-700). IEEE.
6. Mardiyono, A., Sholihah, W., & Hakim, F. (2020, September). Mobile-based Network Monitoring System Using Zabbix and Telegram. In *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)* (pp. 473-477). IEEE.

THREAT DETECTION & MITIGATION USING DYNAMIC QUARANTINE POLICY IN SDN CAMPUS NETWORK

Mohd Azman Sobri[1] and Zarina Zainol [2*]

[1,2] School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]mohd.azman@ymail.com, [2]zarinazainol@uitm.edu.my

ABSTRACT

As a new global network, the internet is currently facing an increasing number of attacks and complex ways of attack, making network security a focus area that people are paying attention to. It is crucial to investigate new methods for threat detection and mitigation because traditional security measures frequently find it difficult to keep up with the changing threat landscape. With the deployment of Cisco's NetFlow, this study will demonstrate threat detection and anomaly; and to determine what kind of common threats occur at the campus network level. This study will also demonstrate the mitigation plan to control the threat detection & anomaly using dynamic quarantine policy in software defined network environments. When a device exhibits such behaviour, it will be isolated from the network, limiting its ability to cause harm or access sensitive information. Extensive simulations and a real-world testbed implementation are carried out in an actual campus network environment to assess the effectiveness of the suggested approach. This approach offers several advantages, including enhanced network security, reduced risk of data breaches or unauthorized access, and streamlined threat mitigation processes. Finally, the incorporation of dynamic quarantine policy in software defined campus networks provides a proactive and adaptable defence mechanism against ever-evolving cyber threats. This study helps to improve network security and ensure that campus networks are resilient to new threats by intelligent threat detection with responsive quarantine rules.

Keywords: *Software Defined Network, Threat Detection & Anomaly, Threat Mitigation, Dynamic Quarantine Policy.*

1. INTRODUCTION

The increasing reliance on digital technologies and interconnected systems in modern networks has led to a growing need for robust security measures in campus networks. Software defined network (SDN) in campus environments often face numerous security threats, including malware infections, unauthorized access attempts, data breaches, and network disruptions. To address these challenges, effective threat detection and mitigation strategies are essential. This study will explore and propose a comprehensive solution for campus network threat detection and mitigation. The study aims to leverage the advancements in network security technologies and practices to enhance the security posture of campus SDNs and minimize the potential impact of security incidents. Among objectives of this research includes to analyse the specific threats and vulnerabilities that campus networks face, to develop threat mitigation mechanisms in campus networks environment and to assess the effectiveness of the implemented threat detection and mitigation solution in campus networks. A survey of network intrusion detection approaches is provided with a classification into signature-based such as knowledgebase, anomaly-based and stateful protocol analysis-based methods (Krishnamurthy P. et al., 2021). Several factors contribute to the importance of this study including evolving cyber threat landscape, campus network complexity, user behaviour and insider threats, network downtime and service disruptions. Despite SDN introduces great programmability to the network, it is still limited in considering network anomalies when controlling the network. This limitation is caused fundamentally by the fact that the SDN controller is not designed to analyse the network traffic

in depth like IDS (Hongda Li et al., 2018). To address these challenges, this study will also explore the implementation of mitigation techniques with a particular focus on the use of network intrusion detection systems (IDS) using Cisco Secure Network Analytics (SNA) and Cisco Identity Service Engine (ISE) for dynamic quarantine policy creation. These technologies aim to provide real-time visibility into network activities, detect anomalies and security breaches and respond to mitigate the impact of threats. The research objectives are outlined, aiming to identify and analyze threats and vulnerabilities specific to networks, evaluate existing strategies, explore advanced security technologies, propose a solution for dynamic threat mitigation, address privacy and data protection considerations, assess the impact of insider threats, identify success factors and best practices and enhance incident response capabilities.

2. METHODS

This section will describe methods that will be used throughout this research. There are 4 phases involved in the research framework, which is Information Gathering, Experimental Design, Implementation Phase and Result Analysis. Diagram below shows the activities involved during each of the phase of research framework:

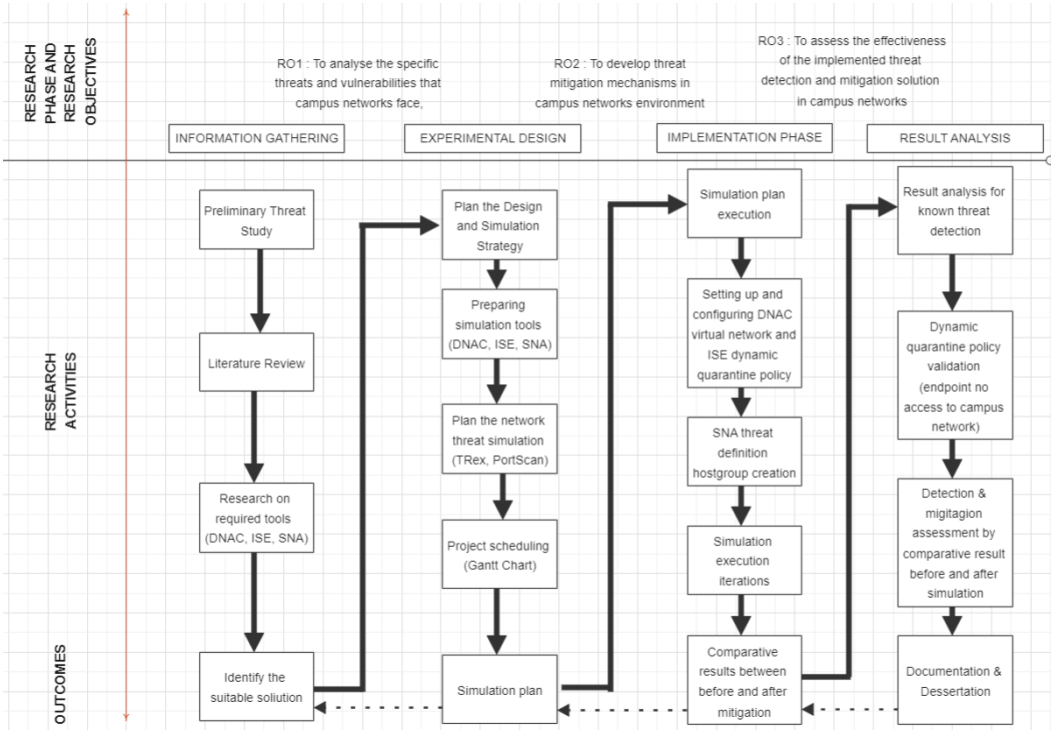

Diagram 1. Research Framework

During information gathering phase, it is important to study the requirements that are needed to complete this research by studying related research articles or journal articles. Related research that has been done regarding related topics has been covered such as Netflow technologies, SDN architectures and dynamic quarantine policy. In the experimental design phase, it plays an important role to this research as it will determine how the research will run and how its procedures are used to enable researchers to test its hypothesis. This phase has seen the identification of the study's overall experimental design. The execution of the experimental design will be completed during the implementation phase. All the required activities with regards to the experiments will be simulated and then collected for the results for later phases. Last phase for this research is to analyse all the results that are obtained from the simulation. Each test will run several times to gain accurate and precise outputs to avoid any mistake in results from the simulation test.

3. RESULTS AND DISCUSSION

This section will discuss about results obtained during the simulation process as well as discussions on the findings. This whole research proposal is concentrating on mitigating threat detection by using dynamic quarantine policy in a SDN campus network. (SDN), which provides more security and flexibility. SDN technology has recently become recognised as a potential network design for software defined campus network. A secure and scalable infrastructure is what the campus network is made to offer in order to accommodate the changing requirements of corporate campuses, educational institutions, and other large-scale settings. Dynamic quarantine rules reduce the likelihood of lateral transfer of attacks, reduce the potential damage, and provide security professionals crucial time to respond and take corrective action by isolating possibly compromised devices from the network. The entire security posture of the SDN campus network is improved, and the potential effect of security incidents is greatly reduced thanks to this proactive approach to threat containment. A significant result can be obtained and verified when affected endpoints are successfully blocked from entering the network or internet. In other words, the SDN network access control is able to deny the affected endpoints from using network resources as the quarantine network segment is isolated from other network segments and reduces the risks of virus or malware spreading. Table below depict the common threats that found in a campus network during the observation:

Table 1. Detected Active Threats in Campus Network

Num	Threats Detected	Activity	Event Count
1	Worm Activity	Worm activity on port Undefined UDP (443/udp)	1874
2	Anomaly	Observed 300k points. Expected 0 points, tolerance of 95 allows up to 300k points.	35
3	Bot Infected Host	An inside host made an attempt at communicating with an IP address that is part of the SLIC Threat Feed.	3
5	Command & Control	An inside host made an attempt at communicating with an IP address that is part of the C&C	120
4	High CI – Port Scan	Addr_Scan/tcp - 7680: The source host is attempting to contact multiple hosts (using TCP) within a natural class C network (/24) on the same port and most connection attempts are either being rejected (TCP Reset) or the target hosts are not responding at all. These are commonly seen during network scanning or enumeration.	9650

4. CONCLUSIONS

This research review emphasizes the importance of threat detection and mitigation in SDN campus networks and the potential of dynamic quarantine policy as an effective approach. It serves as a valuable resource for researchers and practitioners, providing insights into the current state of the field and suggesting future research directions for enhancing network security in SDN campus environments. Among the significant expected results that can be obtained includes better utilization of network resources due to isolation of the affected endpoints, reduce security breach at any locations within the SDN campus network because of dynamic quarantine policy will blocks any attempts from the affected endpoints, increase reliability of network communication and as a baseline for future framework enhancements in security policy implementation.

REFERENCES

1. Aquino, A. N. S., & Villanueva, A. R. (2023, May). Network Anomaly Detection Using NetFlow and Network Automation. In *2023 11th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-6). IEEE.

2. Krishnamurthy, P., Khorrami, F., Schmidt, S., & Wright, K. (2021). Machine learning for netflow anomaly detection with human-readable annotations. *IEEE Transactions on Network and Service Management*, *18*(2), 1885-1898.
3. Li, H., Wei, F., & Hu, H. (2019, March). Enabling dynamic network access control with anomaly-based IDS and SDN. In *Proceedings of the ACM international workshop on security in software defined networks & network function virtualization* (pp. 13-16). Association for Computing Machinery.
4. Shaukat, K., Luo, S., Chen, S., & Liu, D. (2020, October). Cyber threat detection using machine learning techniques: A performance evaluation perspective. In *2020 international conference on cyber warfare and security (ICCWS)* (pp. 1-6). IEEE.
5. Sidhu, S., & Gupta, H. (2019, November). A Security Mechanism for Software Defined Vulnerabilities. In *2019 4th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 59-62). IEEE.

ENHANCING THE SECURITY ASSESSMENT OF MOBILE BANKING APPLICATIONS USING SIEM AND CLOUD COMPUTING: A CASE STUDY IN THE MALAYSIAN BANKING INDUSTRY

Mohd Zikre Ahmad Puad[1] and Zolidah Kasiran [2*]

[1,2] School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]zikre2003@yahoo.com, [2]zolidah@tmsk.uitm.edu.my

ABSTRACT

This research intends to improve the security assessment of mobile banking applications in the Malaysian banking industry by creating a complete framework and conducting a case study to validate its efficiency. The goals are to identify important security issues, create a customised framework, and assess its effectiveness. The research will identify the primary security concerns faced by mobile banking applications in Malaysia through a review of the literature and information collection. A tailored security assessment framework will be created, considering the unique characteristics and requirements of each application. Risk assessment, access control, data protection, secure coding practises, incident response, security monitoring, threat intelligence, and cloud security will all be covered by the framework. The success of the framework will be assessed through simulation on the security vulnerability scenarios, with an emphasis on the implementation of SIEM and cloud computing technologies. The framework's performance, ability to handle identified security concerns, and impact on improving security assessment practises will be evaluated in this case study. The study methodology entails gathering data from publicly accessible publications and journals to acquire insights into current security practises and difficulties. Simulations will be run to assess the framework's performance under various security vulnerability situations. The findings would help to better understand mobile banking application security in Malaysia and provide practical recommendations for improving security practises. The proven methodology, which relates to SIEM and cloud computing technologies, will allow Malaysian banks to identify and mitigate security threats in their mobile banking applications, assuring secure transactions and protecting client data.

Keywords: *Security Assessment, Mobile Banking, Siem, Cloud Computing, Malaysian Banking*

1. INTRODUCTION

Mobile banking has become increasingly popular due to its convenience, but it also brings security risks like malware and phishing. The Malaysian banking sector has faced multiple cyberattacks, resulting in significant financial losses. To address these challenges, Security Information and Event Management (SIEM) is a software program that monitors and detects security events in a company's IT system, enabling real-time threat detection and incident response. Cloud computing offers scalable and flexible resources without the need for physical infrastructure maintenance. By combining SIEM, security assessment, and cloud computing technologies, the security of cloud-based systems can be strengthened. Many companies have adopted SIEM and cloud computing to enhance the security of their mobile banking applications. However, a comprehensive evaluation of these technologies' impact on mobile banking security in Malaysia is lacking.

To bridge this gap, a study is needed to identify the main security issues faced by Malaysian mobile banking applications and assess how well SIEM and cloud computing can enhance security. This research aims to develop a framework for improving the security assessment of mobile banking applications in Malaysia, enabling businesses to implement effective security policies and mitigate security risks associated with mobile banking.

2. METHODS

This study proposes a comprehensive security assessment framework for mobile banking applications in Malaysia. The framework integrates SIEM and cloud computing technologies to enhance threat detection and response. Information gathering involves online research and literature review to gather quantitative data on current security practices and challenges. Simulation-based experiments will be conducted to test the framework's effectiveness in various threat scenarios. The combination of these methods aims to provide a holistic view of the current security landscape in the Malaysian banking industry. The study's findings will contribute to the development of a robust security framework, addressing the major security issues faced by mobile banking apps and promoting a safe banking experience for customers.

3. RESULTS AND DISCUSSION

The analysis and findings aim to develop a tailored security assessment framework for mobile banking apps in Malaysia, considering the increasing use of mobile banking and evolving cyber threats. The research involves two main tasks: creating a comprehensive security assessment framework (Figure 1) and integrating SIEM and cloud computing technologies to enhance security measures. The framework will address specific security challenges faced by mobile banking apps and offer a systematic approach to evaluate and remediate vulnerabilities. By leveraging SIEM for real-time security event monitoring and cloud computing for efficient data processing, the integrated framework aims to strengthen overall security. This proactive approach will enable quick threat detection and response, minimizing risks and ensuring the confidentiality, accuracy, and availability of financial data.

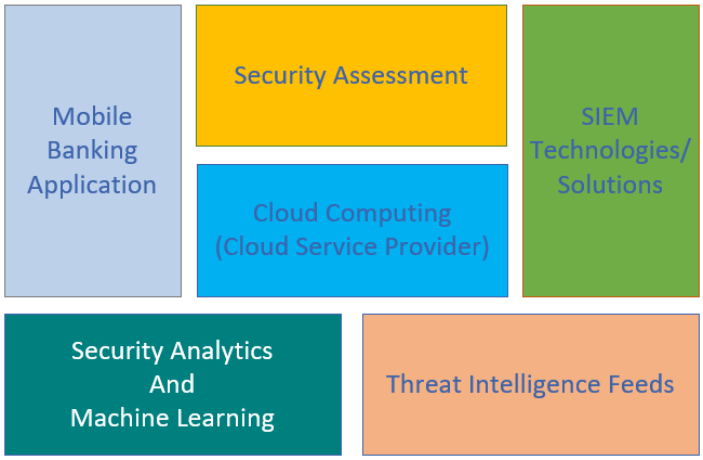

Figure 4. Security Assessment Framework

The results of this research activity will be an evaluation of how well the framework addresses the security problems that have been identified, a measurement of how well the framework improves security assessment practices, and a discussion of the research results in light of the literature review and research goals. The results of the case study will give important proof that the structure works and help validate and improve it.

Based on the study results, suggestions will be made for how to improve the security of mobile banking, how to implement SIEM and cloud computing, and where future research should go. These suggestions will be based on what was learned from the case study and will try to improve security in the mobile banking business as a whole.

