

zenius

Kampus  
Merdeka  
INDONESIA JAYA

# Final Project Presentation

Nomor Kelompok: 05

Nama Mentor: Rauzan Samara

Nama:

- Muhammad Noorrosyid Sulaksono
- Frischa Indah Prananty

Machine Learning Class

Program Studi Independen Bersertifikat  
Zenius Bersama Kampus Merdeka



- 1. Latar Belakang**
- 2. Explorasi Data dan Visualisasi**
- 3. Modelling**
- 4. Kesimpulan**

# Latar Belakang

# Latar Belakang Project

Sumber Data: [Walmart Dataset](#)

Problem: **Regression**

Tujuan:

- Memprediksi harga Weekly Sales dari Walmart berdasarkan Date, Store, Temperature, CPI, Fuel\_Price, dan Unemployment.
- Menentukan model yang terbaik pada Walmart Dataset.
- Menentukan metrics dari model yang digunakan.

# Explorasi Data dan Visualisasi

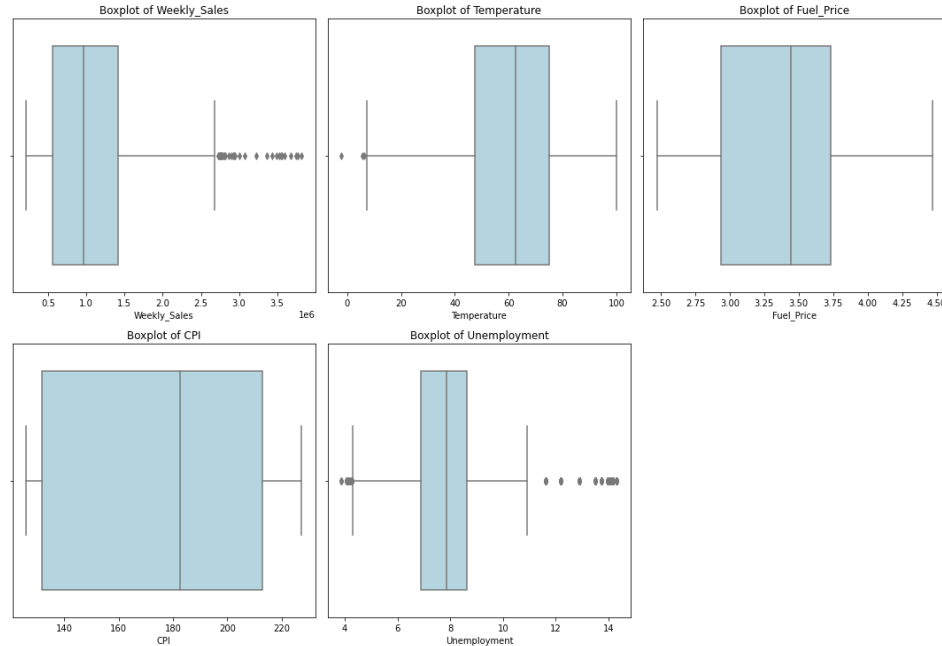
# Business Understanding

Walmart merupakan perusahaan amerika yang mengelola jaringan toserba. Dalam keberjalanannya, hasil penjualan merupakan suatu hal yang sangat penting dalam keberlangsungan bisnis dari toserba.

Dengan banyaknya variasi item yang dijual di Walmart, perlu diperhatikan target konsumen Walmart. Dengan menganalisis target konsumen dari Walmart, dapat diketahui pola dan trend belanja dari para konsumen.

