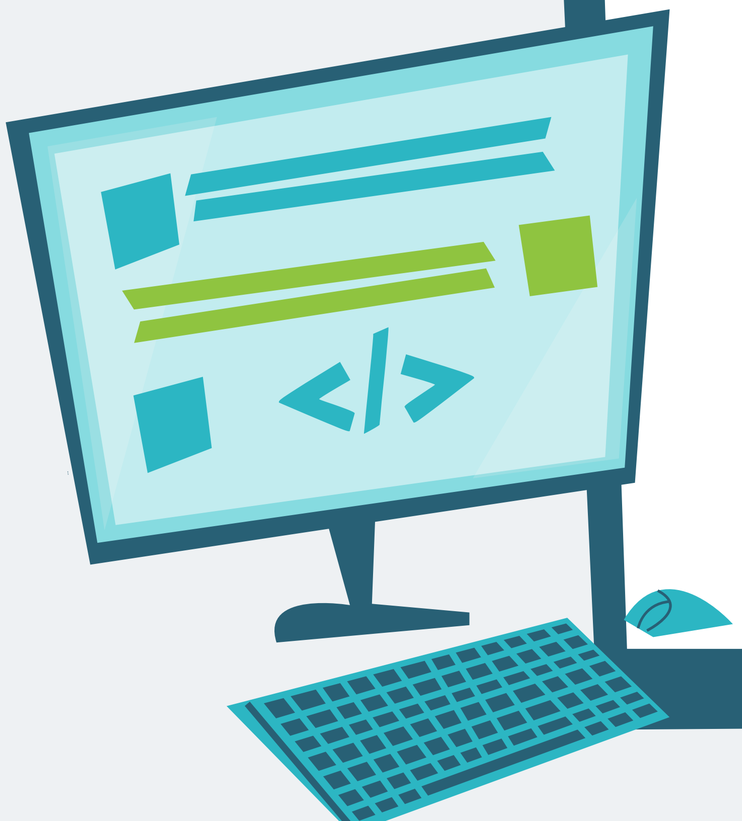
## دعونا نتحدث عن "البيانات مفتوحة المصدر"

تُشير عادةً إلى مجموعات البيانات المتاحة مجانًا للاستخدام العام والتعديل والتوزيع.

غالبًا ما يتم إتاحة الوصول إلى مجموعات البيانات هذه بأقل قدر من القيود، مما يعزز التعاون والابتكار .
فيما يلي بعض المنصات والمستودعات الشائعة للوصول إلى البيانات مفتوحة المصدر:



## 1. مجموعات بيانات KAGGLE:

*KAGGLE* عبارة عن منصة تستضيف مجموعات بيانات لمسابقات التعلم الآلي وعلوم البيانات. ويقدم مجموعة واسعة من مجموعات البيانات في مختلف المجالات التي ساهم بها المجتمع.

## 2. GitHub:

يستضيف *GitHub* العديد من المستودعات التي تحتوي على مجموعات بيانات مفتوحة المصدر. غالبًا ما يشارك المستخدمون مجموعات البيانات المتعلقة بمجالات أو مشاريع بحثية محددة.

Analytics Vidhya

## 3.تحليلات Vidhya:

إنها منصة لعلوم البيانات عبر الإنترنت تقدم مقالات وبرامج تعليمية ودراسات حالة ودورات وشهادات ومسابقات.
وتستضيف مدونة ومنتدى لتعزيز التعلم العملي والتعاون في مجتمع علوم البيانات.

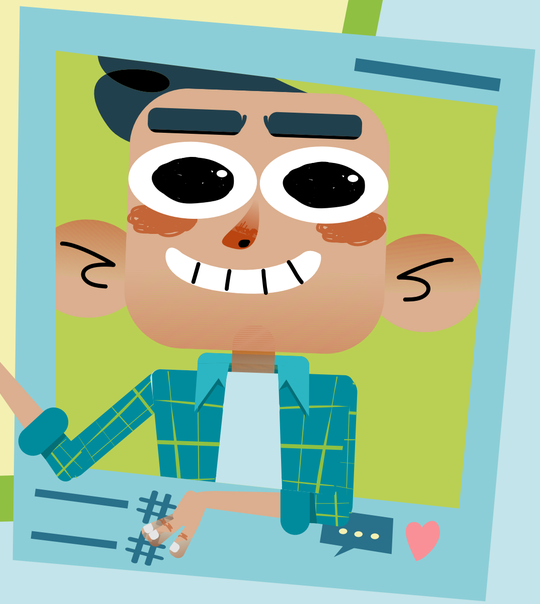
## KDD.4:

يستخرج *KDD* (اكتشاف المعرفة في قواعد البيانات) الرؤى من مجموعات البيانات الكبيرة عبر مراحل مثل الاختيار والمعالجة المسبقة والتحويل واستخراج البيانات والتقييم.
تعمل هذه العملية التكرارية على تحويل البيانات الأولية إلى معرفة قابلة للتنفيذ لاتخاذ قرارات مستنيرة.