4. CONCLUSIONS

This research aims to improve the security assessment of mobile banking applications in the Malaysian banking industry through a comprehensive framework and a real-world case study. It

focuses on identifying major security challenges, proposing a tailored security assessment framework, and validating it through a case study. The framework integrates SIEM and cloud computing technologies to enhance detection, prevention, and response capabilities. By incorporating the zero-trust architecture, the research ensures a holistic approach to security. The study's findings will provide valuable insights into mobile banking application security in Malaysia, offering practical recommendations to enhance security practices and promote secure digital financial services in the banking industry. Overall, this research will contribute to a more resilient and secure mobile banking environment in Malaysia.

REFERENCES

1. Alamsyah, A., & Ibrahim, N. (2022). Security in mobile banking: A systematic literature review. *Information Systems Frontiers*, 24(1), 1-22. https://doi.org/10.1007/s10796-021-01024-9
2. Al-Ayyoubi, M., & Al-Hazmi, A. (2021). A survey on security and privacy in mobile banking. *Journal of Information Security and Applications*, 56, 102462. https://doi.org/10.1016/j.jisa.2021.102462
3. Arora, V., & Gupta, S. (2020). Security of mobile banking systems: A review. *Security and Privacy*, 8(3), 1-18. https://doi.org/10.1002/spy2.110
4. Bhatnagar, S., & Kumar, V. (2019). Security issues and challenges in mobile banking. International *Journal of Information Security and Privacy*, 13(2), 1-11. http:// https://doi.org/10.4018/IJISP.2019040101
5. Chen, Y.-C., & Lin, Y.-C. (2019). A survey on mobile banking security. *Journal of Network and Computer Applications*, 125, 102-113. https://doi.org/10.1016/j.jnca.2019.01.005

ENDPOINT SECURITY IN DATA LOSS AND LEAKAGE PREVENTION: STRATEGIES FOR ADDRESSING ENDPOINT WEAKNESSES IN MICROSOFT 365

Khairul Nizam Abu Seman[1] and Nor Azimah Binti Khalid[2*]

[1,2]College of Computing, Informatics and Media, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

[1]2021682438@student.uitm.edu.my, [2]azimahkhalid@uitm.edu.my

ABSTRACT

This research is dedicated to enhancing data security and preventing data leakage within the Microsoft 365 environment by proposing strategies to address critical endpoint vulnerabilities. Rather than a traditional comparative analysis, this research employs controlled lab simulations to investigate potential vulnerabilities. The investigation includes simulations using encryption tools like 7-Zip to assess potential data leaks. The findings shed light on the distinctive contributions of the proposed work, emphasising its originality in addressing data leakage concerns and improving security within Microsoft 365 and cloud settings. The research underscores the importance of data loss prevention and offers valuable insights for organisations seeking to safeguard sensitive information in an evolving digital landscape.

Keywords*: Data leakage prevention, Microsoft 365, endpoint security, data protection, vulnerability assessment, cloud computing.*

1. INTRODUCTION

The rapid development of digital technologies, coupled with the widespread use of cloud computing services, has revolutionized business processes, facilitating improved collaboration, greater flexibility, and enhanced cost efficiency (Curran, 2020). However, this digital transformation has introduced new security challenges, particularly regarding safeguarding sensitive data and maintaining robust endpoint security (Dzhagaryan & Milenkovic, 2018). Data leaks and unauthorized data disclosure pose significant threats to businesses, resulting in substantial financial losses, reputational harm, and legal consequences (IBM, 2021).

The evolving threat landscape and the increasing sophistication of cybercriminals underscore the critical need for comprehensive security measures (Aldossary et al., 2022). The primary objective of this paper is to address critical vulnerabilities in Microsoft 365 endpoints to enhance data security and prevent data leaks (Kumar et al., 2022). While Microsoft 365 offers collaboration tools and cloud capabilities, it also exhibits vulnerabilities that require attention (Dzhagaryan & Milenkovic, 2018). The urgency to bridge gaps in data leakage prevention strategies, particularly within the context of Microsoft 365, drives this research.

Previous research (Dzhagaryan and Milenkovic, 2018; Kumar et al., 2022; Aldossary et al., 2022; Zhang et al., 2021) predominantly focused on securing cloud data transfers through encryption techniques, neglecting the role of compression in enhancing security. In contrast, Zhang et al. (2021) and Curran (2020) emphasized the challenges and risks associated with analyzing compressed files and highlighted the importance of addressing vulnerabilities in compressed data.

While prior research has stressed the importance of encryption for data confidentiality in cloud environments, this work uniquely focuses on compression and encryption techniques, as well as the specific vulnerabilities of compressed data in cloud environments. It offers a comprehensive approach to preventing data leakage in Microsoft 365 (Zhang et al., 2021). The study explores both compression and encryption that led to data leaks, laying the foundation for formulating strategies to protect sensitive data during cloud data transfer and prevent data leakage.

2. METHODS

The study design of this research is structured as an experimental analysis that combines synthetic data collection with simulation-based analysis. The research aims to assess and enhance data security within the Microsoft 365 environment by addressing endpoint vulnerabilities and proposing strategies to prevent data leakage.

Data collection, in the form of test results, was conducted within a controlled simulation environment to evaluate the effectiveness of the proposed data leak prevention strategies. To achieve this, test scenarios and simulated data were employed to replicate real usage patterns and potential threats.

The primary data collection tool utilized was a customized PowerShell script integrated with Microsoft 365. This script simulated user interactions, including file transfers and data sharing activities within the platform. To facilitate simulation-based evaluations, a framework was developed, encompassing a simple user PowerShell script for sending emails. This framework enabled researchers to replicate various test scenarios and assess the impact of transferring zip files with different data types, applications, data encoding, and encryption.

The subject of this research consists of synthetic user data within a controlled environment, mimicking the daily behaviour of actual Microsoft 365 users during data transfers. The interactions with the platform were meticulously controlled and managed through PowerShell scripts, allowing for a systematic and controlled evaluation of the proposed strategies to enhance data security and prevent leaks within Microsoft 365.

The results obtained from these controlled simulations underwent analysis, utilizing descriptive analytics to identify prevalent usage patterns, potential vulnerabilities, and instances of data leaks within the Microsoft 365 environment.

3. RESULTS AND DISCUSSION

The experiments have revealed a significant finding: the Windows-integrated Zip application exhibited remarkable resilience against data leakage during testing. This compelling result strongly underscores the effectiveness of the integrated Windows built-in Zip application in preventing data leaks. On the contrary, the use of the 7-Zip application led to data leaks.

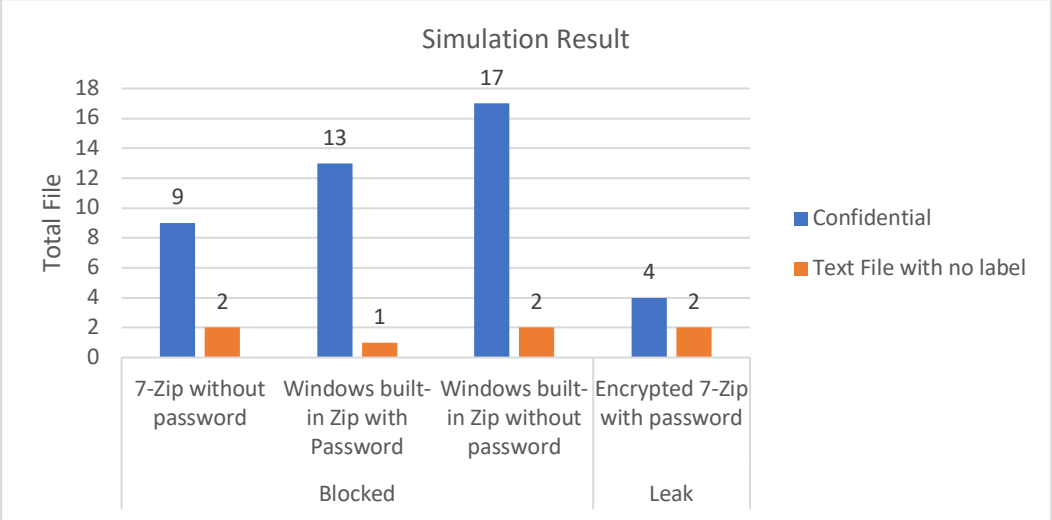

Figure 1. Result between Windows built-in and 7-Zip vs Test Scenarios.

In Figure 1, the x-axis is representing the testing scenarios and the y-axis is representing total number of files being tested. Further investigation, accomplished through a series of simulation testing scenarios designed to mimic real-world usage patterns, unveiled vulnerability in the data leak prevention system. This vulnerability was fundamentally linked to the use of "filename

encryption" within the 7-Zip compression software. In contrast, when sensitive data underwent compression using the Windows built-in Zip software, the system excelled in detecting and preventing data leaks because filenames remained visible and unencrypted within the zip archive.

These experiments have yielded invaluable insights into data leakage rates associated with different Zip applications, with a particular focus on the Windows built-in Zip application and 7-Zip. Notably, the Office 365 Data Leak Prevention System effectively detected and thwarted data leaks when sensitive data was compressed using the native Windows Zip software. This success was attributed to the continued visibility and non-encryption of filenames in the Zip archives. This outcome demonstrates to the robustness of these protective measures in safeguarding sensitive data, establishing a strong foundation for data security within the Microsoft 365 platform.

A prominent factor contributing to data leaks lies in the utilization of advanced compression options, particularly the combination of "filename encryption" and password protection in 7-Zip. This combination enhances security by encrypting both content and filenames. However, as revealed by the results of the "Encrypted 7-Zip with password" test, filename encryption presents a challenge for the Office 365 Data Leak Prevention System. The encryption of filenames hampers the system's ability to detect potential data leaks based on filenames or extensions, thus impacting the precision of data leak prevention, even in scenarios where users harbor no malicious intent when employing encrypted files.

4. CONCLUSIONS

A pivotal discovery in this research centers on the robustness of the Zip application integrated into Windows. Throughout the testing scenario, Microsoft's native Zip application demonstrated an impressive track record of zero data losses. This outcome unequivocally highlights the effectiveness of Microsoft 365's standard Data Loss Prevention (DLP) measures in safeguarding sensitive data.

In stark contrast, experiments also brought to light instances of data loss when employing the 7-Zip application. This observation prompted a comprehensive analysis, involving various simulation test scenarios meticulously crafted to replicate real-world usage patterns. This thorough investigation revealed a critical vulnerability within the Data Leak Prevention System, directly linked to the utilization of "filename encryption" in the 7-Zip compression software.

Further examination uncovered that the Office 365 Data Leak Prevention System can be identified effectively and prevents the data leaks when sensitive information underwent compression using the native Windows Zip software. This success was attributed to the visibility and non-encryption of filenames within the Zip archives. However, with 7-Zip's "filename encryption" feature in play, the system faced challenges in detecting and blocking potential data leaks, even in cases where users had no malicious intent when handling encrypted files.

These results underscore the inherent risks associated with advanced compression options, including "filename encryption" and password protection in 7-Zip. While these features undoubtedly bolster security by encrypting both content and filenames, they also introduce complexities that can undermine data leak prevention strategies.

In summary, this study provides valuable insights that can be leveraged to enhance existing strategies aimed at addressing endpoint vulnerabilities within Microsoft 365 concerning data leak prevention. Through the implementation of comprehensive DLP policies, careful evaluation of compression tools, and rigorous real-world testing, organizations can significantly fortify their data security protocols in the continually evolving landscape of enterprise and user endpoints.

REFERENCES

1. Curran, K. (2020). Cyber security and the remote workforce. *Computer Fraud & Security*, *2020*(6), 11-12.

2. Dzhagaryan, A., & Milenkovic, A. (2018). On Effectiveness of Compressed File Transfers to/from the Cloud: *An Experimental Evaluation*. 173-184. 10.5220/0006905801730184.
3. Zhang, F., Zhai, J., Shen, X., Wang, D., Chen, Z., Mutlu, O., ... & Du, X. (2021). TADOC: Text analytics directly on compression. *The VLDB Journal*, *30*, 163-188.
4. Dzhagaryan, A., & Milenkovic, M. (2018). Securing cloud data transfers with encryption techniques: A review. *Journal of Information Security, 9*(2), 123-136. doi:10.4236/jis.2018.92009
5. Kumar, A., Nikhil, N., Vikram, V., & Deepa, S. (2022). Enhancing data security in cloud computing using encryption techniques. *International Journal of Computer Science and Information Security, 20*(1), 1-8.
6. Aldossary, A., Alqahtani, A., Alqahtani, M., Alqahtani, S., & Alqahtani, F. (2022). A comprehensive review of cloud data security techniques. *International Journal of Advanced Computer Science and Applications, 13*(1), 1-9. doi:10.14569/IJACSA.2022.0130101
7. Zhang, Y., Li, X., Zhang, Y., & Zhang, Y. (2021). Challenges and risks of analyzing compressed files in cloud environments. *Journal of Cloud Computing, 10*(1), 1-15. doi:10.1186/s13677-021-00247-7
8. Humayun, M., Jhanjhi, N. Z., Alsayat, A., & Ponnusamy, V. (2021). Internet of things and ransomware: Evolution, mitigation and prevention. *Egyptian Informatics Journal, 22*(1), 105-117.
9. Ragab, H., Milburn, A., Razavi, K., Bos, H., & Giuffrida, C. (2021, May). Crosstalk: Speculative data leaks across cores are real. In *2021 IEEE Symposium on Security and Privacy (SP)* (pp. 1852-1867). IEEE.

ANALYSIS OF THREAT HUNTING APPROACES FOR ADVANCED PERSISTENT THREAT (APT) ATTACK.

Siti Arpah Bint Ahmad[1*], Faizzul Fadhli Bin Ismail[2], Zarina Binti Zainol[3]

[1,3]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[2]Mesiniaga Tower, Ss 16, 47500 Petaling Jaya, Selangor

[1]arpah340@uitm.edu.my,[2]faizzul@mesiniaga.com.my, [3]zarinazainol@uitm.edu.my

ABSTRACT

Nowadays, organizations are increasingly exposed to cyberattacks. One of the dangerous attacks is an advanced persistent threat (APT). This attack is a focus and extended cyberattack. Once affected, it will remain in the organizations, collecting and exploiting all the important resources of the organizations. To mitigate cyberattacks, organizations typically prepare in advance of potential attacks. Proactive threat hunting refers to the practice of preparing for potential attacks before they occur. This work compares the traditional threat hunting and Cyber threat Intelligence (CTI)based threat hunting to detect the APT attack in an organization. Indicator of compromise (IOC), Indicator of Attack (IOA) and Tactics, Techniques and Procedures (TTP) has been used in both approaches. Result shows that CTI based treat hunting is faster compared to the traditional threat hunting.

Keywords: *Advanced Persistent Threat (APT), Cyber threat Intelligence (CTI), Indicator of Compromise (IOC), Indicator of attack (IOA), Threat hunting (TH), Tactics, Techniques and Procedures (TTP)*

1. INTRODUCTION

Advanced persistent threat (APT), is a long-term, cyberattack strategy that once it gains unauthorized access to a network or system it will reminds undetected for an extended time (Ahmad et.al,2022). Types of APT threats are Nation-State APT, Criminal APT dan Cyber-Espionage APT. Example of Nation-state APT groups are APT28(fancy Bear) and APT29 (Cozy Bear). These APTs are believed to be associate to Russian State-sponsored (Letho,2023). Example of the APT28 attack is using Simple Network Management protocol (SNMP) to access into Cisco routers worldwide. Original usage of SNMP is for network administrators to monitor and configure network devices remotely, however it can be misused by attacker to exploit network devices and penetrate a network (Copyright .C, 2023). To mitigate the threat, organization traditionally relies on cybersecurity analysts' experience and intuition to proactively searching for signs of compromise within an organization's network or systems. This approach is ad-hoc and reactive and it is referred to as traditional (IBM, 2023) threat hunting. As for CTI based threat hunting, automation and machine learning techniques are utilized to proactively identify and mitigate threats(CyberProof, 2023). CTI leverage various source such as malware analysis, dark web monitoring, and threat actor behaviour to proactively identify and mitigate potential threats (Bolla and Talentino,2022). IOCs are artifacts that indicate a system has been breached. These artifacts exist after an attack occurred. IOAs are patterns of system/network behaviour that indicate an attack is in progress (Yuval et al, 2020). IOA comes together with TTP. It is an indicators of progress artifacts related to behaviour of the attacker and an attack execution structured framework (Ajmal, 2021). Threat modelling framework for cybersecurity is a process involves identifying and communicating information about the threats that may impact a particular system or network (Fortinet, 2023). Example of threat modelling framework are ISO 27001, OWASP 2021 and MITRE (Guha, 2023). This work compares traditional versus CTI based threat hunting approach to search for any indicator and potential threats of APT28. MITRE framework has been adopted as the threat modelling framework for Brute Force attack. Attackers

(at various points during the breach) use brute force as a trial-and-error technique to gain access to crack passwords, login credentials, and encryption keys (Fortinet, 2023). Finding APT28 signatures is essential for organizations to identify any traces of their presence or actions in an organization's network. This helps gain insights into APT28's methods and procedures, allowing for better understanding and effective responses (Copyright .C, 2023).