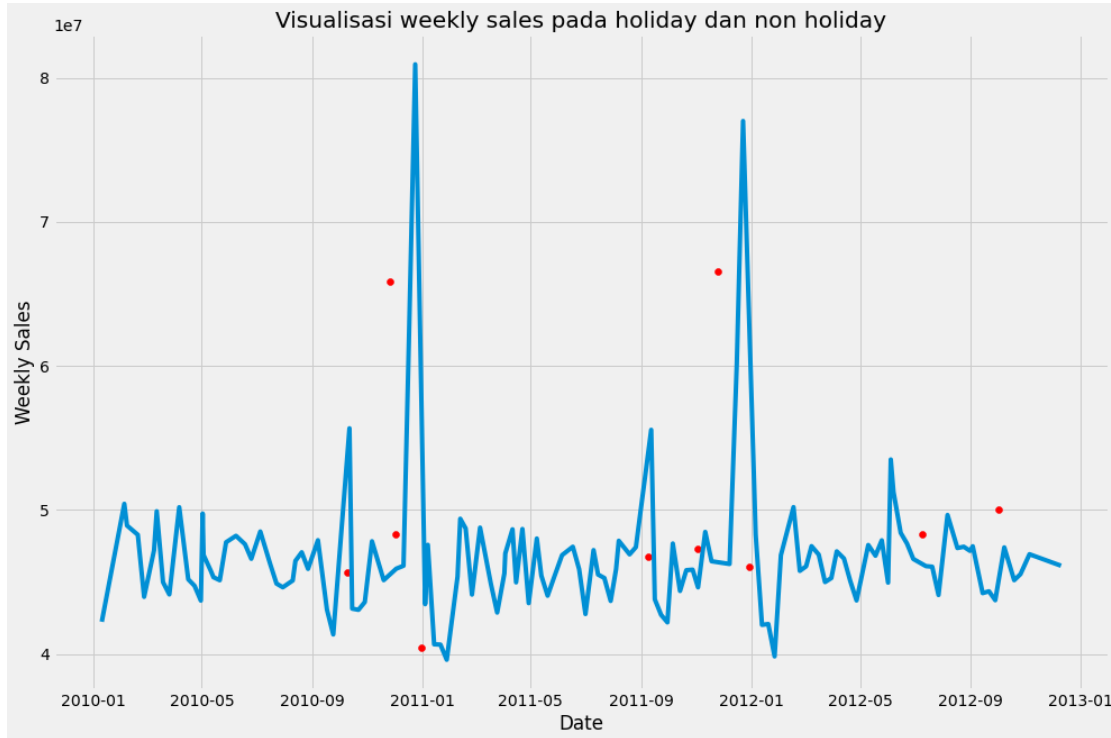
# Data Cleansing

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 6435 entries, 0 to 6434  
Data columns (total 8 columns):  
#   Column      Non-Null Count  Dtype    
---  ---        
0   Store        6435 non-null   int64     
1   Date         6435 non-null   object    
2   Weekly_Sales 6435 non-null   float64   
3   Holiday_Flag 6435 non-null   int64     
4   Temperature  6435 non-null   float64   
5   Fuel_Price   6435 non-null   float64   