# في هذه المرحلة، قد تكون مهتمًا بكيفية جمع هذه البيانات

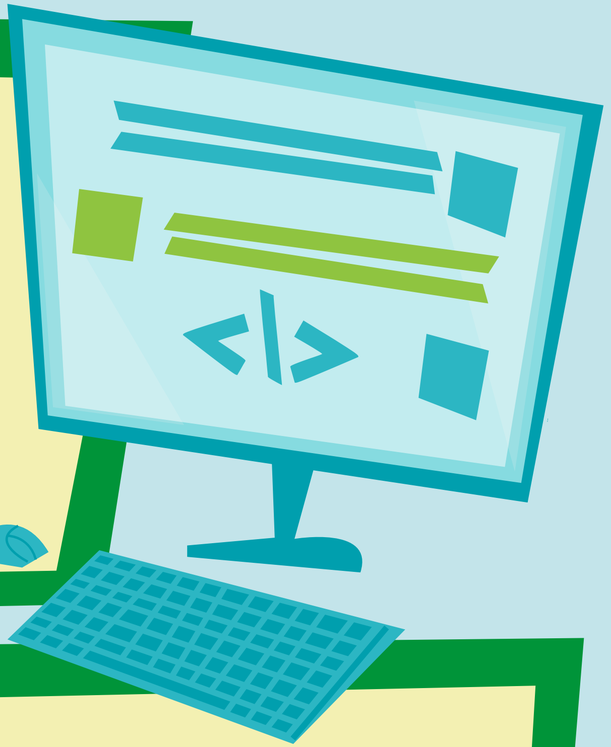
قد يكون تحديد البيانات ذات الصلة بالتحديات التي تواجهك أمرًا صعبًا، وغالبًا ما يتطلب جمع بيانات مستقلة.

فيما يلي بعض الأفكار للمساعدة:



## استخراج البيانات من الانترنت Web scraping:

حدد موقع لاستخراج المحتوى، واستخرج محتوى من صفحة الويب، واحفظ البيانات *HTML* بالتنسيق المفضل لديك.



## الاستطلاعات Surveys :

الاستطلاعات عبر الإنترنت، والاستطلاعات الهاتفية، والمقابلات الشخصية.

## التوظيف الجماعي Crowdsourcing:

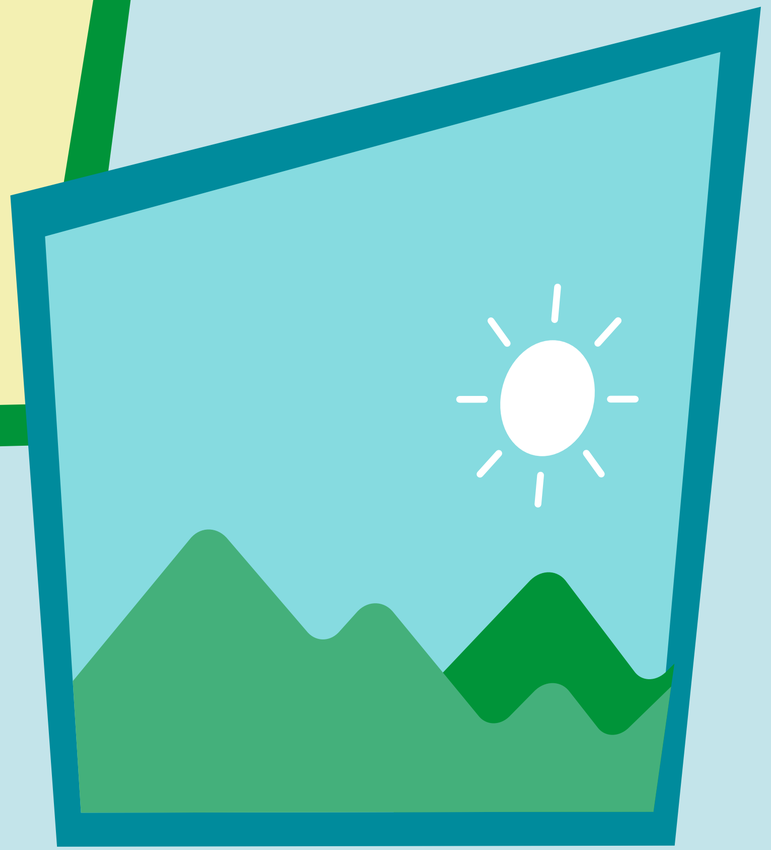
يستفيد من الأفراد الذين يشتركون في مصالح مشتركة لجمع البيانات، حيث يجمع بين أصحاب العمل الحر والمتطوعين. هذه الطريقة فعّالة من حيث التكلفة تيسر العمليات، مما يوفر الوقت والنفقات للشركات.