2. METHODOLOGY

This work compares traditional threat hunting and CTI based threat hunting to search for any indicator and potential threats of APT28. For traditional threat hunting open-source tools such as Wireshark, Snort, and open-source intrusion detection systems have been used to search for suspected IP addresses, URL and Domain. For CTI based threat hunting, BitDefender XDR and SentinelOne XDR tools have been used. For both approaches signs of compromise are based on industry reports, threat intelligence systems, cybersecurity research organizations, security vendor repositories, and government agencies. IOCs and IOAs used for this work are IP address, a URL, Domain, and File hash. TTPs used for this work are Brute Force termed as T1110. Types of Brute force attacks used by this work are MITRE T1110.001, MITRE T1110.003 and MITRE T1140.

3. RESULTS AND DISCUSSION

The results of the hunt showed no indicators of APT28 in the organization. Figure 1 shows that CTI-based threat hunting is faster than the traditional approach.

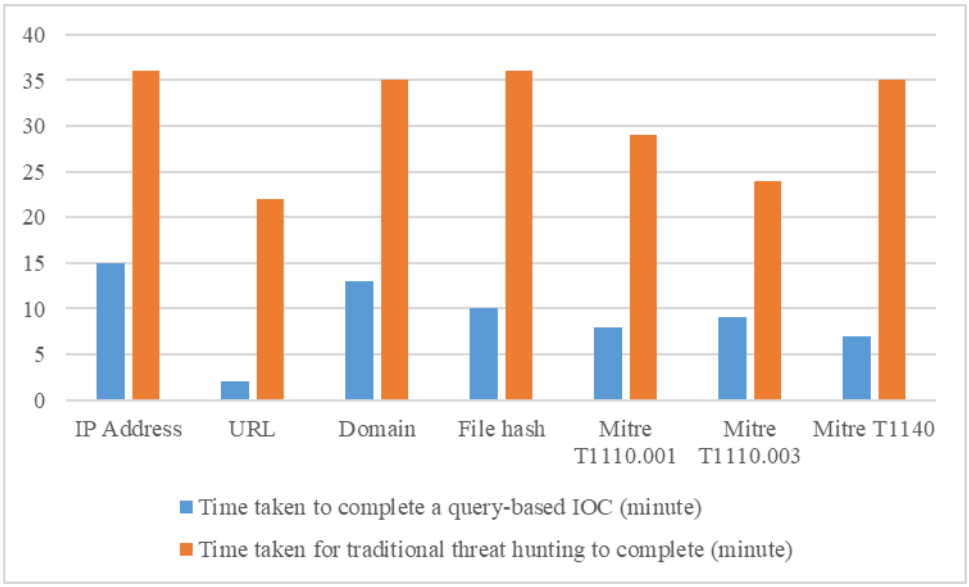

Figure 1. Time taken between traditional and CTI implemented threat hunting.

4. CONCLUSIONS

This work compares traditional versus CTI threat hunting approach for APT28 attack in an organization. Result shows that CTI based threat hunting is faster in getting the IOC, IOA and TTPs for the potential of APT28 signatures. The result of the hunt indicates that the APT28 does not exist in the organization. Even though no evidence of APT28 was found, organizations should remain vigilant and continue monitoring for potential signs of their presence. Threat actors like APT28 continuously evolve, so refreshing CTI sources and reviewing threat hunting strategies regularly is essential.

REFERENCES

1. Ahmad. et al. (2022). Advanced Persistent Threat (APT)-Style Attack Modeling and Testbed for Power Transformer Diagnosis System in a Substation. *2022 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2022*, *2*, 6–10. https://doi.org/10.1109/ISGT50606.2022.9817518
2. Ajmal et al. (2021). Offensive security: Towards proactive threat hunting via adversary emulation. *IEEE Access*, *9*, 126023–126033.https://doi.org/10.1109/access.2021.3104260
3. Bolla, A., & Talentino, F. (2022, April 13). Threat hunting driven by Cyber Threat Intelligence. Webthesis. https://webthesis.biblio.polito.it/22631/
4. Copyright, C. (2023). Advisory APT28 Exploits Known Vulnerability to Carry Out Reconnaissance and Deploy Malware on Cisco Routers. *National Cyber Security Centre.*
5. CyberProof. (2023). *CyberProof*. Retrieved from What is treat hunting: https://www.cyberproof.com/cyber-101/threat-hunting/
6. Fortinet (2023), *Threat Modeling*, Retrieved from https://www.fortinet.com/resources/cyberglossary/threat-modeling#:~:text=Definition%20of%20Threat%20Modeling,they%20may%20impact%20the%20network.
7. Fortinet (2023), *What Is a Brute Force Attack?* Retrieved from https://www.fortinet.com/resources/cyberglossary/brute-force-attack#:~:text=A%20brute%20force%20attack%20is,and%20organizations'%20systems%20and%20networks.
8. Guha.A (2023), Unlocking Synergy: ISO 27001, OWASP 2021, and the MITRE Framework, retrieved from, https://www.linkedin.com/pulse/unlocking-synergy-iso-27001-owasp-2021-mitre-framework-abhirup-guha#:~:text=Together%2C%20they%20enable%20organizations%20to,respond%20to%20real%2Dworld%20threats.
9. IBM. (2023). *IBM Newsletter*. Retrieved from What is threat hunting?: https://www.ibm.com/topics/threat-hunting
10. Lehto, M. (2023). Cyber Warfare and War in Ukraine_0. *Journal of Information Warfare*, *22*(1), 61–75.
11. Puzis, Rami & Zilberman, Polina & Elovici, Yuval. (2020). ATHAFI: Agile Threat Hunting and Forensic Investigation. Retrieved from https://www.researchgate.net/publication/339814086_ATHAFI_Agile_Threat_Hunting_And_Forensic_Investigation/link/5fe0e8f1299bf14088331488/download

EMOTIONAL UX IN OPTICAL WEB DESIGN AMONG URBAN WORKING PEOPLE

Mohd Noor Hafizee Yahaya[1] and Anitawati Mohd Lokman[2*]

[1,2] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 404500 Shah Alam, Selangor, Malaysia

[1]fizieyayasuho@gmail.com, [2]anitawati@uitm.edu.my

ABSTRACT

This study explores how the use of emotion in web design can persuade urban workers to make an online purchase. As technology and the Internet have become an integral component of their company strategy since covid hit Malaysia, optical firms are not left behind in online shopping trend. Since urban workers make up the majority of optical business owners, web design is crucial to ensuring that the company can thrive both online and offline. Problems with cluttered websites can make users not emotionally attached to the websites and thus making them not complete the website process. The objective of this study is to propose a guide for designing emotional UX in optical website design. Thisstudy will use a Kansei checklist to collect the data and analyse it using a multivariate analysis. The results offer insightful information for web designers who are creating optical websites for urban workers by utilising emotional UX strategies. The anticipated outcome of this study is that it willreveal a guide for website designers to use emotional UX to create optical websites.

Keywords: *Kansei Approach, Optical Website, Urban Working People*

1. INTRODUCTION

This study is about how emotional UX in web design can influence urban working people to buy a product through websites. Web searching and online shopping is a day-to-day process that is carried out by many individuals with various different tools (Park, 2019). Optical businesses are not left behind from process of web searching, as technology and the internet has become part and parcel of their business strategy since covid hits malaysia (Photiadis, 2022). Most optical businesses are urban working people thus web design plays an important role to make sure the business can survive online and offline (Ali et al., 2023). A critical component for the adoption and success of websites and electronic commerce has been highlighted as good web design (Prasetia, 2020). In order to create a successful e-commerce website, the goal of this article is to analyse, from a marketing perspective, the key factors that may affect online consumers' perceptions and behaviours. Up until now, there are no studies about emotional design for optical websites web design (Photiadis, 2022). The comparison between emotional UX and web design for optical websites is still unknown. This study will investigate how urban working people will use the website to buy spectacles.

Additionally, this research outlines the outcomes of employing K.E. (Kansei engineering) and multivariate analysis on the data collected during the emotion measurement procedure. Ultimately, the research concludes by proposing a guide for designing emotional UX in optical website design an innovative website designed to elicit and evoke specific target emotions.

2. METHODS

In this research, the results and analysis of the data obtained from the emotion measurement conducted using the web-based Kansei checklists. The results and analysis are presented in the following order: first, by conceptualising emotion in website design by using correlation coefficient analysis (CCA), and then follow by the structure emotion using principal component analysis (PCA). Lastly the requirement analysis using structural equation modeling-partial least squares (SEM-PLS): statistical analysis technique. Principal component analysis (PCA) was utilized to analyze the distribution of variables (Kherif, 2020). The research employed JMP version-14 to perform CCA and PCA, facilitating the understanding and explanation of the emotional structure derived from the evaluation results. It is important to note, as stated by Bartholomew (2002) and Khedher et al. (2015), that the interpretation of PC is subjective, which should be taken into consideration during the discussion. For instance, some may believe that the first PC sufficiently represents the overall emotional structure, while others may suggest the inclusion of additional PC.

This research analysis commences by examining the distribution of variables along axes for the urban working people groups, which resulted from the assessment using the web-based kansei. Subsequently, a similar discussion on distribution will be conducted based on different assessment methods.

3. RESULTS AND DISCUSSION

The conducted analyses have facilitated the development of a proposed guide for web design in optical website, which serves as a guide for designing Kansei Websites. The structure of emotion, as determined through CCA and Principal Component Analysis (PCA), helped conceptualize emotions. Additionally, the design requirements included in the guide were derived from the elements that exhibited the relationship between emotion and element, as indicated by the SEM-PLS.

Table 1. Guide for designing emotional UX in optical website design.

Factor number	Element of emotion	Example Specimen	Design element			
			Background colour	Fontsize	Number of images	location search button
1	Beautiful	6 and 1	Beige or white	Medium	More than 3	Top Middle
2	Professional	8 and 7	Red or grey	Large	2 and 1	Top right

In order to make the most of this guide, designers need to carefully choose a suitable blend of design elements that will evoke the intended emotion. Striking a balance between creativity and adherence to the guides is crucial for the product's success. However, the achievement of developing a Kansei Website hinge upon two crucial factors: first, the idea derived from the outcome of emotion evaluation, and second, the incorporation of the designer's technical skills. Throughout the design process, it is imperative to consult the guide and consider the technical specifications as advised by the web design expert.

4. CONCLUSIONS

In conclusion, this study focused on exploring the emotional user experience (UX) in optical web design among urban working individuals. Three types of analyses were employed to achieve the research objectives. The CCA analysis was used to identify emotional design elements in optical website user interface design. The SEM-PLS analysis evaluated the emotional UX in the same context. Lastly, the PCA analysis facilitated the proposal of a comprehensive guide for designing emotional UX in optical website design. The findings of this study contribute to a deeper understanding of emotional aspects in web design and provide practical insights for enhancing user experiences in the optical industry.

REFERENCES

1. Galbraith, J., Moustaki, I., Bartholomew, D. J., & Steele, F. (2002). The analysis and interpretation of multivariate data for social scientists. In *Chapman and Hall/CRC eBooks*. https://doi.org/10.1201/9781420057454
2. Khedher, L., Ramírez, J., Górriz, J. M., Brahim, A., Segovia, F., & Alzheimer′ s Disease Neuroimaging Initiative. (2015). Early diagnosis of Alzheimer′ s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing,* 151, 139-150.
3. Kherif, F., & Latypova, A. (2020). *Principal component analysis. In Machine Learning* (pp.209-225). Academic Press.
4. Mohd-Ali, B., Low, Y. C., Shahimin, M. M., Arif, N., Abdul-Hamid, H., Abdul-Halim, W. H.W., & Mohidin, N. (2023). Comparison of vision-related quality of life between wearing Orthokeratology lenses and spectacles in myopic children living in Kuala Lumpur. *Contact Lens and Anterior Eye*, *46*(1), 101774.
5. Park, E. (2019). Motivations for customer revisit behavior in online review comments: Analyzing the role of user experience using big data approaches. *Journal of Retailing and Consumer Services, 51*, 14–18. https://doi.org/10.1016/j.jretconser.2019.05.019
6. Prasetia, R., & Syah, T. Y. R. (2020). Impact of web design, assurance, customer service and brand image on intention to adopt internet banking and customer loyalty at Bank Central Asia (BCA). *International Journal of Economics, Business and Accounting Research (IJEBAR), 4*(01). https://jurnal.stie-aas.ac.id/index.php/IJEBAR
7. Photiadis, T., & Papa, V. (2022). 'What's up with ur emotions?' Untangling emotional user experience on Second Life and Facebook. *Behaviour & Information Technology*, *42*(1), 46–59. https://doi.org/10.1080/0144929x.2021.2013537
8. Zhao, R., Zhan, L., Yao, M., & Yang, L. (2020). A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2. 5. *Sustainable Cities and Society*, 56, 102106.

ACCESSIBILITY STUDY OF AUDIO-TACTILE MAP FOR VISUALLY IMPAIRED USERS AT MALAYSIAN ASSOCIATION FOR THE BLIND (MAB)

Nur Adlina Huda Badrul Hisham[1], Nurulhuda Noordin[2] and Fariza Hanis Abdul Razak[3*]

[1,2,3]*College of Computing, Informatics and Mathematics, Univeristi Teknologi MARA, Shah Alam, Selangor, Malaysia.*

[1]n.adlinahuda@gmail.com, [2]drnurul@uitm.edu.my, [3]fariza062@uitm.edu.my

ABSTRACT

A blind person should have spatial awareness, mobility, and orientation abilities to navigate safely in the real world. For them to travel in their daily life, they will face challenges especially when they are in an unfamiliar place, where they need to explore the place by using their cane and their senses. Audio Tactile Map (ATM) is one of the assistive technologies that can help blind people safely reach their destination. Despite being useful, the ATM was found to be used by the users only ONCE. This study aims to evaluate the accessibility of the audio-tactile map at the Malaysian Association for the Blind (MAB to find out why the ATM was not used after the first use. A qualitative research method was adopted using interviews and user studies with five participants from Malaysian Association for the Blind (MAB). Thematic analysis was used to analyze qualitative data. From this research, the audio-tactile map at MAB was found accessible for visually impaired users, however, some improvements should be made especially the symbols and the language used on the ATM. We also learned that the ATM should not only be accessible but also useful in order to make users continue using the ATM: ATM is not just a wayfinding tool but also an information provider.

Keywords: *accessibility, audio-tactile map, visually impaired user, assisted technology, wayfinding*

1. INTRODUCTION

Maps are one of the ways of helping in navigating the people in this world. Visual maps such as printed and digital maps are usually used with the sense of sight, which is using the eyes to understand the map of a place. Information given in the map is usually in a visual map. Therefore, it will be unsuitable for visually impaired users since it is inaccessible for them to navigate themselves to their destination.

