

# Learning Progres Review Week #12

*By Optimistic team*



## Session 34

# Regression



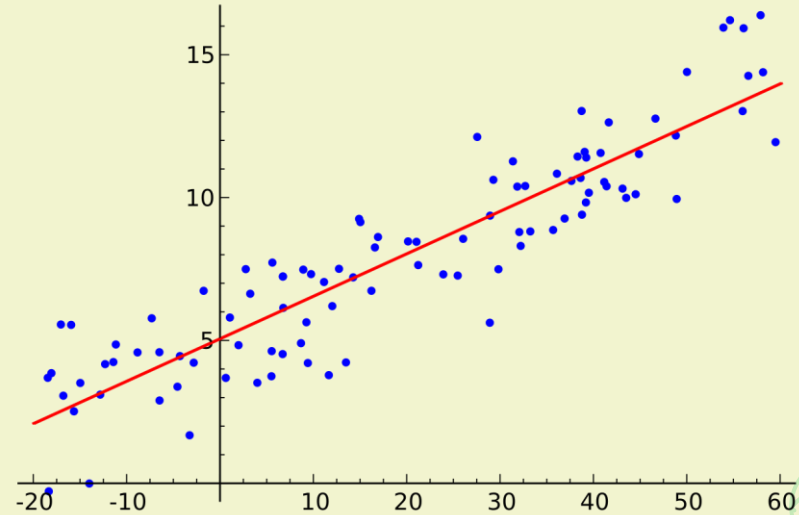
# Regression

Regresi adalah model yang menyatakan hubungan antara variabel terikat (yang menjadi target) dan variabel bebas (yang menjadi predictor) dengan satu atau lebih variabel bebas. Dimana biasanya variabel terikat dilambangkan dengan  $Y$ , dan variabel bebas dilambangkan dengan  $X$ . Contoh proyek yang menggunakan model *Regression* adalah memprediksi harga rumah.



# Linear Regression

*Linear regression* adalah salah satu model regression yang merupakan model yang paling sederhana yang sering digunakan untuk melakukan pemodelan *regression*. *Linear Regression* terbagi menjadi dua, yaitu *Simple Linear Regression*, dan *Multiple Linear Regression*. Bedanya *simple linear regression* memiliki 1 variabel bebas, sedangkan *multiple linear regression* memiliki lebih dari 1 variabel bebas.



# Linear Regression- Regularized

## 1. Lasso Regression

*Least Absolute Shrinkage and Selection Operator* atau yang biasa disebut dengan *Lasso Regression* adalah metode regresi yang melibatkan pembatasan ukuran absolut dari koefisien regresi.

```
from sklearn.linear_model import Lasso

lasso_model = Lasso().fit(X_train,y_train)

y_train_pred = lasso_model.predict(X_train)
y_test_pred = lasso_model.predict(X_test)
```

# Linear Regression- Regularized

## 2. Ridge Regression

*Ridge regression* merupakan modifikasi dari metode kuadrat terkecil yang menghasilkan penduga bias dari parameter regresi, *ridge regression* digunakan untuk mengatasi masalah multikolinearitas.

```
from sklearn.linear_model import Ridge

ridge_model = Ridge().fit(X_train,y_train)

y_train_pred = ridge_model.predict(X_train)
y_test_pred = ridge_model.predict(X_test)
```

# Linear Regression- Regularized

## 3. Elastic-Net Regression

*Elastic-Net Regression* merupakan metode seleksi variabel yang inovatif untuk estimasi regresi linear. *Elastic-Net Regression* memiliki tujuan untuk menemukan koefisien yang meminimalkan jumlah kuadrat kesalahan dengan menerapkan penalti pada koefisien.

```
from sklearn.linear_model import ElasticNet

elastic_net_model = ElasticNet(alpha=1, l1_ratio=0.5, normalize=False).fit(X_train, y_train)

y_train_pred = elastic_net_model.predict(X_train)
y_test_pred = elastic_net_model.predict(X_test)
```

# Tree-based Regressor

## Decision Tree

*Decision tree* adalah alat pendukung dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas, dan kemungkinan konsekuensi. *Decision tree* menyediakan cara untuk menyajikan algoritma dengan pernyataan kontrol bersyarat.








