# PCA

## PCA (Principal Component Analysis) : او تحليل المكونات الرئيسية

هو تقنية إحصائية تُستخدم لتبسيط مجموعة بيانات معقدة متعددة الأبعاد عن طريق تحويلها إلى مجموعة بيانات ذات أبعاد أقل مع الاحتفاظ بأكبر قدر ممكن من التباين في البيانات الأصلية والتعرف على الأنماط، لاختزال الأبعاد واستخراج الميزات

في مجالات مثل التعلم الآلي، وتحليل البيانات، والرؤية الحاسوبية PCA يستخدم

**الأهداف الرئيسية لـ ☻ PCA:**

**اختزال الأبعاد .1 (Dimensionality Reduction):** تبسيط مجموعة البيانات عن طريق تقليل عدد المتغيرات (الأبعاد) مع الحفاظ على معظم المعلومات الأصلية .

**استخراج الميزات .2 (Feature Extraction):** استخراج الميزات الرئيسية التي تساهم بشكل كبير في التباين في البيانات .

**مزايا PCA:**

* **تبسيط البيانات :** تقليل الأبعاد مع الحفاظ على معظم المعلومات الأصلية.
* **تقليل الضوضاء :** إزالة الأبعاد التي تحتوي على ضوضاء أو معلومات غير مهمة .
* **زيادة الكفاءة :**  تحسين كفاءة الخوارزميات بالتعامل مع بيانات أقل تعقيدًا.

باختصار، PCA هو أداة قوية لاختزال الأبعاد واستخراج الميزات من البيانات المعقدة، مما يساعد في تحليل البيانات وتحسين أداء الخوارزميات المختلفة. A close-up of a computer screen

Description automatically generated

# Cross validation-:

، أو التحقق المتقاطع، هو أسلوب إحصائي يُستخدم لتقييم أداء النموذج التنبؤي وضمان صلاحيته وفعاليته. يُستخدم هذا الأسلوب لتجنب المشاكل التي قد تنشأ من تقسيم البيانات إلى مجموعات تدريب واختبار غير ممثلة بشكل جيد، مما يمكن أن يؤدي إلى تحيز في التقييم.

# انواعه

## **K-Fold Cross-Validation:**

هو تقنية تُستخدم لتقييم أداء نموذج التعلم الآلي بطريقة أكثر دقة وموثوقية. الفكرة الرئيسية هي تقسيم البيانات إلى عدة مجموعات ، واستخدام كل مجموعة كبيانات اختبار مرة واحدة فقط، بينما تُستخدم المجموعات الأخرى لتدريب النموذج وهكذا مع كل مجموعه حيث كل مجموعه لها دورها بأن تكون لأجل ال تيست والباقي لأجل الترين

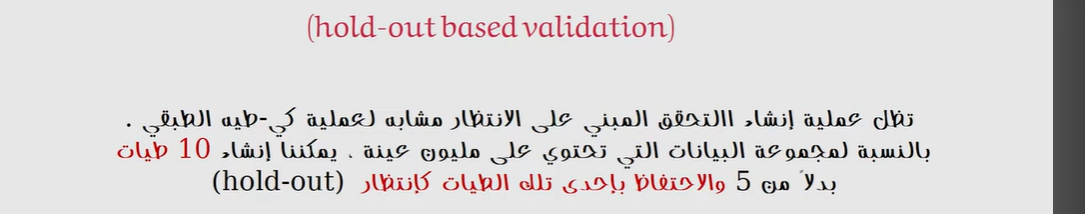
A screenshot of a computer

Description automatically generated

##  hold-Out Cross Validation (LOOCV):

* حالة خاصة من k-Fold Cross Validation حيث k يساوي حجم العينة (عدد البيانات).
* في كل تجربة، يتم استخدام كل البيانات ما عدا واحدة لتدريب النموذج، ويتم اختبار النموذج على البيانات المتبقية.
* يتم تكرار هذه العملية لكل عينة من البيانات.





##  Stratified k-Fold Cross Validation:

* يشبه k-Fold Cross Validation، لكن يتم الحفاظ على نسبة التصنيفات المختلفة في كل طية (fold) مماثلة للنسبة الأصلية في مجموعة البيانات.
* يُستخدم عادة عندما تكون البيانات غير متوازنة.



## Time Series Cross Validation:

* + يُستخدم مع بيانات السلاسل الزمنية حيث الترتيب الزمني مهم.
  + يتم تقسيم البيانات إلى أجزاء بحيث يتم تدريب النموذج على البيانات السابقة واختباره على البيانات المستقبلية.

### لماذا يُستخدم Cross Validation؟

1. **تقييم الأداء**:
   * يساعد على تقدير دقة النموذج على بيانات غير مرئية.
   * يعطي فكرة عن كيفية أداء النموذج على بيانات جديدة.
2. **تجنب Overfitting**:
   * يساعد على التأكد من أن النموذج لا يتعلم فقط الأنماط الخاصة بمجموعة التدريب، بل يعمم بشكل جيد على بيانات جديدة.
3. **اختيار النموذج**:
   * يمكن استخدامه لمقارنة أداء نماذج مختلفة أو معلمات مختلفة لنفس النموذج.

باستخدام Cross Validation، يمكن للباحثين والمحللين ضمان بناء نماذج قوية وموثوقة قادرة على التعميم والتنبؤ بشكل جيد على بيانات جديدة.