Due to the inaccessibility of visual maps, audio-tactile maps have been introduced to the blind. For visually impaired users, tactile graphics are crucial for them to understand the graphics (Götzelmann, 2018). The cognitive burden brought by physiological limitations and the complexity of the graphics makes visually impaired users have difficulties in using tactile graphics (Götzelmann, 2018). They may struggle to estimate distances (Brayda et al., 2019). It is important to develop and enhance the map for them to use in their daily life. From there, they can use it to reach their destination. However, the users need to understand the usage of the map for them to safely navigate without any errors. Audio-tactile maps (ATMs) are alternatives to tactile maps as they provide audio descriptions that can enhance user experience in their wayfinding. Despite its usability, ATMs are found to be unused after the first use. Therefore, this research aimed to evaluate the accessibility of ATMs.

2. METHODS

The study used a qualitative research approach using interview and user testing as main data collection methods. The sampling design was both non-probability sampling, adopted from Creswell (2013), and purposive sampling adopted from Creswell (2013). The participants were five (5) visually impaired users from Malaysian Association for the Blind (MAB) who were above 18 years old. The instrument of the user testing was adapted from V. Trinh et al. (2023) while the questions of the interview were adapted from Griffin et al. (2020). Both interviews and user testing were used to help evaluate the accessibility of the ATMs at MAB. This research framework was adopted from Beebe & Clark (2005), which includes planning, data collection, data analysis, and documentation.

3. RESULTS

Data collected from interviews and user testing were analysed using thematic analysis (Braun & Clarke, 2012). Table 1 shows the themes that have been analysed based on data from interviews and user testing.

Table 1. Themes of usability and accessibility

Usability	Accessibility
1. Findability 2. Understandability 3. Visualizability 4. Effectiveness 5. Satisfaction 6. Safe 7. User Control	1. Language choices 2. Distinct tactile feels 3. Detailed audio description

4. CONCLUSION

From the analysis and findings, the ATMs were basically accessible for visually impaired users, however, there are some enhancements needed to make the ATMs more accessible so that the visually impaired users will find them useful: more language choices, distinct tactile feels, and detailed audio description. ATMs are not just a wayfinding tool, but also an information provider to these users so that they will continue using the ATMs.

REFERENCES

1. Beebe, N. L., & Clark, J. G. (2005). A hierarchical, objectives-based framework for the digitalinvestigations process. *Digital Investigation*, *2*(2), 147-167.
2. Braun, V., & Clarke, V. (2012). *Thematic analysis*. American Psychological Association.
3. Brayda, L., Leo, F., Baccelliere, C., Vigini, C., & Cocchi, E. (2019). A Refreshable Tactile Display Effectively Supports Cognitive Mapping Followed by Orientation and Mobility Tasks: A Comparative Multi-modal Study Involving Blind and Low-vision Participants. In *Proceedings of the 2nd Workshop on Multimedia for Accessible Human Computer Interfaces* (pp. 9-15).

4. Creswell JW. (2013). *Qualitative inquiry and research design: Choosing among five approaches.(*3rd ed.). Thousand Oaks, CA: Sage.
5. Götzelmann, T. (2016). LucentMaps: 3D printed audiovisual tactile maps for blind and visually impaired people. In *Proceedings of the 18th international ACM Sigaccess conference on computers and accessibility* (pp. 81-90).
6. Griffin, E., Picinali, L., & Scase, M. (2020). The effectiveness of an interactive audio-tactile map for the process of cognitive mapping and recall among people with visual impairments. *Brain and behavior*, 10(7), e01650.
7. Trinh, V., Manduchi, R., & Giudice, N. A. (2023). Experimental Evaluation of Multi-Scale Tactile Maps Created with SIM, a Web App for Indoor Map Authoring. *ACM Transactions onAccessible Computing*, 16(2), 1-26.

EXPLORING OLDER ADULT EXPERIENCES AS RESISTORS TOWARDS TECHNOLOGY

Anis Hanani Azman[1] and Emma Nuraihan Mior Ibrahim[2*]

[1,2]School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]anishanani.az@gmail.com, [2]emma@tmsk.uitm.edu.my

ABSTRACT

The increasing ubiquity of technology in modern society has led to growing research on the elderly user experience, particularly in older adults' interaction with technologies. However, this adaptability often prioritizes younger individuals, leaving older adults facing barriers in technology adoption. To understand the factors influencing technology resistance in older adults, this research adopts a life course theory perspective, which emphasizes the impact of life events, social institutions, and personality on individuals' attitudes and decisions. Using a qualitative inquiry methodology, the study engaged in semi-structured interviews with retired Malaysian older adults to gain insights into their experiences with technology. Thematic analysis revealed three key dimensions identified from the literature review: life experiences, personality, and social institutions, along with a new theme, "life transitions." The findings highlight the influence of life events on technology exposure and interaction, the role of personality traits in shaping technology adoption, and the impact of social influences on behavioral intentions. In response to the barriers identified, the research provides recommendations for designing age-appropriate and user-friendly technologies, such as step-by-step instructions, intuitive interfaces, interfaces personalization, auditory personalization, and collaborative design. By applying life course theory to study older adults' barriers towards technology, this research underscores the importance of understanding the cumulative effects of life experiences and broader social structures in shaping attitudes towards technology. The study acknowledges its limitations, particularly the narrow focus on retired Malaysian older adults, limiting the generalizability of the findings to other demographics. Future research should expand the participant pool to include diverse age groups and cultural backgrounds and employ longitudinal research to explore the dynamics of technology barriers over time.

Keywords: *adoption barriers, life course theory, older adults, technology engagement, thematic analysis.*

1. INTRODUCTION

People's lives have been greatly impacted by technological advancement, and this process is currently speeding up rapidly. While some older adults appreciate new technologies, others try to avoid them. Research on the elderly user experience has become increasingly common over the past few years particularly in older adults' interaction with technologies. Recent studies show how older adult ICT users have adopted and incorporated these technologies into their daily life. According to Angelini et al., older adults prefer to utilise personalised technology, so they do not need to ask for help from others. The authors emphasised the significance of creating technology that helps older adults in terms of companionship, offering support and knowledge for long-term and sustainable improvements in health-related technology.

This research positioned and contextualised the notion of barriers and resistors from the lens of life course theory as discussed in (Elder & Giele, 2009; Levy et. al., 2005). This research positioned and contextualised the notion of barriers and resistors from the lens of life course theory as discussed in (Elder & Giele, 2009; Levy et. al., 2005). In this digitized world,

technology plays pivotal role in the older adult's life. Despite numerous works on technology acceptance and adoption in the literature, it is still lacking to our knowledge on the factors contributing to barriers and resistors towards technology from this older adult's perspectives. To gain insight into older adults' experiences with technology and their corresponding opinions, this research employs qualitative inquiry methodology.

2. METHODS

This research uses narrative research through semi-structured interview. Narratives from older adults help to convey the scope of the study problem and comprehend older adults' feelings towards technology such as challenges, perspectives and motivations. The primary data source for this research is the interview, and the interview questions were adapted and modified from Park et al. (2023) and Elder et al. (2003). The secondary data source is a literature review from UiTM Online Database. The sampling technique used for this research is purposive sampling as it is frequently used in research to identify and select cases that possess a wealth of relevant information regarding the topic of study. Alongside knowledge and experience, the availability, willingness to participate, and ability to eloquently, expressively, and thoughtfully articulate views and ideas play a crucial role in producing quality research. The selected participants for this research are five older adults aged older than 60 who have retired from working. The session started with participants telling a little bit about themselves including their previous employment, how long has it been since they retire and what field they worked for. Then, participants were asked questions according to four sections – Aging Experiences and Perspectives, Physical, Cognitive Changes and Technology Interactions, Social Relationships and Challenges, and Future Technology. In the first section, images of smartphones from different stages of evolution were shown to stimulate participants' ideas and past experiences in using them.

3. RESULTS AND DISCUSSION

According to the participant's statements, life events have had an impact on how early they were exposed to technology, how widely it was used in their community, and how they personally interacted with it. older adults who adopted smartphones later in life may hold different attitudes towards technology compared to those who grew up with it.

Personality traits can have a big impact on the potential technological impediments older adults may encounter. These characteristics can affect how much assistance they receive, how open they are to trying new things, how tech-savvy they are, and how patient they are with themselves. For example, one participant was found sceptical of new technologies and is less likely to use them, even if he has the support of others. Another participant, who are not tech-savvy found it difficult to use complex technology, even if she is willing to try. It was also found in a different participant who is impatient and give up easily when they find technology to be too difficult to understand.

Social influence, the point to which a person believes that significant individuals think they should utilise technology and performance expectations have a considerable impact on behavioural intentions (Venkatesh et al., 2003). Older adults rely heavily on their social institutions, specifically their family members, for assistance and support in using technology. They express the need for their family members, including children and grandchildren, to teach them how to use technology and help them navigate through various tasks.

This research found an emergence of new theme defined as life transition that revealed distinctive and important features of older adults' resistance to technology. This suggests that older adults who have gone through a major life change, like retirement, have a harder time adjusting to new technologies that weren't common during their younger years, and older adults who go through changes in their social lives have less motivation to adopt new technologies

that could help them stay in touch with loved ones. Some older adults may encounter difficulties embracing new technologies because they are unfamiliar with the tools and programmes available. Life events like retirement or the death of a spouse, which can reduce possibilities for social engagement and exposure to new technologies, could make worse this lack of familiarity. Figure 1 shows the thematic map illustrated from the research findings.

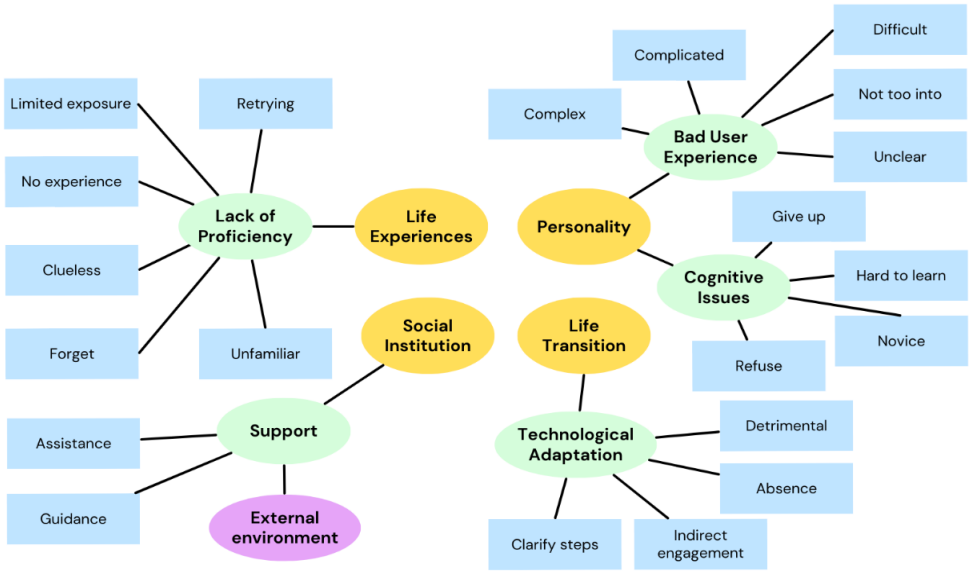

Figure 1. Thematic map.

4. CONCLUSIONS

The first objective is to identify the factors within the context of older adult's personality, life experiences and social institutions influence their decisions to resist technology based on their life course experiences. It is achieved through in the planning and data collection phase which is through the literature review and semi-structured interview with retired older adults. From the first initial thematic analysis process, barrier factors were as found in literature which are life experiences, personality and social institution. After themes reviewing, a non-emergence theme were found which is life transitions. These barrier factors need to be considered when designing technology for older adults in the future.

The second objective is to provide recommendations on the guidelines for designing age-appropriate and user-friendly technologies that cater to the life course experiences of older adults. Some possible recommendations include step-by-step instructions, intuitive interfaces, interfaces personalisation, auditory personalisation, and collaborative design. The recommendations provided in the second objective of this research align closely with the principles of the life course theory. Although most of the technology had provide these accessibilities, this research emphasised on the importance for designers to review and include these when creating technology. By considering the impact of life experiences and transitions on technology adoption and usability, these guidelines strive to create technology that empowers older adults, supports their unique needs, and promotes lifelong digital engagement.

While this study provides useful insights on barriers to utilising technology among older adults, it is important to note that other age groups, cultural backgrounds, and technical experiences may generate different results. The study only included retired adults over the age of 60, limiting the results' application to other demographic groups.

REFERENCES

1. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
2. Elder, G. H., Jr., & Giele, J. Z. (2009). Life course studies: An evolving field. In G. H. Elder, Jr. & J. Z. Giele (Eds.), *The craft of life course research* (pp. 1–24). The Guilford Press.
3. Levy, R., Widmer, E. D., & Kellerhals, J. (2005). Towards an interdisciplinary perspective on the life course. In Ghisletta, P., Lévy, R., & Widmer, E. D. (Eds.), *Towards an Interdisciplinary Perspective on the Life Course* (1st ed.). Elsevier.
4. Lozoya, S. V. M., Guirado, M. a. Z., González, A. Z., & Lopez, A. B. M. (2022). Use of technologies and self-efficacy in older adults. *Revista Iberoamericana De Tecnologías Del Aprendizaje*, *17*(2), 125–130. https://doi.org/10.1109/rita.2022.3166870
5. Park, H. K., Chung, W., & Ha, J. (2023). Acceptance of technology related to healthcare among older Korean adults in rural areas: A mixed-method study. *Technology in Society*, 72, 102182. https://doi.org/10.1016/j.techsoc.2022.102182
6. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478. https://www.jstor.org/stable/30036540
7. Wei, W., Gong, X., Li, J., Tian, K., & Xing, K. (2023). A study on community older people's willingness to use smart home—an extended technology acceptance model with intergenerational relationships. *Frontiers in Public Health*, 11. https://doi.org/10.3389/fpubh.2023.1139667

EMOTIONAL USER EXPERIENCE (UX) AND ITS INDICATORS IN ISLAMIC LIFESTYLE APP AMONG YOUTH

Puteri Yusmiera Mohamad Yusuf[1] and Anitawati Mohd Lokman[2*]

[1,2] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 404500 Shah Alam, Selangor, Malaysia

[1]puteriyusmiera.work@gmail.com, [2]anitawati@uitm.edu.my

ABSTRACT

Studies on user experience in website and mobile app design have mostly concentrated on usability and functionality over the years. The research on product emotion, however, covered how usable design is not as effective as emotional design. As technology grows, Islamic content can be easily accessed nowadays. Developers of mobile apps brings idea and developed a mobile app that focused on content of Islam to bring muslims closer not only to the app, but to the Creator itself. Despite this, there is little research that focused on the emotional experience of mobile apps that give attention to youth preferences. Thus, this paper presents the result of indicators of emotional experience perceived by youth towards Islamic Lifestyle App. The study used a quantitative research approach, and the data was collected from 107 respondents using Google Form questionnaire and analysed by Multi-Variate Analysis method. The results showed positive indicators for Islamic Lifestyle App mobile design elements however, to involve a group of experienced mobile developers to ensure mobile app design fulfilled the emotional elements, can be considered for the future work.

Keywords: *Emotional User Experience, User Experience, Islamic Lifestyle App, Youth*

1. INTRODUCTION

In recent years, there has been a significant increase in the development of mobile applications. As a result, evaluating the usability of mobile applications is an important aspect of technological advancement and application. The rapid and growing number of mobile apps in the Google Play and Apple stores has encouraged developers to develop elevated applications in order to compete in the intense competition app markets (Hussain et al., 2017). The usability of mobile apps is one of the most important aspects of one's quality. Not only that, but the architecture of these applications must also account for a variety of design constraints, such as limited resources, connectivity issues, data entry models, and mobile device display resolutions that vary (Nayebi et al., 2012). The author added, for all types of software or mobile apps, it is critical to consider three aspects of usability, which are easier to use, easy to learn and lastly, it meets or exceeds user expectations.

As the mobile phone market expands and access improves, phone applications or apps will become an important part of user interfaces into Islamic and Muslim-related content (Bunt, 2010). The concept of emotion is inherently difficult to study. Humans struggle to express their feelings and distinguish between various emotions (Agarwal & Meyer, 2009). Not only that, emotions, and their function in the user experience (UX) have been discussed for a while (Garcia & Hammond, 2016). Since emotional experiences leave a lasting effect on our long-term memory, emotional design produces extraordinary user experiences for websites. Recent scientific research suggests that emotional factors be considered when building websites since emotions affect how people see, think about, and make decisions while interacting with websites (Turumogan et al., 2019). The Qur'an would be an important entry point into Islamic apps. While memorization of the Qur'an is regarded as a morality, the Qur'an has traditionally been accessed by listening to or delivering a (portion of) recitation and/or reading the text.