## زيادة البيانات Data augmentation:

زيادة تنوع البيانات المتاحة للتدريب بشكل كبير دون جمع بيانات جديدة بشكل فعلي. تنطبق زيادة البيانات أيضا على أنواع أخرى من البيانات.

## يتم إنشاء البيانات الاصطناعية Synthetic data:

المرئية في المقام الأول، برمجيًا باستخدام محركات العرض التي تولد الصور والتعليقات التوضيحية. تعتبر هذه البيانات المرنة والقابلة للتطوير ذات قيمة لتدريب نماذج التعلم الآلي ومحاكاة السيناريوهات المختلفة.



# دولاب موازنة البيانات
# :Data flywheel

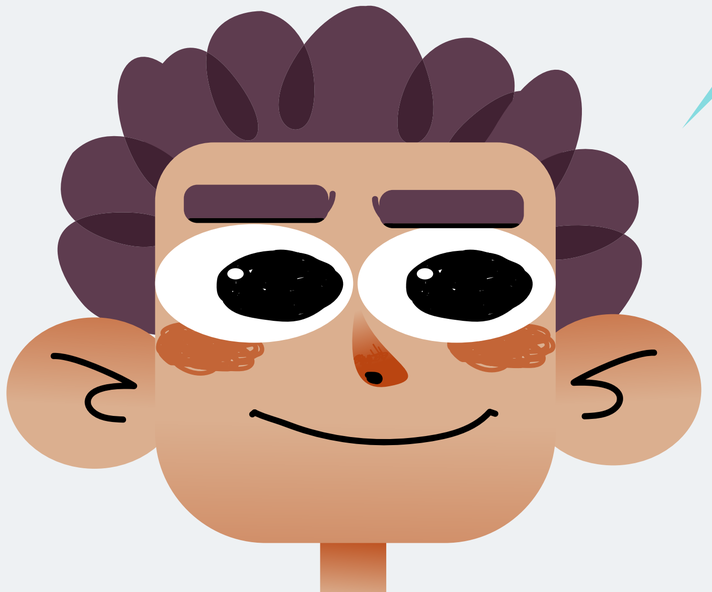
مفهوم الدائرة البيانية مثير للاهتمام، حيث يشدد على عملية دورية

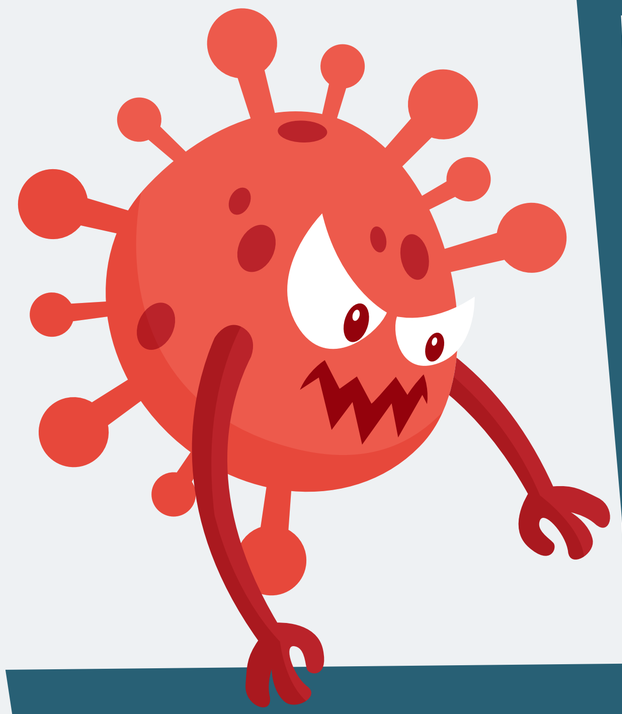
اعرض نموذجك أمام المستخدمين ← اجمع المزيد من البيانات ← قم بتحسين النموذج ← تقديم منتج أفضل للمستخدمين ← الحصول على المزيد من المستخدمين