# Syntax

---

```
from sklearn.tree import DecisionTreeRegressor

dt_regressor = DecisionTreeRegressor().fit(X_train, y_train)

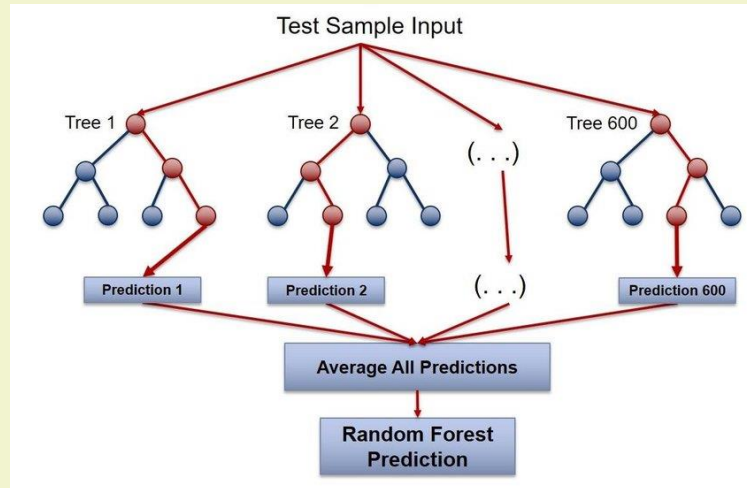
y_train_pred = dt_regressor.predict(X_train)
y_test_pred = dt_regressor.predict(X_test)
```



# Tree-based Regressor

## Random Forest

*Random forest* adalah suatu algoritma yang digunakan pada klasifikasi data dalam jumlah yang besar. Klasifikasi *random forest* dilakukan melalui penggabungan pohon (tree) dengan melakukan training pada sampel data yang dimiliki.






# Syntax

---

```
from sklearn.ensemble import RandomForestRegressor

rf_regressor = RandomForestRegressor().fit(X_train, y_train)

y_train_pred = rf_regressor.predict(X_train)
y_test_pred = rf_regressor.predict(X_test)
```



## Evaluation Metric

Mean squared error

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n e_t^2$$

Root mean squared error

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$$

Mean absolute error

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |e_t|$$

Mean absolute percentage error

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right|$$

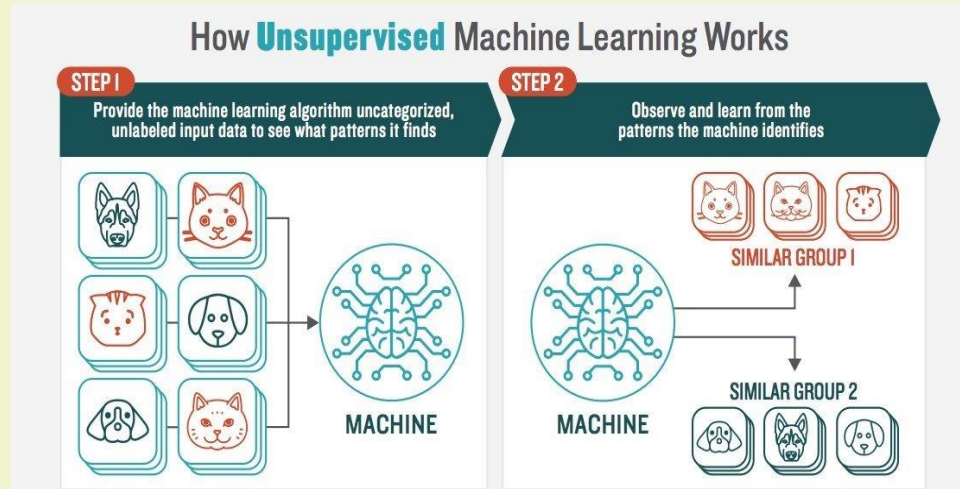
## Session 35

# Unsupervised Learning



# Unsupervised Learning

*Unsupervised Learning* merupakan teknik machine learning dimana kita tidak perlu mengawasi modelnya namun mengizinkan model bekerja untuk menemukan informasi yang dibutuhkan, terutama berkaitan dengan data tanpa label.