Millions of applications are available for download to any smartphone or tablet, including Islamic mobile applications for religious lifestyle productivity. Most applications have become commonplace in everyone's lives (Zainal et al., 2017) such as MuslimPro, Quran Companion, OnePath Network and many more. However, there are less studies conducted on Islamic Lifestyle App that focus on youth and the indicators of Islamic Lifestyle App that can strengthen the relationship between the app and youth itself. Thus, this paper presents a study of Islamic Lifestyle App, which will be carried out to identify the indicators of emotional UX that can enhance the engagement of youth towards the app itself.

2. METHODS

This study was conducted in a quantitative approach. The primary research data for this study was collected using Kansei Checklist adapted from Lokman (2009) and literature review from the UiTM online database as the secondary research data. The quantitative data collection was conducted using Google Form questionnaires. Non- Probability (Convenience Sampling) was the sampling design chosen for this research as the researcher uses any subjects that are available to participate in the research study. The targeted population for this research was general Islamic apps users in the age range of 25-40 years old, that is considered as Youth in Malaysia (Yunus, 2014). The data collected from the online questionnaires were analyzed by three different analysis techniques: first, Correlation Coefficient Analysis (CCA) was used to examine the emotional connections between Kansei words; second, Factor Analysis (FA) was used to examine the data collected during the evaluation session and identify the factor that would have the greatest impact on element design; and third, Partial Least Squares Analysis (PLS) was used to determine the recommendation for the desired Islamic Lifestyle App. The findings of this analysis will empower the app developers and stakeholders with actionable knowledge to enhance the app's features, content, and user interface, ultimately fostering a more fulfilling and personalized experience for the app's users.

3. RESULTS AND DISCUSSION

After performing CCA analysis, it is evident that out of 12 selected Kansei Words, 3 of the Kansei Words are positively correlated and the rest of it remains negatively correlated. The analysis continued with FA and the results revealed that two factors have dominant influence to emotions. On the other hand, PLS analysis results showed that Prominent Colour – Green is the element of good design and Iconography – Bottom leads to elements of bad design for Islamic Lifestyle App.

4. CONCLUSIONS

The research aimed to study on emotional UX elements of Islamic Lifestyle App among youth and provide element design recommendations. According to the analysis, the application must be improved, especially in terms of emotional design. The study’s conclusion offers recommendations for future enhancements to each UX component for Islamic Lifestyle App. It is advised to enlist a team of expert mobile app developers to assess Islamic Lifestyle App for future work. This is because of their sufficient expertise and originality in app creation. Additionally, during the study phase, the researcher can work with the app development team or stakeholders. Understanding the app's design choices, intended user base, and potential areas for improvement to improve emotional UX may all be benefited by their findings.

REFERENCES

1. Agarwal, A., & Meyer, A. (2009). Beyond usability: Evaluating emotional response as an integral part of the user experience. Conference on Human Factors in Computing Systems - Proceedings, 2919–2930. https://doi.org/10.1145/1520340.1520420
2. Garcia, S. E., & Hammond, L. M. (2016). Capturing & measuring emotions in UX. Conference on Human Factors in Computing Systems - Proceedings, 07-12-May-, 777–785. https://doi.org/10.1145/2851581.2851605
3. Garcia, S. E., & Hammond, L. M. (2016). Capturing & measuring emotions in UX. *Conference on Human Factors in Computing Systems - Proceedings*, 07-12-May-, 777–785. https://doi.org/10.1145/2851581.2851605
4. Hussain, A., Mkpojiogu, E. O. C., Musa, J., & Mortada, S. (2017). A user experience evaluation of Amazon Kindle mobile application. *AIP Conference Proceedings,* 1891(October 2017). https://doi.org/10.1063/1.5005393
5. Nayebi, F., Desharnais, J. M., & Abran, A. (2012). The state of the art of mobile application usability evaluation. *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering: Vision for a Greener Future*, CCECE 2012, May 2014. https://doi.org/10.1109/CCECE.2012.6334930
6. Turumogan, P., Baharum, A., Ismail, I., Noh, N. A. M., Fatah, N. S. A., & Noor, N. A. M. (2019). Evaluating users' emotions for kansei-based Malaysia higher learning institution website using kansei checklist. *Bulletin of Electrical Engineering and Informatics, 8*(1), 328–335. https://doi.org/10.11591/eei.v8i1.1448
7. Yunus, P. M. D. F. (2014). Youth. 20th Century Britain: Economic, Cultural and Social Change, 214–230. https://doi.org/10.5749/minnesota/9780816690411.003.0002
8. Zainal, A., Ahmad, N. A., Razak, F. H. A., & Nordin, A. (2017). Older adults' requirements of islamic mobile applications: A multimethod exploration. *Advanced Science Letters, 23*(5), 4236–4240. https://doi.org/10.1166/asl.2017.8335

DESIGNING MALAYSIA HERB PLANTS APPLICATION USING PICTIVE TECHNIQUE

Izah Shahirah Mohamad[1] and Rozianawaty Osman[2*]

[1, 2]College of Computing, Informatics and Mathematics, University Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

[1]2021255398@student.uitm.edu.my, [2]rozianawaty@uitm.edu.my

ABSTRACT

To preserve and sustain the knowledge of the plants, there is a need for a digital repository of all plants' information. Yet, the information to be stored in the system needs to be explored. The objective of this study is to identify the essential information required for Malaysian herb plants and to determine how best to represent this information digitally. To achieve this goal, a co-design study utilizing the PICTIVE technique is conducted, involving collaboration among ethnobotanists, researchers, and pharmacologists. In the study, various pieces of information were found as important, including images, local names, scientific names, botanical family, characteristics, traditional uses, medicinal applications, parts of the plant used, chemical compounds present, geographical location, habitat, precautions, and reference sources. To effectively convey this information, various ways are proposed, such as utilizing maps to show locations, displaying habitat through images, and using infographics to display plant information. The results are anticipated to be used as an input for the development of the herb plant information system.

Keywords: *Malaysian Herb, Information System, PICTIVE, Information Preservation*

1. INTRODUCTION

Ethnobotany, as defined in the work of Siraj (2022), is a field within the life sciences that explores the intricate relationship between humans and plants. It places particular emphasis on investigating, observing, and identifying the diverse array of plants utilized for addressing health concerns in both humans and livestock. Additionally, Ethnobotany involves the study of how specific groups of people, often those sharing common ethnic backgrounds, cultural histories, and traditions, employ certain herbs. This utilization extends to various aspects of life, encompassing not only medical treatments but also the enhancement of daily living through the crafting and incorporation of these plants into items such as soaps and fibers, as elucidated by Schultes (2008).

Botanists, researchers, and pharmacologists require in-depth knowledge and comprehensive information about specific herbal plants, as highlighted by Mohd Talkah et al. (2013). However, the current online herbal database lacks comprehensiveness and is unsuitable for public sharing, possibly containing outdated data (Roshaizi et al., 2022, Zakaria et al., 2020 & Sahri et al., 2016). Additionally, the database is presented in a scientific format, making it less accessible to the general public. Furthermore, as pointed out by Roshaizi et al. (2022), there is a lack of appreciation for ethnobotanical knowledge due to limited interest, digital records, and information regarding the practical benefits of ethnobotany in everyday life. It is crucial to document ethnobotanical knowledge to prevent its gradual disappearance, as emphasized by Abdul Rahman et al. (2019).

This study sought to determine the necessary information for Malaysian herb plants and their digital representation. It utilized data from ethnobotanical research and existing applications, employing the participatory design method with ethnobotanists, researchers, and pharmacologists to identify required information. The findings were presented digitally through a prototype of the Malaysia Herb Plants Application, encompassing all essential herbal details.

2. METHODS

The aim of this study is to create an application prototype that focused on Malaysian herb plants. It employs participatory design, consisting of two phases: (1) identifying key herb plant information using content analysis, and (2) designing the digital presentation using the PICTIVE technique (Muller, 1991) as shown in Figure 1.

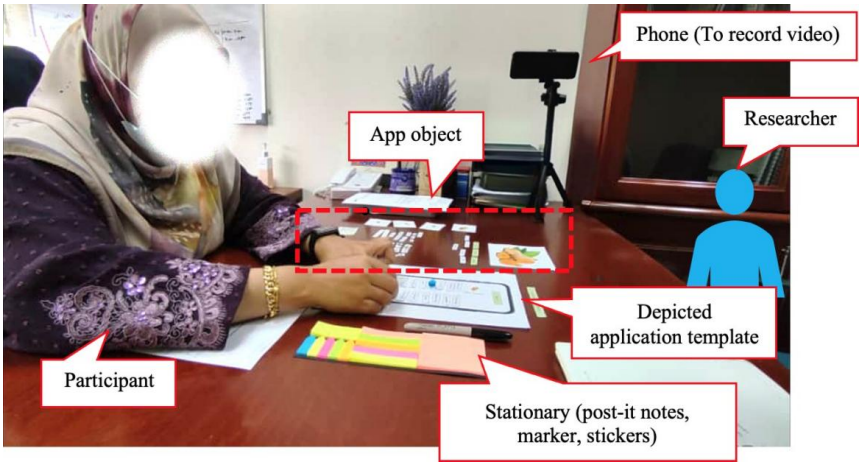

Figure 1. The setting for the data collection in phase 2

Phase 2 of this study includes an individual session, which is preceded by a participant selection process carried out through email or phone communication. Only individuals possessing expertise in ethnobotany were eligible for participation in this phase. The session itself was divided into two parts: the Introduction and Information Display sections.

In Part 1, participants received an explanation of the research's objectives, the tasks they needed to complete, and their consent was obtained. Following that, demographic data such as age, gender, education, and occupation, as well as information about their experience with mobile applications and web-based systems, was collected.

In Part 2, participants were presented with an image of an herb and were then asked to describe how they would prefer the herb's information to be displayed. They were also required to share their design ideas within the designated space (refer to Figure 1). Throughout these tasks, participants' responses were recorded, and the suggested ways to present the herb's information were analyzed across all participants.

3. RESULTS AND DISCUSSION

The findings show that herb information chosen to be displayed include images, local name, scientific name, family, characteristics, traditional and medicinal uses, plant parts used, chemical compounds, location, references, and ethnicity. Additionally, participants recommended the inclusion of habitat details (P3, P4, and P5), precautionary information (P1 and P5), a common name (P3), herb origin (P4 and P5), care instructions (P4), and the incorporation of a herb recognition feature (P2) in the application. Participants also want the information to be presented in a concise and comprehensive manner. To achieve this, they propose the utilization of infographics, an interactive map showcasing herb distribution, and the incorporation of icons. Figure 2 shows the Malaysia Herb Plants Application's prototype that was designed based on the participants' input. It consists of four pages which are Herb List, Herb's Details, Herb ID, and Herb Map.

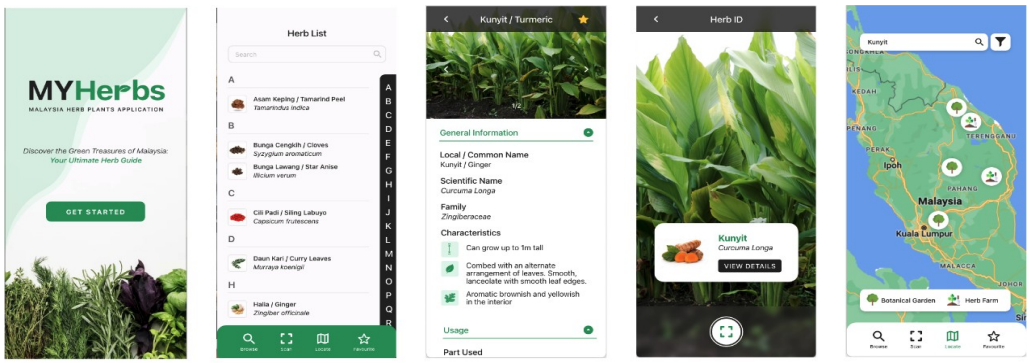

Figure 2. The screenshots of Malaysia Herb Plants Application's prototype.

4. CONCLUSIONS

This study has provided valuable insights into the necessary information and the optimal digital presentation methods for the application. The findings from this study are expected to contribute significantly to the development of the herb plant information application or system. Further research is recommended to gather more data with more participants and validate the information selected in this study.

REFERENCES

1. Abdulrahman, M. D., Ali, A. M., Khandaker, M. M., Nur Fatihah, H. N., & Mat, N. (2019). Review on traditional knowledge and utilisation of medicinal plants. *Biosci. Res*, *16*(2), 3298-3313.
2. Mohd Talkah, S., Zulkiflee, I., & Shamsir, M. S. (2013). Development and deployment of an ethnobotanical and phytochemical knowledge database of Malaysia. *Jurnal Teknologi, 64*(2).79-84 https://doi.org/10.11113/jt.v64.2049
3. Muller, M. J. (1991). Pictive - An exploration in participatory design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Reaching through Technology - CHI '91*. (pp. 225-231). https://doi.org/10.1145/108844.108896
4. Roshaizi, N. A., Omar, M. S. S., & Mustapha, A. (2022). Development of Malaysian Ethnobotanical Online Database. *Journal of Sustainable Natural Resources, 3*(1), 30-43.
5. Sahri, Z. (2016). The design and implementation of Malaysian indigenous herbs knowledge management system based on ontology model. *Knowledge Management International Conference*, (pp. 439-44)
6. Schultes, R.E. (2008). *Ethnobotany. In: Selin, H. (eds) Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures*. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-4425-0_9566.
7. Siraj, J. (2022). *Ethnobotany*. IntechOpen. doi: 10.5772/intechopen.104754
8. Zakaria, S. M., Amri, C. N. A. C., & Shahari, R. (2020). Ethnobotany and traditional knowledge of Acanthaceae in Peninsular Malaysia a review. *Pharmacognosy Journal*, *12*(6).

APPLICATION OF SEMI-ACTIVE SUSPENSION WITH FIREFLY ALGORITHM AND GENETIC ALGORITHM TO IMPROVE RIDE COMFORT

Muhammad Khairul Ikhwan bin Khalid[1], Mohd Zaki Zakaria[2*] and Irwan Mazlin[3]

[1,2,3]College of Computing, Informatics and Mathematics, Univeristi Teknologi MARA, Shah Alam, Selangor, Malaysia.

[1]2021862288@student.uitm.edu.my , [2]zaki@tmsk.uitm.edu.my2, [3]irwanmazlin@gmail.com

ABSTRACT

This research investigates the performance of different suspension systems in improving vehicle ride quality and handling. The study focuses on passive suspension, active suspension, and semi-active suspension systems. Various control strategies and algorithms are examined, including fuzzy logic control and optimization algorithms such as the Firefly Algorithm and Genetic Algorithm. The objective is to optimize the suspension parameters and improve the overall performance of the suspension system in terms of ride comfort, stability, and handling. Experimental evaluations and simulations are conducted to assess the effectiveness of the different suspension systems and control strategies. The results demonstrate that semi-active suspension systems with intelligent control algorithms, such as fuzzy logic and optimization algorithms, can significantly enhance vehicle ride quality and handling, providing a balance between comfort and stability. The findings of this research contribute to the development of advanced suspension systems and control strategies for improved vehicle dynamics and overall driving experience.

Keywords: *Quarter car, magnetorheological damper, Fuzzy Logic controller, Firefly algorithm, Genetic Algorithm.*

1. INTRODUCTION

The role of electronic control systems in modern cars is vital for enhancing driving aspects like safety, handling, and comfort, with the suspension system playing a significant role in ride quality. Springs and shock absorbers are key components, where springs support the car body's movement while shock absorbers dampen the spring's energy, providing a smoother ride (Chen, 2022). Electronic control systems optimize suspension performance by adjusting settings based on sensors monitoring speed, road conditions, and driver inputs. A researched semi-active suspension system combines a magnetorheological damper (MRD) and spring, utilizing magnetorheological fluids to adjust damping force. This study aims to optimize semi-active suspension systems using the Firefly Algorithm (FA) and Genetic Algorithm (GA), exploring their role in fine-tuning control parameters for improved ride comfort and handling. The research also compares the proposed algorithms with traditional methods to assess their efficiency and effectiveness in enhancing vehicle ride comfort. The study's findings can contribute to the automotive industry by developing improved suspension systems for a more comfortable ride.