تتيح هذه الحلقة التكرارية التحسين السريع، مما يعزز الدورة المستمرة التي تعمل على تسريع تطوير المنتج وتسهيل التحول المُبكِّر للمنتج

في عصر اليوم الغني بالبيانات، يعد إعطاء الأولوية لخصوصية البيانات لكلٍ من المؤسسات والأفراد أمراً ضرورياً.

وهذا يعني التعامل مع البيانات الشخصية ومعالجتها وحمايتها بشكل مناسب لاحترام الحقوق الفردية عبر الإنترنت.





من الضروري أيضًا أن تظل يقظًا ضد التهديدات المتطورة مثل الوصول غير المصرح به، والهجمات الإلكترونية، والانتهاكات، ورسائل التصيد الاحتيالي، والبرامج الضارة، التي تعرض سرية البيانات وأمانها للخطر.

يعد تنفيذ أفضل الممارسات لخصوصية البيانات أمرًا ضروريًا لحماية المعلومات الحساسة والحفاظ على الثقة مع أصحاب المصلحة.

# تشمل الممارسات الرئيسية ما يلي:

## كلمات المرور القوية

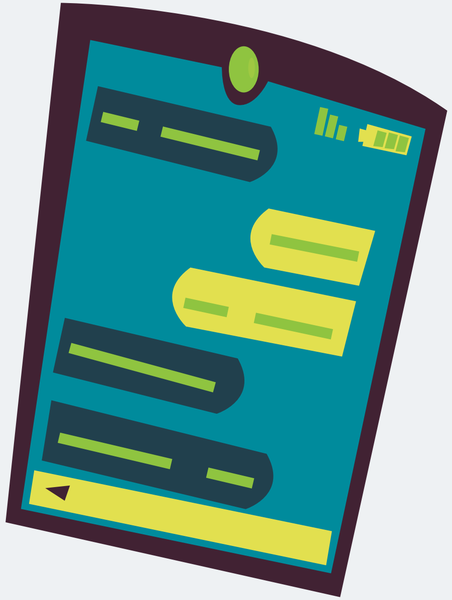
تعزيز استخدام كلمات مرور قوية وفريدة مكونة من أحرف وأرقام ورموز، وتحديثها بانتظام لتحسين الأمان.


## التحكم في مشاركة البيانات

تنفيذ ضوابط وصول صارمة لتقييد مشاركة البيانات على الموظفين المصرح لهم، بناءً على الأدوار والمسؤوليات الوظيفية المحددة.

## المصادقة الثنائية(2FA)

لطبقة أمان إضافية، تتطلب من المستخدمين (2FA) تمكين التحقق من هويتهم بعامل ثان، مثل رمز من تطبيق أو رسالة نصية، إلى جانب كلمة المرور الخاصة بهم.



## تحديثات البرنامج

قم بتحديث جميع البرامج بانتظام، بما في ذلك أنظمة التشغيل وبرامج مكافحة الفيروسات والتطبيقات، لتصحيح نقاط الضعف والحماية من التهديدات الأمنية المحتملة.

## احتياطات محاولات التصيد الاحتيالي

قم بتثقيف المستخدمين حول مخاطر التصيد الاحتيالي وتشجيع الشك تجاه رسائل البريد الإلكتروني أو الرسائل غير المرغوب فيها. تنفيذ حلول تصفية البريد الإلكتروني لاكتشاف محاولات التصيد الاحتيالي وحظرها.


## الشبكات الآمنة فقط

العامة *Wi-Fi* اتصل بشبكات آمنة وموثوقة، وتجنب شبكات للأنشطة الحساسة. استخدم الشبكات الخاصة الافتراضية عند الوصول إلى البيانات عن بعد لتشفير الاتصالات وتعزيز (*VPN*) الخصوصية.

من خلال دمج هذه الاحتياطات في ممارسات الأمن السيبراني الخاصة بك، يمكنك تحسين الوضع الأمني لكل من المستخدمين الفرديين والأنظمة التنظيمية بشكل كبير.