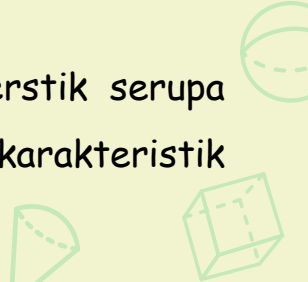


# Karakteristik dan Pengaplikasian

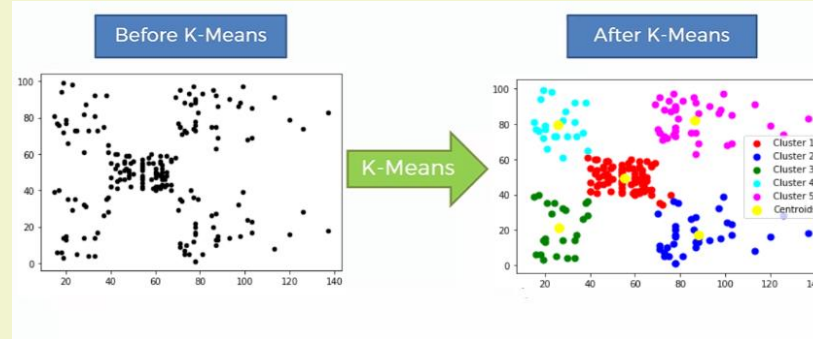
## Karakteristik:

- Dataset tidak memiliki label/target.
- Bertujuan untuk menemukan *pattern* serta membuat model.
- Bukan untuk memprediksi, namun untuk mengelompokkan.

## Pengaplikasian:

- *Customer segmentation*, untuk membagi *customers* menjadi beberapa grup untuk dicari polanya kemudian digunakan untuk kebutuhan *marketing* atau membuat strategi bisnis lainnya.
  - *Anomaly detection*, untuk mendeteksi data termasuk data normal atau *outlier*.
  - *Recommender system*, untuk mengelompokkan *users* yang memiliki karakteristik serupa kemudian *users* akan diberikan rekomendasi suatu produk berdasarkan karakteristik tersebut.
- 

# K-Means Clustering



- Mengelompokkan data yang sifatnya sama ke dalam satu grup.
- Grup direpresentasikan dengan simbol K.

## Cara kerja:

- Menentukan jumlah K cluster.
- Memilih K dari data secara acak sebagai centroid.
- Menentukan objek data yang berdekatan dengan K.
- Update centroid berdasarkan rata-ratanya di setiap cluster.
- Mengulang proses hingga data menjadi konvergen atau tidak ada perubahan lagi ketika proses perulangan.





# K-Means Clustering

Mengevaluasi performa model clustering, dengan cara:

- *Inertia*

Menghitung jarak antara centroid dengan setiap data di cluster yang sama. Rumus:  $\sum (x_i - c_k)^2$   
jika inertia kecil, artinya jarak data dalam satu cluster sangat dekat.

- *Silhouette Score*

$$SC = \frac{b - a}{\max(a, b)}$$

Dimana:

a = rata-rata jarak antar data dalam satu cluster.

b = rata-rata jarak satu cluster dengan cluster lain.





# K-Means Clustering


## Cara menentukan K:

- Menggunakan *Elbow Method* dimana dilakukan dengan mengiterasi nilai random/prediksi K terhadap nilai inertia dan dipilih nilai yang minimum dan berada di titik siku-siku.
- Menggunakan *Silhouette Score* dimana dipilih berdasarkan SC tertinggi.