2. METHODS

This chapter is dedicated to creating mathematical representations of passive and semi-active suspensions, incorporating an MR damper and fuzzy logic controller within the quarter car model. The investigation primarily centers on variables such as vertical displacement,

velocity, and acceleration. The analysis involves two key subsystems: the tire subsystem and the suspension subsystem. Additionally, advanced intelligent techniques like the Firefly algorithm and Genetic algorithm are employed to optimize semi-active suspension parameters, aiming to enhance both ride comfort and handling.

Where, ks=spring stiffness Cs= suspension damping coefficient Fmr = force from actuator. ms= Body mass mu=tire mass Then, the passive system is adapted with MR damper with fuzzy logic controller to generate desire force to improve the system stability based on the velocity of sprung mass and velocity of unsprang mass (Li, 2019). Below is the equation represent the MR damper:

$$F = c_1 \dot{y} + k_1(x - x_0)$$

$$z = -\gamma \mid \dot{x} - \dot{x} \mid\mid z \mid\mid z \mid_{n-1} + \beta(\dot{x} - \dot{y}) \mid z \mid_n + A(\dot{x} - \dot{y})$$

$$y = \frac{1}{c_0 - c_1}(az + c_0\dot{x} + k_0(x - y))$$

$$a = a_a + a_b u$$

$$c_1 = c_{1a} + c_{1b} u$$

$$c_0 = c_{0a} + c_{0b} u$$

$$u = -\eta(u - v)$$

The parameters a, β, and A within the model serve the purpose of adjusting the linearity during unloading and ensuring a smooth transition from the pre-yield to post-yield region. The stiffness of the accumulator is denoted as k1, while the viscous damping observed at higher velocities is represented as co. To introduce roll-off at low velocities, a dashpot, referred to as c1, is integrated into the model. The stiffness at higher velocities is regulated by k0, and x0 denotes the initial displacement of spring k1, associated with the nominal damper owing to the accumulator. Z is the revolutionary variable, and f is used to signify the predicted damping force. Importantly, constants c0 and c1 remain unaffected by the voltage applied to the MR damper as shown.

Next, implement an controller to control and adjust the damping coefficient of the shock absorbers or dampers in real-time based on various input parameters. In this study, fuzzy logic controller is used with two input variables: sprung mass velocity and unsprung mass velocity, with the output being the damping coefficient used to compute the desired force, Fd. These input variables are categorized into three sections: positive (p), zero (z), and negative (n), while the output variable (damping coefficient) is divided into three sections: small (s), medium (m), and large (l). To optimize the membership functions, a heuristic approach is applied, selecting the best membership function for body acceleration by minimizing the least mean square error. Sinusoidal membership functions are chosen to ensure smoother control, maintaining flatness at the top of the trapezoidal shape. Similar membership functions are used for relative velocity and sprung velocity.

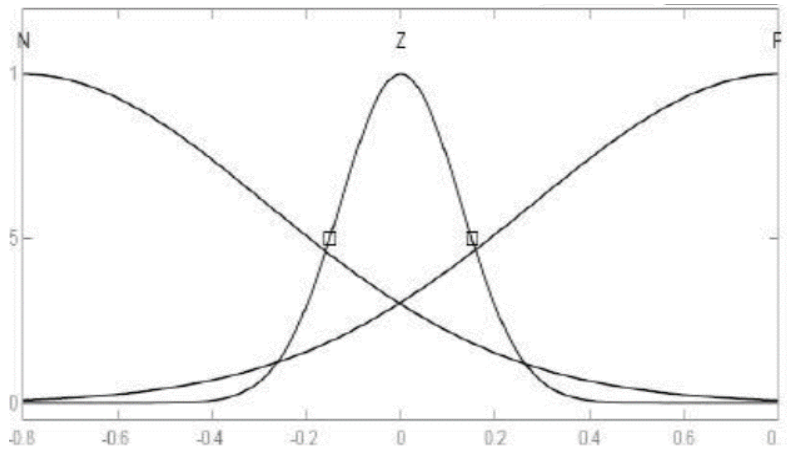

Figure 1: Fuzzy logic membership functions

$If\ \dot{z}\ is\ (P) and\ \ddot{z}\ is\ (N)\ then\ output\ is\ Cd$

$If\ \dot{z}\ is\ (P) and\ \ddot{z}\ is\ (N)\ then\ output\ is\ Cd.$

L=$(Cp.Cd1.Cd2.Cd3.Cd4.Cmax)$

Table 1: Prescript output value of fuzzy system

L	Value
Cp	700
Cd1	3000
Cd2	6000
Cd3	9000
Cd4	12000
Cmax	15000

Lastly, the implement of optimization algorithm to enhance the fuzzy logic performance to control Fd for the suspension system. In this study, the firefly algorithm and genetic algorithm has been developed with body acceleration as objective function. The Firefly Algorithm operates based on three fundamental principles. Firstly, all individuals within a firefly population lack gender distinction, meaning that each firefly can attract others. Attraction among fireflies hinges on their individual levels of brightness. Fireflies with lower brightness are naturally inclined to move towards those emitting stronger light, causing them to gravitate towards brighter peers. In the absence of a brighter firefly, they move randomly. Attractiveness is directly linked to brightness and diminishes as the distance between fireflies increases. The brightness of each firefly corresponds to the objective function being optimized. Below is the pseudo code for FA and GA.

```
1.  Define objective function f(a), where a = (a1,..... ad)
2.  Produce an underlying population of fireflies
3.  Formulate the light intensity L
4.  Specify the absorption coefficient β
5.  While (t< Max_Gen)
6.  For i=1 to n (all n fireflies)
7.    For j=1 to n (all n fireflies)
8.        If ( Lj > Li ), move firefly i towards firefly j
9.      End if
10.       Examine new solutions and update light
          intensity;
11.   End for j
12.  End for i
13.  Rank the fireflies and find the current best
14.  End while
```

Figure 2: FA pseudocode

The genetic algorithm (GA) operates on three fundamental principles. Firstly, all individuals in the population are encoded as binary strings, representing potential solutions with gene sequences. These genes correspond to specific problem parameters. Within the GA, all individuals possess the capacity to attract others through fitness evaluation, determined by their objective function value, signifying their problem-solving performance. Individuals with higher fitness are naturally drawn to those with superior performance and have a higher likelihood of reproductive selection for the next generation, mirroring the concept of survival of the fittest. During reproduction, crossover and mutation operators create offspring inheriting traits from parents, facilitating exploration of the solution space. Even less fit individuals may reproduce to maintain diversity, preventing premature convergence and encouraging exploration. The GA iteratively applies selection, crossover, and mutation for defined generations or until a termination condition, converging toward optimal solutions through successive generations, with fitness directly linked to solution quality.

parameter(s): S – set of blocks
output: superstring of set S

Initialization :
$t \leftarrow 0$
Initialize P_t *to random individuals from* S^*
EVALUATE-FITNESS-GA(S, P_t)

while *termination condition not met*
do
- *Select individuals from* P_t *(fitness proportionate)*
- *Recombine individuals*
- *Mutate individuals*
- EVALUATE-FITNESS-GA$(S, modified\ individuals)$
- $P_{t+1} \leftarrow$ *newly created individuals*
- $t \leftarrow t + 1$

return *(superstring derived from best individual in* P_t*)*

procedure EVALUATE-FITNESS-GA(S, P)
S – *set of blocks*
P – *population of individuals*
for each *individual* $i \in P$
do
- *generate derived string* $s(i)$
- $m \leftarrow$ *all blocks from* S *that are not covered by* $s(i)$
- $s'(i) \leftarrow$ *concatenation of* $s(i)$ *and* m
- $fitness(i) \leftarrow \frac{1}{\|s'(i)\|^2}$

Figure 3: FA pseudocode

3. RESULTS

The main objective of this study is to minimize the vertical amplitude of both body acceleration and displacement. The effectiveness of the proposed FL-FA (Fuzzy logic controller) and FL-GA is assessed by comparing it with the FL (Modified Fuzzy logic suspension system) and a passive suspension system. A bump with amplitude of 0.1m was

applied to all vehicle system. Below figure show the performance of the body acceleration and displacement, respectively.

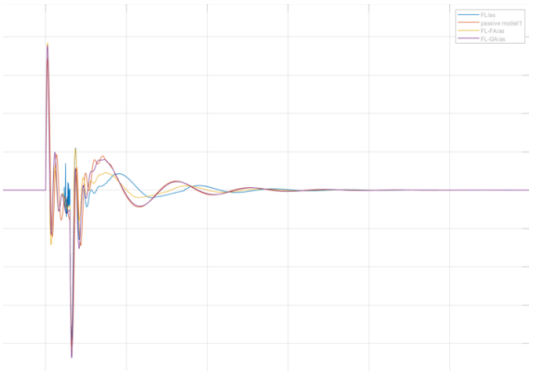

Figure 2: Body Acceleration over time

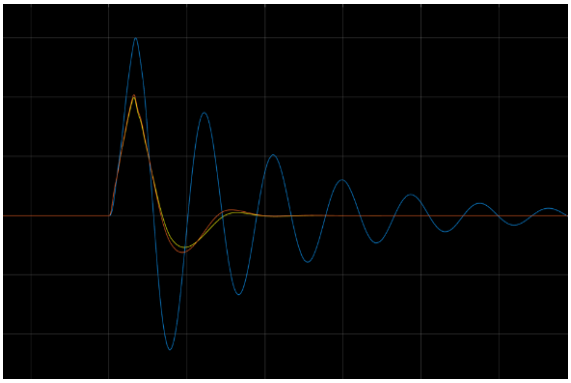

Figure 3: Body Displacement over time

Table 2: MSE result of the system

Mean Square Error				
Parameter of interest	**Passive**	**Fuzzy Logic**	**FL-FA**	**FL-GA**
Body displacement	0.00016	0.04115	0.03236	0.005443
Body acceleration	0.001608	0.0003884	0.0002478	0.0006954

4. CONCLUSION

In conclusion, this thesis comprehensively addresses the performance evaluation of passive, Fuzzy Logic (FL), and FL controllers tuned with Firefly Algorithm (FA) and Genetic Algorithm (GA) in semi-active suspension systems. The validity of equations and models used throughout the research has been verified, establishing a foundation for future studies. The FL-FA controller effectively reduces overshoot in body displacement and settling time for bump input. However, it shows a limited improvement in settling time for body acceleration compared to the FL controller. Additionally, both FL-FA and FL-GA controllers slightly outperform the passive suspension system in bump input scenarios, with FL-FA and Fuzzy Logic controllers significantly reducing body displacement overshoot. Ultimately, FL-FA outperforms the conventional FL controller, benefiting from intelligent tuning to enhance system stability and performance. A comparison between FL-FA and FL-GA reveals that FL-FA offers better comfort and handling, although it results in longer settling times for body acceleration. Future research prospects include the real-time implementation of optimized fuzzy logic control strategies and gain values, potentially integrating these advances into actual vehicle suspension systems.

REFERENCES

1. Chen, X., Song, H., Sixia, Z., & Liyou, X. (2022). Ride comfort investigation of semi-active seat suspension integrated with quarter car model. *Mechanics & Industry*, *23*, 18. https://doi.org/10.1051/meca/2022020.

2. Yajuan, C., You, F., Li, X., & Fu, L. (2019). Research and Optimization of Vehicle Handling Stability Based on Multi - body Dynamics. *IOP Conference Series: Materials Science and Engineering*, 688, 033009. https://doi.org/10.1088/1757-899x/688/3/033009
3. Qazi, A., De Silva, C. W., Khan, A., & Khan, M. S. (2014). Performance Analysis of a Semiactive Suspension System with Particle Swarm Optimization and Fuzzy Logic Control. *The Scientific World Journal*, 2014, 1–12. https://doi.org/10.1155/2014/174102
4. Li, Z., Gong, Y., Li, S., & Wang, W. (2019). Magnetic hysteresis compensation control of a magnetorheological damper. Frontiers in Materials, 6. https://doi.org/10.3389/fmats.2019.00299
5. Talib, M. H. A., Darus, I. Z. M., & Samin, P. M. (2018). Fuzzy logic with a novel advanced firefly algorithm and sensitivity analysis for semi-active suspension system using magneto-rheological damper. *Journal of Ambient Intelligence and Humanized Computing*, *10*(8), 3263–3278. https://doi.org/10.1007/s12652-018-1044-4
6. Griffin, E., Picinali, L., & Scase, M. (2020). The effectiveness of an interactive audio-tactile map for the process of cognitive mapping and recall among people with visual impairments. *Brain and behavior*, 10(7), e01650.
7. Trinh, V., Manduchi, R., & Giudice, N. A. (2023). Experimental Evaluation of Multi-Scale Tactile Maps Created with SIM, a Web App for Indoor Map Authoring. *ACM Transactions on Accessible Computing*, 16(2), 1-26.

KNOWLEDGE MODELLING FOR TYPE-2 DIABETIC DIET KNOWLEDGE-BASED SYSTEM

Nazzatul Farahidayah Mohd Nazziri[1] and Shamimi A. Halim[2*]

[1,2] School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]2021415316@student.uitm.edu.my, [2] shamimi@fskm.uitm.edu.my

ABSTRACT

Effective knowledge modeling is vital for supporting type 2 diabetes patients in managing their diets. Formal models are essential tools for expressing ideas, relationships, principles, and limitations. Knowledge-based systems (KBS), particularly utilizing the CommonKADS knowledge engineering methodology, are instrumental in structuring and disseminating complex information. In dietary guidance for individuals with type 2 diabetes, a pressing challenge is providing precise, personalized advice amidst varying factors such as food choices, evolving research, and shifting standards. The internet's abundance of information highlights this issue by offering unstructured and fragmented diabetes-related knowledge, causing inconsistencies and contradictions and making decision-making more challenging. This study employs CommonKADS assessment knowledge model to manage complex data concerning type 2 diabetes patients' meals, facilitating systematic and evidence-based dietary recommendations. Based on the model, the system is developed and tested with feedback to improve the system. The outcome of this study empowers healthcare professionals, nutritionists, and type 2 diabetes patients with personalized nutritional advice. It also improved self-management skills and subsequently enhanced long-term health outcomes. The use of knowledge modeling, particularly through CommonKADS, is essential for regulating dietary plans for type 2 diabetes patients, culminating in a personalized dietary guidance system designed to enhance health and well-being.

Keywords: *Type 2 diabetes, Knowledge Modelling, CommonKADS, Knowledge-based Systems*

1. INTRODUCTION

The prevalence of obesity and type-2 diabetes (T2D) has emerged as a significant concern in developed countries, giving rise to various health risks. T2D, characterized by insulin resistance and elevated blood sugar levels, can lead to cardiovascular disease and other complications. Effective T2D management demands a comprehensive approach, including dietary modifications (Butt, 2022). Nevertheless, existing online food resources often lack standardization and fail to cater to the specific nutritional needs of T2D patients (Moschonis et al., 2023; Agastiya et al., 2022; den Braber et al.). The absence of robust knowledge management can constrain the effectiveness of knowledge-based systems and limit the availability of real-time feedback and monitoring tools (Dal Mas, et al., 2020). A strong grasp of nutrition and dietary requirements is essential for crafting personalized and efficacious diet plans. The CommonKADS methodology is harnessed to address this challenge (Allen, et al., 2023) and develop a knowledge-based system for managing and dispensing personalized dietary advice to T2D patients (Batarseh, et al., 2021). Generally, CommonKADS or Common Knowledge Acquisition and Design Structuring is a renowned methodology in the field of knowledge engineering and artificial intelligence. It provides a structured framework for the systematic development of knowledge-based systems, facilitating the modeling and

organization of knowledge to solve complex problems efficiently (Cahyaningsih, et al., 2022). In the context of T2D, CommonKADS assessment knowledge model able to facilitate the organization and utilization of scattered knowledge pertaining to T2D diets. By encompassing variables like carbohydrate intake, fiber intake, glycemic index, and portion control, the knowledge-based system can generate precise and tailored diet plans. This study aims to construct and validate CommonKADS assessment knowledge models for T2D patient diets, with the ultimate goal of enhancing the management and control of blood glucose levels through an effective knowledge-based system.

