

Nama : Fahmi Adhiwangsa

NIM : 1103204142

Homework

- StatQuests ([PCA](#), [LDA](#), [SVD](#))
- Learn [PCA](#), [LDA](#) and [SVD](#) in ScikitLearn
- Learn Pandas Framework

PCA (Principal Component Analysis) adalah suatu metode yang digunakan dalam bidang statistika dan data *science* untuk mengekstraksi fitur penting dari suatu data dengan cara mencari variansi terbesar dari data tersebut, kemudian menemukan fitur baru yang saling tidak terkorrelasi dengan satu sama lain. Setelah itu, fitur-fitur tersebut disebut dengan PC (*Principal Component*) atau komponen utama, dan dapat digunakan untuk melakukan reduksi dimensi data atau untuk meningkatkan akurasi dari suatu model *machine learning*. PCA biasanya digunakan untuk menangani data yang memiliki banyak fitur dan dimensi, seperti data citra atau data sinyal. Hasil akhir dari PCA adalah kita bisa mengetahui faktor apa saja yang paling berperan dalam menjelaskan fenomena yang ada di dataset dengan tetap mempertahankan karakteristik data (*maximum variance*). PCA juga dapat mengurangi *overfitting* pada suatu model *machine learning* dan mempercepat proses pelatihan suatu model.

LDA (Linear Discriminant Analysis) adalah suatu metode dalam *machine learning* yang digunakan untuk melakukan klasifikasi pada data dengan mengurangi dimensi data tersebut. LDA merupakan teknik reduksi dimensi yang digunakan sebagai tahap pra-pemrosesan pada *machine learning* dan *pattern recognition*. Berikut adalah beberapa informasi mengenai LDA:

- LDA digunakan untuk memproyeksikan data ke dalam ruang dimensi yang lebih rendah, sehingga memudahkan dalam melakukan klasifikasi data.
- LDA bekerja dengan mencari linear *combination* dari fitur-fitur yang ada pada data yang dapat membedakan antara kelas-kelas yang ada pada data tersebut.
- LDA mengasumsikan bahwa data yang digunakan memiliki distribusi normal dan memiliki *kovarians* yang sama antar kelas.
- LDA dapat digunakan untuk mengurangi dimensi data dengan memilih fitur-fitur yang paling relevan dalam membedakan antar kelas pada data tersebut.

- LDA dapat digunakan pada data dengan jumlah fitur yang lebih besar daripada jumlah sampel.
- LDA dapat digunakan pada data dengan lebih dari dua kelas.

SDV (Synthetic Data Vault) adalah suatu metode dalam *machine learning* yang digunakan untuk membuat data sintetis yang mirip dengan data asli. SDV digunakan untuk mengatasi masalah privasi dan keamanan data, di mana data asli tidak dapat digunakan untuk tujuan tertentu karena mengandung informasi sensitif atau rahasia.