## Kelebihan K-Means Algorithm:

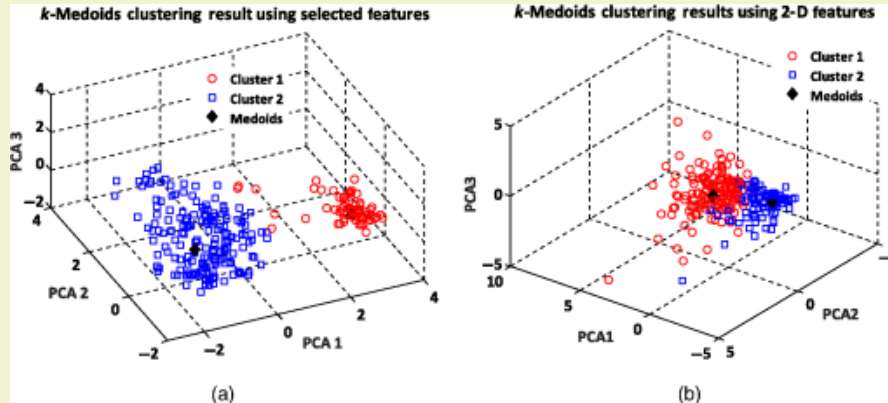
- Relatif simple dan intuitif.
- Tepat digunakan untuk data yang besar.
- Model mudah diadaptasikan untuk data baru.
- Cluster dapat digeneralisasi dengan berbagai bentuk dan ukuran.

## Kekurangan K-Means Algorithm:

- Pemilihan K dengan cara manual.
  - Sensitif terhadap *outlier*.
  - Apabila initial values kurang optimal, hasil K-Means kurang baik.
- 

# K-Medoids Clustering

- Berbeda dengan K-Means Algorithm, K pada K-Medoids Algorithm disebut medoids. Perbedaannya adalah K-Means meminimalisasi jumlah clusters dimana titik tengahnya berupa rata-rata (bukan data asli), sedangkan K-Medoids meminimalisasi jumlah cluster dimana titik tengahnya merupakan data asli.



# Cara Kerja K-Medoids Clustering

- Menentukan jumlah cluster K.
- Memilih medoids awal secara random di setiap cluster.
- Menghitung jarak masing-masing data ke medoids kemudian menandai jarak yang terdekat untuk dihitung totalnya.
- Melakukan iterasi medoids.
- Menghitung total simpangan ( $S$ ). Jika  $a$  adalah jumlah jarak terdekat antara data ke medoids awal dan  $b$  adalah jumlah jarak terdekat antara data ke medoids baru, maka total simpangan adalah  $S = b - a$ . Apabila  $S < 0$ , tukar data medoids dengan data lain untuk dijadikan medoids baru.
- Mengulang proses hingga data menjadi konvergen atau tidak ada perubahan lagi ketika proses perulangan.




## Kelebihan dan Kekurangan K-Medoids Clustering

---

### Kelebihan K-Medoids Algorithm:

- Relatif mudah dipahami dan diimplementasikan.
- Tidak sensitif terhadap *outlier*.
- Konvergen terhadap steps yang tetap.

### Kekurangan K-Medoids Algorithm:

- Jika berbeda pemilihan medoids di awal, maka keseluruhan hasilnya akan berbeda.
  - Tidak direkomendasikan untuk data besar karena proses komputasi yang cukup rumit.
- 

## Session 36

# Communication & Presentation Skill



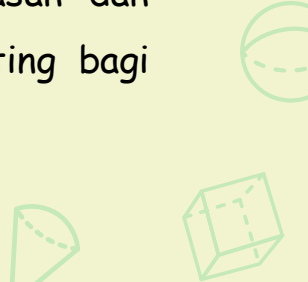


# Konteks

---

Konteks adalah sebuah elemen penting yang membuat ide dan fakta lebih mudah dapat dipahami dan berguna. Konteks yang merupakan elemen dasar dari komunikasi memberikan sebuah tempat bagi orang-orang dan ide untuk bertemu bersama. Jika diterapkan dengan benar, maka konteks akan memberikan ide anda menjadi lebih hidup.