2. METHODS

The study comprises five key stages, commencing with the collection and acquisition of information. These stages encompass the creation of task and domain knowledge models, the application of the CommonKADS Adaptation Guideline, the construction of inference knowledge, and the subsequent development and validation of prototypes. A critical aspect of the methodology involves the adaptation of CommonKADS generic assessment task template as the use of the CommonKADS methodology requires a deep understanding of the inferences embedded within the generic task template. To address this, we adopt a CommonKADS knowledge model adaption underlined by Halim, S. A. (2021). Figure 1 shows the steps that we have adopted in adapting the knowledge of T2D with the CommonKADS generic assessment task template.

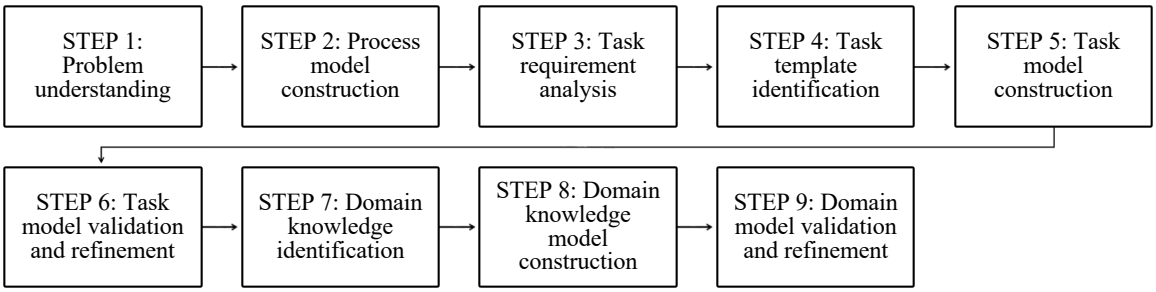

Figure 1. CommonKADS adaption guideline.

3. RESULTS AND DISCUSSION

The inference task model is a key component of the CommonKADS methodology, providing a framework for reasoning and generating decisions based onavailable information. Figure 2 shows Type-2 Diabetic Patient Dietary Assessment Task Model which consist of six (6) inference models (in oval shape) which are i) abstract, ii) specify patient, iii) select, iv) specify diet, v) evaluate and vi) match. As a model validation, several test cases were used and think-aloud strategy are practiced with dieticians. Below is example of test case and Table 1 shows the input and output for each of the inferences.

"Jake, an 11-year-old boy who was recently diagnosed with type 2 diabetes. Despite his young age, Jake has been struggling with hypertension and is also underweight, making it crucial to address his nutritional needs to maintain a healthy balance."

Table 1. Input and output for each inference.

Input	Inference	Output
Gender: Male, Age: 11, Weight: 35, Taking Insulin: No, Blood Pressure: 120, Blood Sugar: 9, Smoking: No, Hypertension: Yes, Activity Days per week:3, Stress Level:2,	Abstract	Insulin sensitivity: reduced insulin sensitivity
Insulin sensitivity: reduced insulin sensitivity, Calories Need: Weight Segment, Age Segment, Blood Pressure Segment	Specify Patient	Calories: 1918, Weight Segment: Underweight, Age Segment: Children, Blood Pressure Segment: High
Calories: 1918, Weight Segment: Underweight, Age Segment: Children, Blood Pressure Segment: High	Select	10 Dietary set for children
10 Dietary set for children	Specify Diet	Dietary set #120
Dietary set #120	Evaluate	Dietary set #120 still suitable or same with the previous dietary set
Dietary set #120	Match	Confirmation with the facts: Dietary recommendation

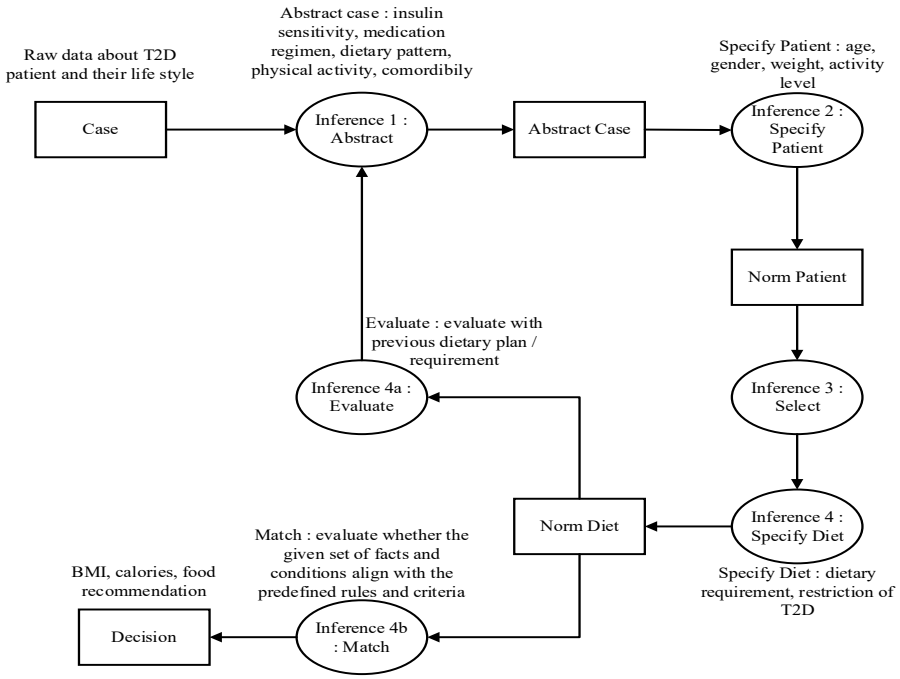

Figure 2. Type-2 Diabetic Patient Dietary Assessment Task Model

4. CONCLUSIONS

The significance of this study lies in its contribution to the field of personalized dietary recommendations for T2D patients. By adapting the CommonKADS methodology in developing the knowledge-based system, the study offers a structured and evidence-based approach to generate accurate and tailored dietary guidance. The study provides valuable insights into the assessment knowledge model, inference tasks, and knowledge representation techniques for effective decision-making. Based on the findings, it is recommended to further evaluate and refine the task model with more test cases. Continuous updates and improvements to the model will ensure that the knowledge-based system remains up-to-date with the latest knowledge and guidelines in T2D management.

REFERENCES

1. Adu, M. D., Malabu, U. H., Malau-Aduli, A., & Malau-Aduli, B. S. (2018). Users' preferences and design recommendations to promote engagements with mobile apps for diabetes self-management: Multi-national perspectives. *PLOS ONE*, *13*(12), e0208942. https://doi.org/10.1371/journal.pone.0208942
2. Halim, S. A. (2021). *Enhancement Of Commonkads Assessment TaskFor Domain Expert Maintainable Knowledge-Based System* [Doctoral dissertation, Universiti Teknologi MARA].
3. den Braber, N., Vollenbroek-Hutten, M. M., Oosterwijk, M. M., Gant, C. M., Hagedoorn, I. J., van Beijnum, B. J. F., ... & Laverman, G. D. (2019). Requirements of an application to monitor diet, physical activity and glucose values in patients with type 2 diabetes: the diameter. *Nutrients*, *11*(2), 409.
4. Agastiya, I. M. C., Kurianto, E., Akalili, H., & Wicaksana, A. L. (2022). The impact of telehealth on self-management of patients with type 2 diabetes: a systematic review on interventional studies. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *16*(5), 102485.
5. Dal Mas, F., Garcia-Perez, A., Sousa, M. J., da Costa, R. L., & Cobianchi, L. (2020). Knowledge translation in the healthcare sector. A structured literature reviews. *Electronic Journal of Knowledge Management*, *18*(3), 198-211.
6. Allen, B. P., Ilievski, F., & Joshi, S. (2023). Identifying and Consolidating Knowledge Engineering Requirements. *arXiv preprint* arXiv:2306.15124. https://doi.org/10.48550/arXiv.2306.15124
7. Batarseh, F. A., Freeman, L., & Huang, C. H. (2021). A survey on artificial intelligence assurance. *Journal of Big Data*, *8*(1), 60.
8. Cahyaningsih, E., Silalahi, N. L. E., Rohajawati, S., & Avianti, Y. M. (2022, November). COMMONKADS for Knowledge Based System Development: A Literature Study. In *2022 International Conference on Information Technology Systems and Innovation (ICITSI)* (pp. 213-218). IEEE.

CARDIAC REHAB EXERCISE RECOMMENDATION

Nur Fadhlin Sakina Rosdi[1], Azliza Mohd Ali[2*] and Sharifalillah Nordin[3]

[1,2,3] *School of Computing Sciences, College of Computing, Informatics and Mathematics*
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]2021640548@student.uitm.edu.my, [2]azliza@tmsk.uitm.edu.my, [3]sharifa@tmsk.uitm.edu.my

ABSTRACT
Cardiac rehabilitation (CR) is a complete therapy approach for cardiovascular disease (CVD) patients that focuses on providing supervised aerobic exercise training. Despite the fact that both cardiac rehabilitation and physical activity are safe and improve outcomes, only a small percentage of patients with heart failure participate in either. Self-motivation, enjoyment of physical activity, and self-efficacy are all important factors in physical activity sustainability. If this problem is being ignored, there will be high chances that the heart failure disease to reoccur thus increase the rate of death in any region. In an effort to address this problem, recommendation is designed and developed in order to give exercise recommendation to the cardiac patient based on their preferences. Recommendation system take into account the preferences of a particular user, compare it to what other users with similar preferences liked or disliked and try to predict the information that would satisfy the user the most. This recommendation system can assist the cardiac patient to find their best preferred exercises. This project presents the design and the development of Collaborative Filtering technique with the purpose of giving recommendation to the low risks cardiac patient based on the past patient's rating and preferences. A Web-Based system is built to make it easier for the end user to view and communicate with the system. This project delivered a recommendation system that able to recommend exercises based on the user's preferences using the Collaborative Filtering through cosine similarity method. This project can be enhanced by adding more historical patient in order to obtain more accurate recommendation.

Keywords*: Cardiac Rehabilitation, Cardiovascular Disease,* Exercise Recommendation

1. INTRODUCTION

Cardiovascular disease (CVD) has overtaken all other causes of death and disability in the world. In addition to shortness of breath, chest pain, and a reduction in quality of life related to health, Cardiovascular disease (CVD) also places a heavy financial burden on healthcare systems (Tsao et al., 2022). The American Heart Association (AHA) and the American College of Cardiology (ACC) both recommend cardiac rehabilitation (CR) as a class 1A treatment because it lowers mortality and subsequent cardiovascular events. Although many eligible patients are not enrolled in or finish CR programmes, it is still a relatively underutilised intervention (Chindhy et al., 2020). Cardiac rehabilitation (CR), which aims to lower cardiovascular risk, mortality, and readmission rates. Research conducted by (Klompstra et al., 2022) emphasizes the importance of self-motivation, enjoyment of physical activity, and self-efficacy in sustaining regular physical activity. In comparison to those who engaged in continuously supervised exercise or standard care, patients who were given the freedom to choose their own exercise activities after a period of supervised exercise training showed an increase in submaximal exercise capacity over the following 12 months (Conraads et al., 2012).

Despite the fact that both cardiac rehabilitation and physical activity are safe and improve outcomes, only a small percentage of patients with heart failure participate in either. Self-motivation, enjoyment of physical activity, and self-efficacy are all important factors in physical activity sustainability. However, in reality, healthcare professionals frequently control CR

programs, and patient preferences are poorly understood and need more research. Health care professionals are advised to consider patient preferences and make more of an effort to include them in the decision-making process when creating individualized CR programs for their patients (Dalal et al., 2021). To give this privilege to user, a system called cardiac rehab exercises recommender system using Collaborative Filtering technique able to fully understand what patients need. Collaborative filtering can introduce users to previously unknown exercises or workout routines. By taking into account their preferences and choices, the system can identify exercises that are popular or highly rated among users with similar exercise profiles. This aids users in discovering a diverse range of exercises.

2. METHODS

This research consists of eight stages which are the preliminary study, knowledge acquisition, knowledge representation, modelling, system design, system development, system testing and documentation. These eight phases are very important in developing cardiac rehab exercise recommender system. In the preliminary study, topic related to cardiac rehab and recommender system are studied thoroughly through the literature review gatherings. Next phase is data acquisition where data is gathered through MyCardiacRehab admin and past student's data. Once the data is gathered, data representation is done in order to ease the calculation process that happened in the next phase which is the modelling. In modelling phase, all steps to acquire the rating prediction is listed down which includes the cosine similarity calculation and rating prediction calculation. The output is then used as the recommendation to the new patient. Next is system design and develop where it includes all the process of developing user interface. Next, testing is done in order to compare the accuracy of the rating produced by the algorithm.

3. RESULTS AND DISCUSSION

For exercise prediction on new patient, calculation for cosine similarity for existing patient 1 until patient 10 is done in order to find the similarity between existing patient and patient 11. Taking patient 11 as the new patient, the algorithm is applied in finding the similarity between patient 1 until 10 and new patient which is patient 11. Before calculation is done, all exercise that has been rated randomly by patient 1 to 10 is compiled. In order to get the similarity value, all the rating is loaded in the algorithm and cosine similarity is applied. Algorithm is build using Python by applying the core concept of collaborative filtering. Firstly, data from excel is imported to Python. Attribute selected is the exercise rating from existing 10 patients. Once all the data for patients 1 until patient 10 is fitted in one data frame, cosine similarity is calculated in order to find the pattern between existing patient and new patient. In this case, patient 11. For this model, every patient executed the same exercise for week 1 and gave different ratings. From the ratings given, cosine similarity is calculated to give recommendation for Week 2 belongs to patient 11. The function of cosine similarity is to calculate the similarity or distance between two vectors in a high-dimensional space based on their angle. The cosine similarity function computes the cosine of the angle between two vectors and returns a value between -1 and 1. If the cosine similarity is 1, the two vectors are perfectly similar and point in the same direction. If the cosine similarity is -1, the two vectors are completely dissimilar, pointing in opposite directions. If the cosine similarity is zero, the two vectors are orthogonal or independent of one another. Below is the result for each of the cosine similarity.

Table 1. Cosine Similarity.

	Patient 11	**Patient 12**	**Patient 13**	**Patient 14**
Patient 1	0.868	0.874	0.908	0.754
Patient 2	0.907	0.899	0.830	0.861
Patient 3	0.881	0.886	0.820	0.798
Patient 4	0.884	0.859	0.827	0.810
Patient 5	0.939	0.916	0.854	0.871
Patient 5	0.855	0.888	0.819	0.779
Patient 7	0.869	0.901	0.907	0.763
Patient 8	0.923	0.865	0.865	0.756
Patient 9	0.916	0.872	0.852	0.775
Patient 10	0.899	0.748	0.675	0.749

Table 2. Predicted Rating.

Activity Category	**Activity Name**	**Predicted Rating**
Strength	Prone Back Extension	8.679
Strength	Tricep Kickback	6.636
Warm Up	WarmUp6	7.654
Warm Up	WarmUp3	6.340
Warm Up	WarmUp4	6.303
Cardio	Squat with Double Side Steps	9.006
Cardio	Frontal Kick	8.994

Once prediction is done, user will input the actual rating for each of the exercises executed. This input will be used to calculate the Mean Absolute Error, it measures how well a prediction model or system performs in terms of the absolute magnitude of its errors. A lower MAE indicates greater accuracy and closer alignment of predicted and actual values.