6   CPI          6435 non-null   float64   
7   Unemployment 6435 non-null   float64   
dtypes: float64(5), int64(2), object(1)  
memory usage: 402.3+ KB
```



- Tidak terdapat Null Values.
- Tipe Date yang perlu diubah.
- Adanya Outlier pada Weekly Sales, Temperature, dan Unemployment.

# Exploratory Data Analysis



## Sales tertinggi :

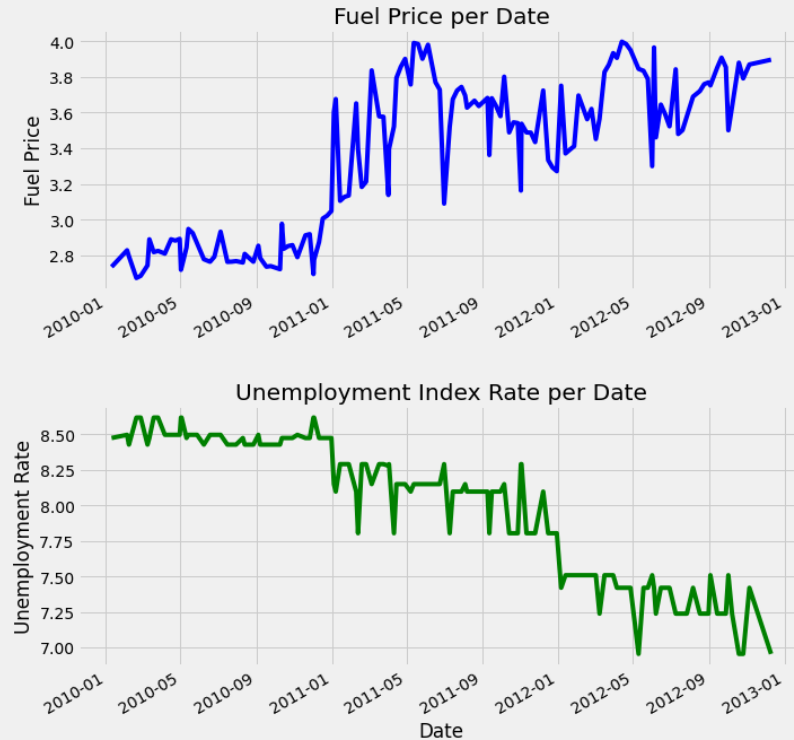
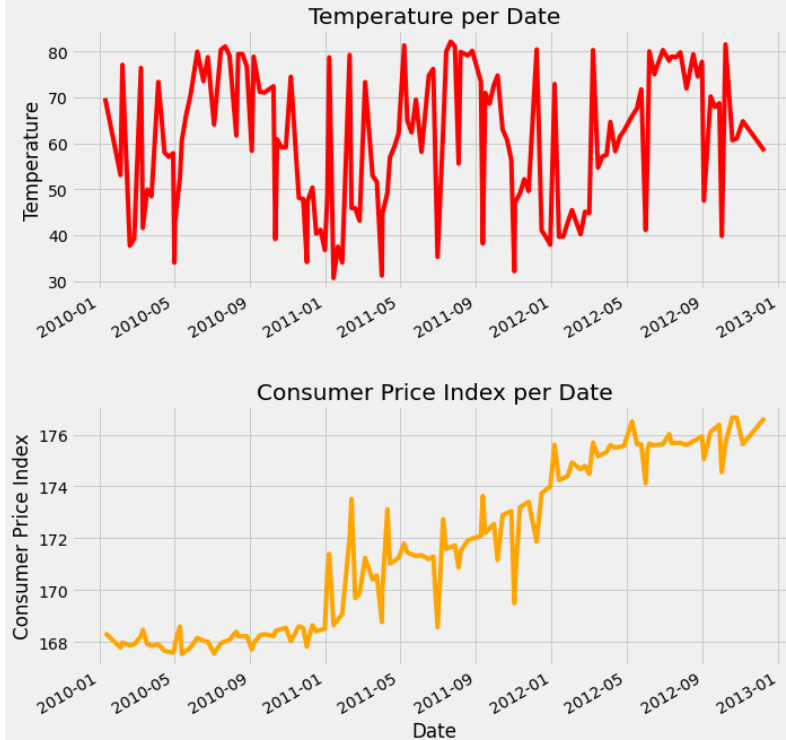
- **Holiday** : 26-Nov-2010 dan 25-Nov-2011 (Thanksgiving)