Konteks memberikan audiens fondasi untuk penjelasan dan memberitahukan mereka mengapa ide tersebut penting bagi mereka.



## Kiat mengenali audiens saat presentasi :

Pentingnya mengenali audiens saat presentasi, kita bisa melakukan penyesuaian agar presentasi mampu mempengaruhi audiens dengan efektif. Jika kita melihat beberapa komponen yang saling terkait dalam sebuah presentasi, salah satu kunci penting presentasi adalah audiens. Hal ini karena kita tampil memberikan presentasi dan menciptakan sebuah proses komunikasi, adalah untuk mereka yang menjadi audiens.





## Mengetahui Audiens

1. Banyak audiens yang akan hadir.
2. Latar belakang audiens.
3. Apa yang mereka suka untuk didengar, dan apa yang tidak mereka suka.
4. Tujuan mereka mendengarkan presentasi.



# Konten

- **Prinsip Piramida**

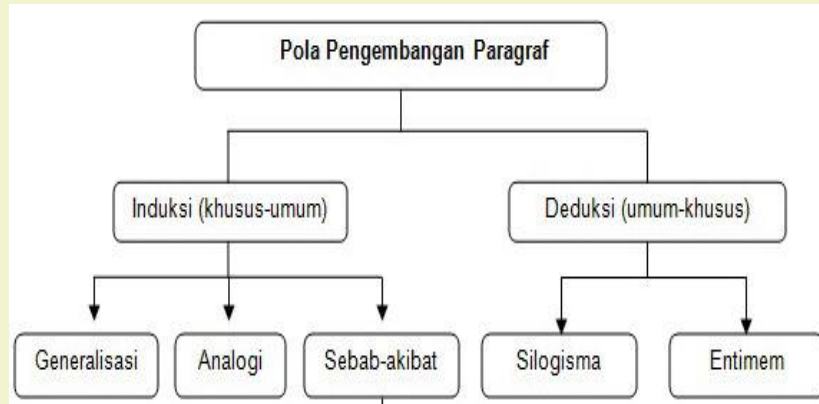
Prinsip Piramida adalah struktur hierarkis untuk membuat logika dan alur cerita yang didukung data. Alur cerita yang didukung oleh data tersebut harus disiapkan terlebih dahulu. Singkatnya, presentasi Anda harus dimulai dengan pengantar yang menyatakan masalah dan jawaban anda, dan seluruh presentasi ada disini untuk mendukung jawaban anda.



## Konten

- **Induktif dan Deduktif**

Tulisan Deduktif merupakan pola pengembangan paragraf berpola umum-khusus, yaitu pernyataan yang bersifat umum diletakkan sebagai kalimat utama yang kemudian diikuti oleh kalimat-kalimat penjelas yang bersifat sebuah pernyataan khusus. Sedangkan Tulisan Induktif merupakan paragraf yang kalimat utamanya berada pada akhir kalimat sehingga pola pengembangan paragrafnya menjadi khusus-umum.






## Konten

---

- **MECE**

Mece adalah sebuah singkatan dari *mutually exclusive collectively exhaustive*, merupakan sebuah prinsip yang digunakan oleh banyak perusahaan manajemen konsultan untuk membantu menggambarkan cara mengatur informasi. Prinsip Mece menyarankan bahwa untuk memahami dan memperbaiki masalah besar apa pun, anda perlu memahami pilihan anda dengan mengurutkannya ke dalam kategori yaitu:

- Saling eksklusif- item hanya dapat masuk ke dalam satu kategori pada satu waktu.
  - Kolektif lengkap - semua item dapat masuk ke dalam satu kategori.
- 

## Design

Tips dalam mendesain presentasi:

1. Menegaskan informasi.
2. Tidak menggunakan 3D charts.
3. Menghindari animasi.
4. Memilih warna yang cocok.



*Thank you!*

## **Our Team**

- 1. Aldiva Wibowo**
- 2. Asprizal Rizky**
- 3. Gilang Rahmat R**
- 4. Lutfia Humairosi**
- 5. Millenia Winadya P**