Table 3. Mean Absolute Error

Week	**Day**	**MAE**
2	1	1.4221
	2	1.1617
	3	0.7951

Table 3 shows that Day 1, MAE value 1.4221 represents a level of error that is moderate. It indicates that the predicted ratings deviate from the actual ratings by around 1.42 units on average. While it is not extremely high, it indicates that there is room for improvement in prediction accuracy. Meanwhile for Day 2, the MAE of 1.1617 indicates that the predicted ratings deviate from the actual ratings by approximately 1.1617 units on average. On a scale of 0 to 10, this deviation represents a relatively minor error. To put it into context, the MAE of 1.1617 is less than half the range of the rating scale. While this suggests that there is room for improvement in prediction accuracy, it also suggests that predicted ratings are reasonably close to actual ratings on average. Lastly, for Day 3 Week 2, a MAE of 0.7951 indicates a low level of error. It implies that the predicted ratings deviate from the actual ratings by approximately

0.7951 units on average. With a rating range of 0 to 10, a MAE of 0.7951 indicates that the predicted ratings are, on average, fairly close to the actual ratings. The deviation is less than one-tenth of the rating scale, indicating that the prediction was reasonably accurate. In summary, a MAE of 0.7951 indicates that predicted ratings are reasonably close to actual ratings on a scale of 0 to 10. It denotes a reasonably accurate prediction with a moderate margin of error.

4. CONCLUSIONS

In a conclusion, health is not something one should take lightly. Cardiovascular rehabilitation is critical for assisting people with cardiovascular conditions in their recovery and overall well-being. Patient's ratings of the exercises prescribed to them are critical in this context because they provide important information about the patients' perceptions of and experiences with the exercises. One of the most important reasons for the importance of patient ratings on exercise is their role in tailoring treatment to individual needs. Healthcare professionals can make informed changes to the exercise programme by soliciting feedback and understanding patient preferences and responses to various exercises. This customized approach ensures that the prescribed exercises are appropriate, tolerable, and enjoyable for the patient. When patients actively participate in exercises they enjoy, their adherence and engagement improve, resulting in better health outcomes and overall satisfaction. For development process, researcher decided to use Python Programming Language to build the system by using Anaconda as the medium. The usage of this programming language is due to the ability of developing web.

REFERENCES

1. Chindhy, S., Taub, P. R., Lavie, C. J., & Shen, J. (2020). Current challenges in cardiac rehabilitation: strategies to overcome social factors and attendance barriers. Expert review of cardiovascular therapy, 18(11), 777-789.
2. Conraads, V. M., Deaton, C., Piotrowicz, E., Santaularia, N. J., Tierney, S., Piepoli, M. F., Pieske, B., Schmid, J. P., Dickstein, K., Ponikowski, P., & Jaarsma, T. (2012). Adherence of heart failure patients to exercise: barriers and possible solutions. European Journal of Heart Failure, 14(5), 451–458. https://doi.org/10.1093/eurjhf/hfs048
3. Dalal, H. M., Doherty, P., McDonagh, S. T., Paul, K., & Taylor, R. S. (2021). Virtual and in-person cardiac rehabilitation. bmj, 373.
4. Klompstra, L., Deka, P., Almenar, L., Pathak, D., Muñoz-Gómez, E., López-Vilella, R., & Marques-Sule, E. (2022). Physical activity enjoyment, exercise motivation, and physical activity in patients with heart failure: A mediation analysis. Clinical Rehabilitation, 36(10), 1324-1331.
5. Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., ... & American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2022). Heart disease and stroke statistics—2022 update: a report from the American Heart Association. *Circulation*, *145*(8), e153-e639.

CHATBOT FOR CAT VETERINARY CONSULTATION

Nur Hanani Hasbullah[1] and Siti Nur Kamaliah Kamarudin[2*]

[1,2] School of Computing Sciences, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[1]2021213476@student.uitm.edu.my, [2]snkamaliah@uitm.edu.my

ABSTRACT

A Chatbot is an integrated structure that is intended to initiate a conversation with human customers or a variety of chatbots by text. These chatbots generally converse through auditory or textual methods, and they can effortlessly mimic human languages to communicate with human beings in a human-like way. It may be a way for pet owners to find home treatment steps to treat their own cats, as animals, in general, are unable to verbalize their pain. This may save the cost and time to travel to see a veterinarian. Hence, in this research, MikoBot was developed with the aim to help answer questions related to common knowledge about cats, such as the type of cat illnesses, home treatment, and cat care. The objective of this project is to determine the common cat illnesses that can be self-treated by the pet owner and to develop a functional chatbot prototype (MikoBot) for veterinary consultations specifically for cats. Data was collected from a user questionnaire, and in 30% of the results, users have tried to use MikoBot to consult with a veterinarian or get information about cats. The method to create the chatbot uses Javascript, HTML, and CSS code, with a connection to the real-time database and API using Firebase. MikoBot managed to give real-time responses to users and thus can be improved in the future by incorporating more knowledge in the database, as well as expanding to other animals apart from cats.

Keywords: *Chatbot, Consultations, Database, Firebase, Veterinary*

1. INTRODUCTION

In today's world, technology plays a vital role for all people in this world. The technology provides easy access to information, making tasks more manageable, communicating with each other efficiently, and much more. Without technology, we may not have experienced the current life with the advance of technology. Moreover, internet access has a significant impact on society. People can acquire all the resources and information, thus enabling them to engage with others only through the Internet. The most important thing in technology is communication. How people can interact and socialize with other people all around the world. The technology provides social networking such as WhatsApp and Telegram, and people can do video conferencing or FaceTime with someone on the other side of the world (Allen, 2019).

According to Huang and Chueh (2021), using the chatbot for communicating is more accessible as it is a built-in instant messaging application. The benefits are that chatbots can be used as an instant messaging program and can be integrated into everyday communication settings. Chatbot is highly being used by people in their daily life nowadays (Badlani et al., 2021), as the advancement of technology allow for Chatbots to learn from conversation and knowledge databases to integrate with models for predicting new message inputs and generate relevant message outputs (Suta et al., 2020).

Subsequently, this research aims to develop chatbots for veterinary consultations related to cats as pets. As every pet owner desires their pet cats to get a healthy life, body check-ups are essential, especially for pets with health complications that require more attention than other normal, healthy pets. Based on List Wire (2021), visiting a veterinarian could be stressful, inconvenient, and expensive due to overbooked appointment schedules, long waiting times,

especially for emergency cases, and ad-hoc visits. Additionally, consultations could be costly, especially if the case is complicated. With the help of online consultations through Chatbots, pet owners can opt for fewer and shorter consultations with veterinarians and save monetary cost. Hence, in this study, a chatbot prototype (MikoBot) was developed with the aim of being able to produce the necessary standard information regarding pet cats that the user chooses to view.

2. METHODOLOGY

Six phases were involved in the chatbot development: preliminary study, data acquisition, design and implementation, development, testing, and documentation. These phases were completed in order to achieve the three objectives aimed for the study. The first objective was to determine the list of typical cats' illnesses that can be self-treated by pet owners. The objective was achieved by gathering related information via a literature study based on relevant journals and articles on veterinary consultations, where the problem and background of the study were specified. Information was also collected through surveys with respondents from pet cat owners. The chatbot's development to achieve the study's second and third objectives was completed using Javascript, HTML, and CSS coding. Subsequently, the MikoBot prototype was developed, and the Google Firebase platform was used to link with the real-time database.

Table 1. Survey on Pet Owners – Feline (Cat) Information.

No	Feline (Cat) Information
1	Do you have your own pet cat?
2	Do you take care of your cat in the house or outside of the house?
3	How many cats do you have?
4	Has your cat ever been sick?
5	If yes, please state what kind of illnesses that your cat faced?
6	How did you treat your cat's illnesses?
7	If you chose to go for home treatment, what kind of illnesses can be treated and monitored by yourself without having to go see a veterinarian?
8	If you choose home treatment, please state one (1) or more treatments that you have tried before?
9	Have you sought out information online about cats before?
10	What kind of information about cats did you look for?
11	Do you have any veterinary clinics near your place?

3. RESULTS AND DISCUSSION

Figure 1 shows the architecture of the developed MikoBot prototype, while the flow of a simple startup conversation is shown in Figure 2. The result of the conversation is shown in Figure 3 and Figure 4, where it shows how MikoBot answered questions from a user. In Figure 4, the user stated a few symptoms of the cat's illness and how MikoBot can determine the cat's illnesses.

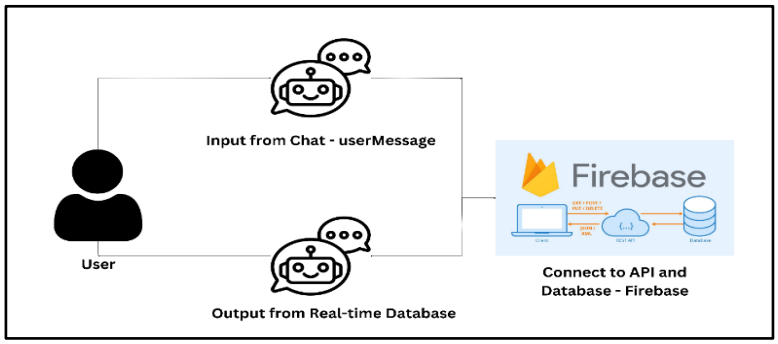

Figure1. System Architecture

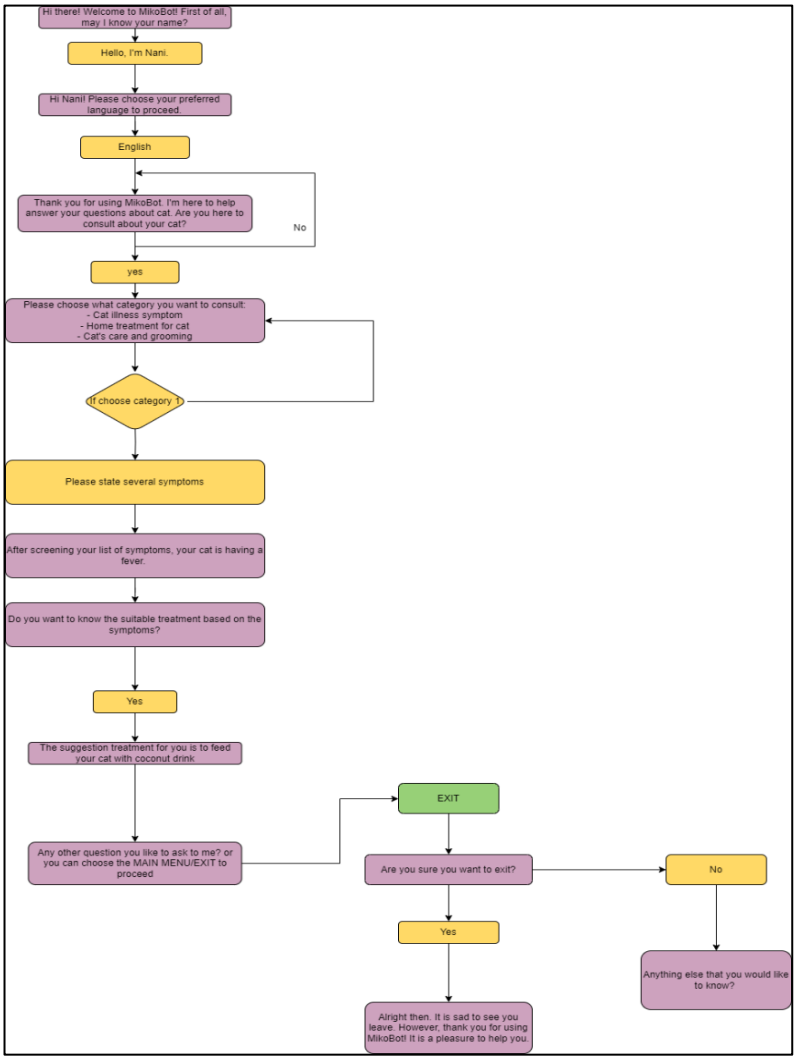

Figure 2. Conversation flow

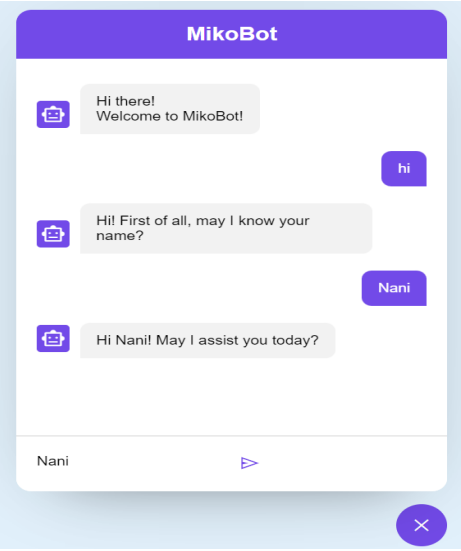

Figure 3. Chatbot prototype.

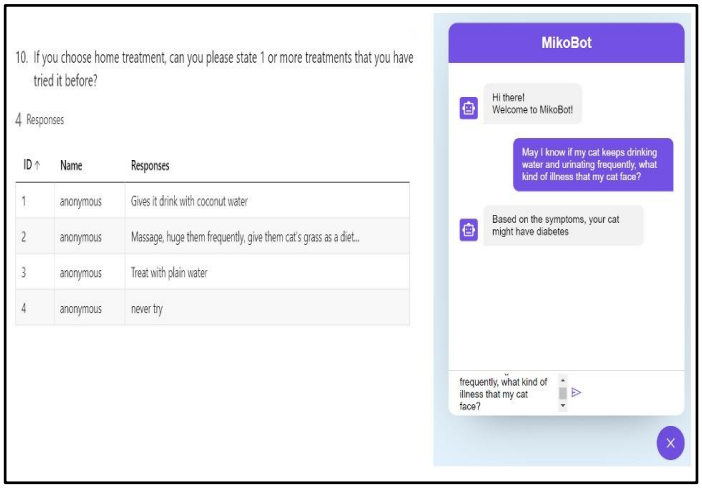

Figure 4. Survey results on home treatment and chatbot's prototype

4. CONCLUSIONS

This research aimed to develop MikoBot, a functional chatbot prototype for pet cats' veterinary consultation. Pet owners would find obtaining necessary, general information about their pet cats easier through MikoBot. This project was scoped only on felines (cats) as pets. This study developed a functional chatbot despite some limitations, such as the limited knowledge of the database and the knowledge being scoped to only cats. Conclusively, this research contributed to the body of knowledge in chatbot technologies and veterinary science studies. Although there were a lot of challenges and unexpected changes during the development process, the MikoBot prototype was successfully developed.

REFERENCES

1. Bharti, U., Bajaj, D., Batra, H., Lalit, S., Lalit, S., & Gangwani, A. (2020, June). Medbot: Conversational artificial intelligence powered chatbot for delivering tele-health after covid-19. In *2020 5th international conference on communication and electronics systems (ICCES)* (pp. 870-875). IEEE.
2. Huang, Duen Huang, and Hao En Chueh. (2021). Chatbot Usage Intention Analysis: Veterinary Consultation. *Journal of Innovation and Knowledge*, 6(3),135–44.
3. Hajzler, I., Nenadović, K., & Vučinić, M. (2023). Health changes of old cats. Journal of Veterinary Behavior, 63, 16–21. https://doi.org/10.1016/j.jveb.2023.04.002
4. Lappin, M. R. (2018). Update on flea and tick associated diseases of cats. *Veterinary Parasitology*, 254, 26–29. https://doi.org/10.1016/j.vetpar.2018.02.022 35
5. K9 Magazine (2022, June 23). How Do Online Veterinary Consultations Work? *K9 Magazine*. https://www.k9magazine.com/online-veterinary-consultations/
6. Mathews, K. A. (2008). Neuropathic pain in dogs and cats: If only they could tell us if they hurt. Veterinary Clinics of North America: Small Animal Practice, 38(6), 1365–1414. https://doi.org/10.1016/j.cvsm.2008.09.001
7. Suta, P., Lan, X., Wu, B., Mongkolnam, P., & Chan, J. H. (2020). An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics Research*, *9*(4), 502-510.

COLLEGE OF COMPUTING, INFORMATICS AND MATHEMATICS
KOMPLEKS AL-KHAWARIZMI
UNIVERSITI TEKNOLOGI MARA
40450 SHAH ALAM
SELANGOR, MALAYSIA

kppimuitmofficial kppim_uitm_official KppimUiTM