- **Non Holiday** : 24-Dec-2010 dan 23-Dec-2011

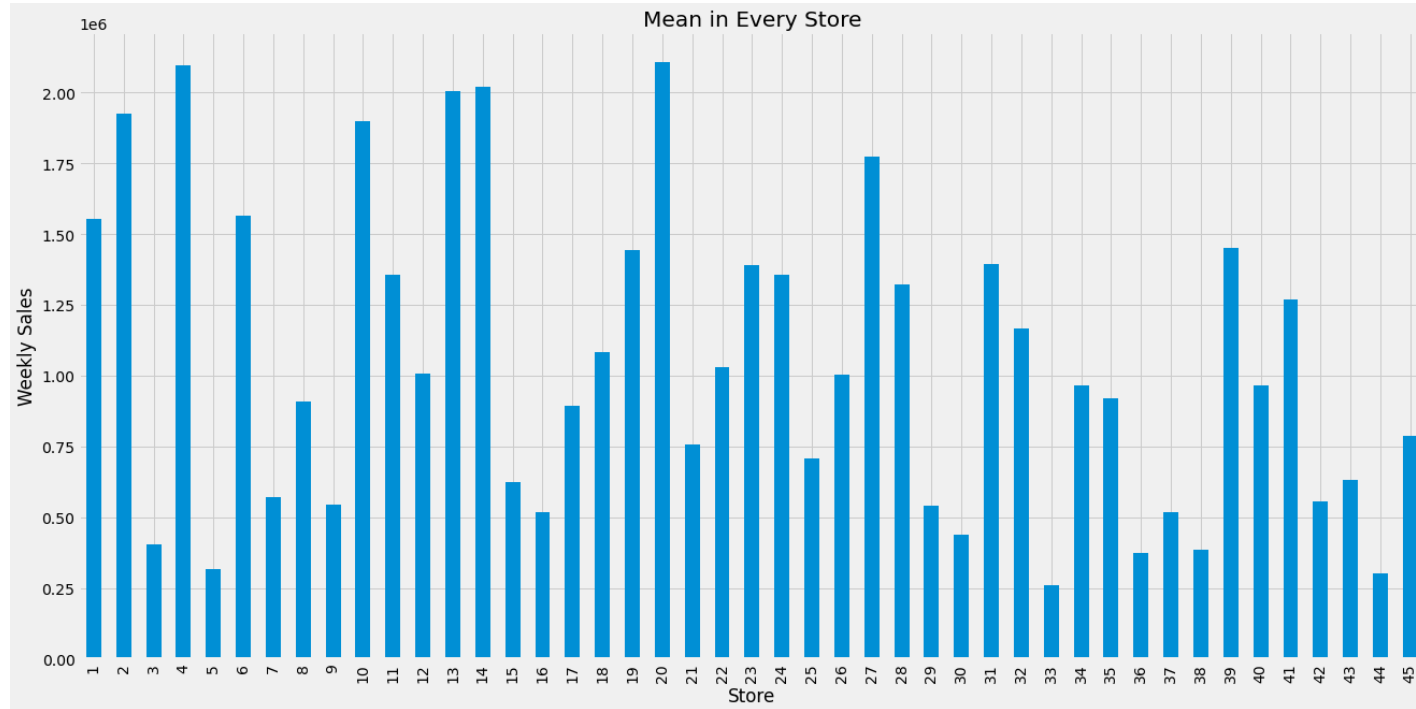
## Sales terendah :

- **Holiday** : 31-Dec-2010 (Christmast)
- **Non Holiday** : 28-Jan-2011 dan 27-Jan-2012

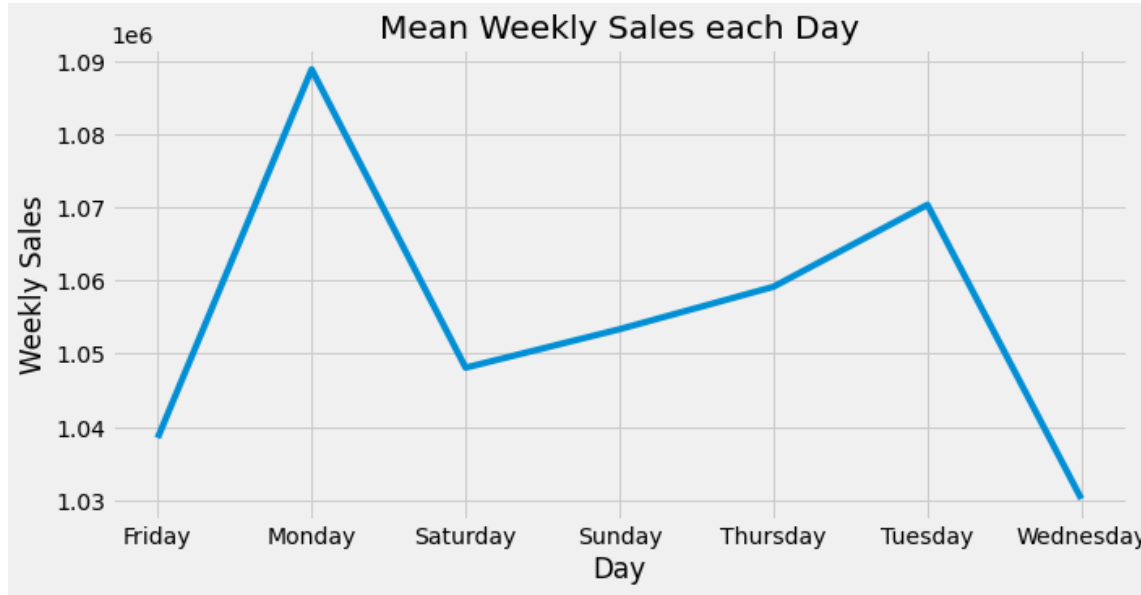


## Visualization of Temperature, Fuel Price, Consumer Price Index, and Unemployment Rate per Date



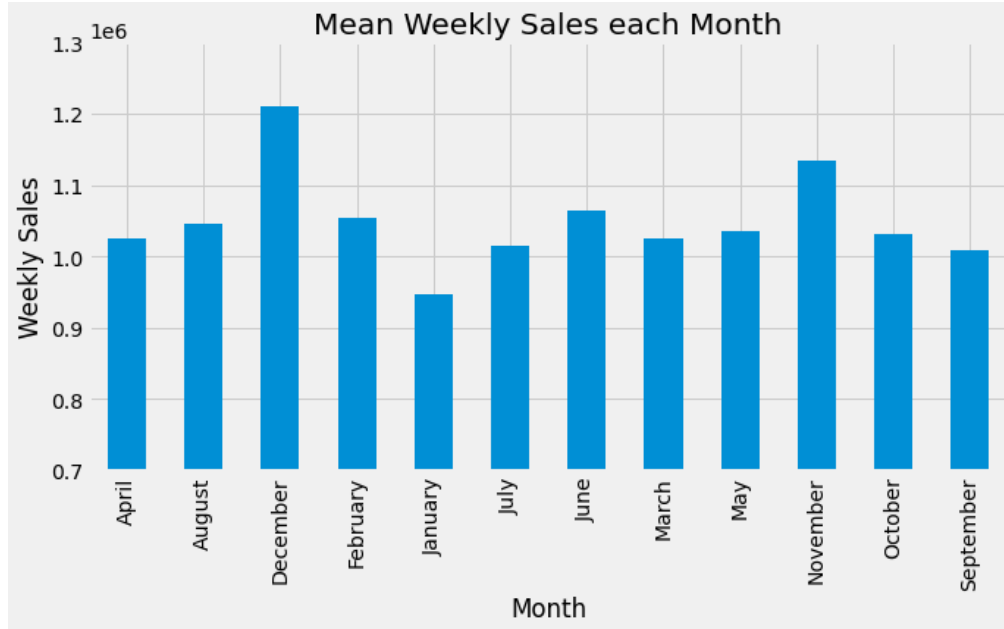


- **Sales Tertinggi** : Store 4 dan Store 20
- **Sales Terendah**: Store 33



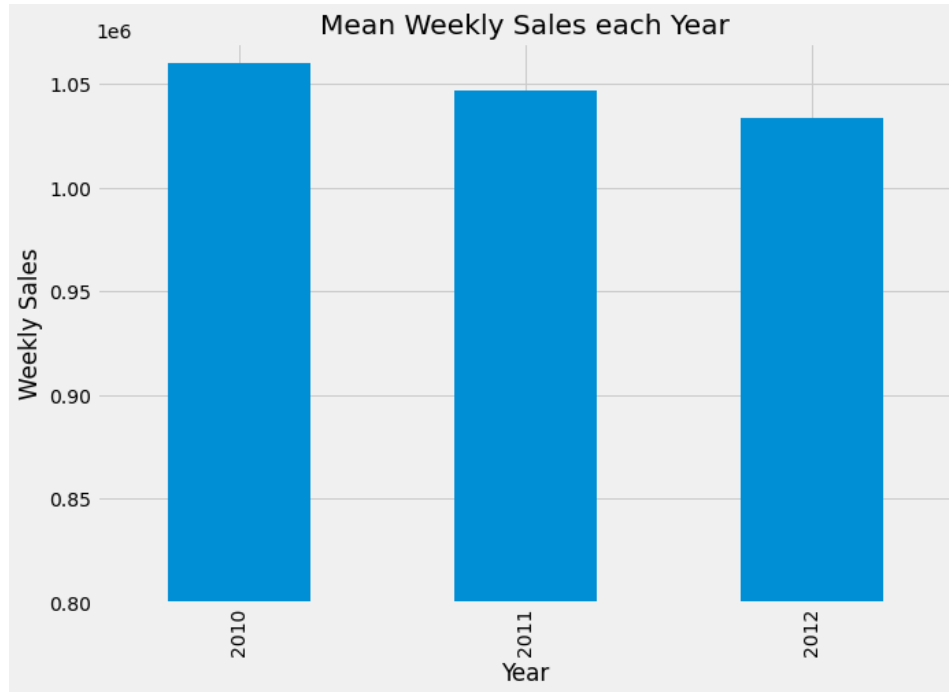
- Rata-rata Penjualan Tertinggi: Senin
- Rata-rata Penjualan Terendah: Rabu

\*Hasil penjualan rata-rata di tiap harinya **TIDAK** memiliki perbedaan yang jauh.



- Rata-rata Penjualan Tertinggi: Desember
- Rata-rata Penjualan Terendah: Januari

\*Hasil penjualan rata-rata di tiap bulannya **TIDAK** memiliki perbedaan yang jauh.



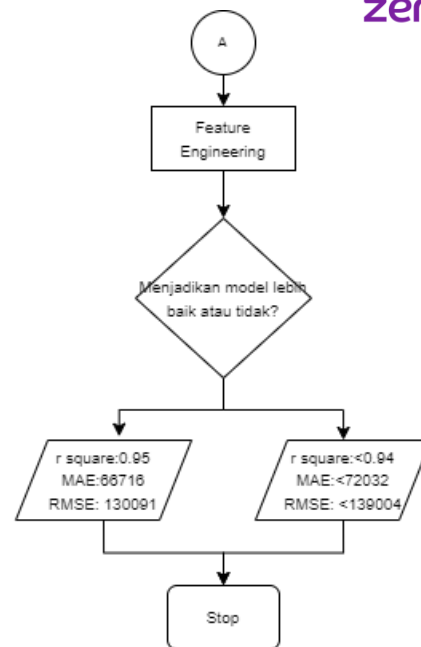
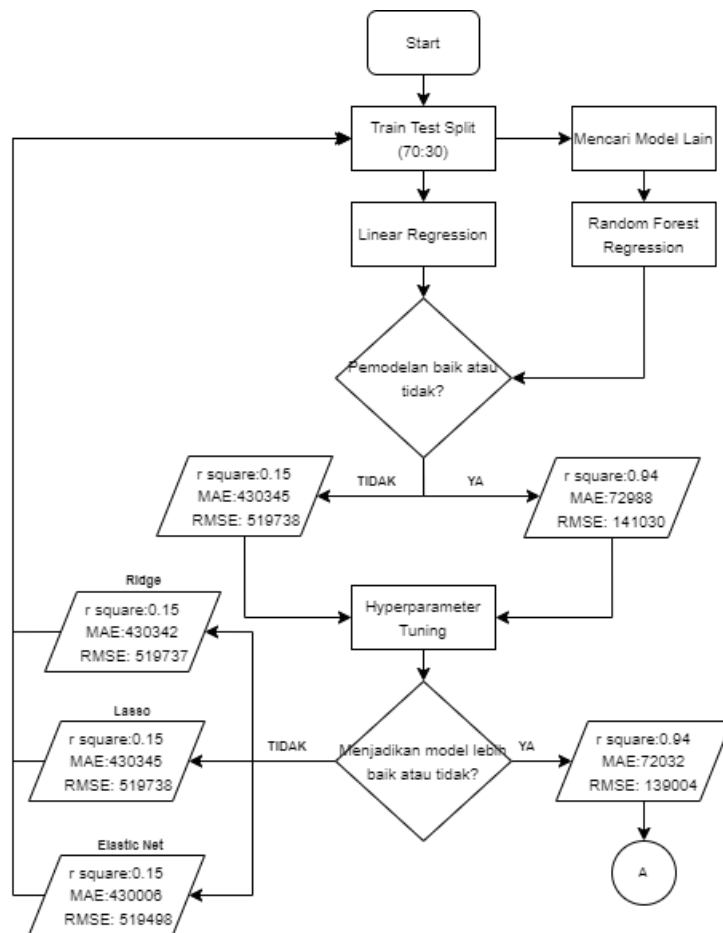
- Rata-rata Penjualan Tertinggi: 2010
- Rata-rata Penjualan Terendah: 2012

\*Hasil penjualan rata-rata di tiap tahunnya **TIDAK** memiliki perbedaan yang jauh.

# Modelling



# Workflow



# Linear Regression

## Metrics :

- r square score: 0.15
- MAE: 430345
- RMSE: 519738
- Test Train Score Error: -11.8 % (Underfitting)



# Hyperparameter Tuning on Linear Regression

Ridge :	Lasso :	Elastic Net :
<ul style="list-style-type: none"><li>- r square score: 0.15</li><li>- MAE: 430342</li><li>- RMSE: 519737</li><li>- Test Train Score Error: -11.8 % (Underfitting)</li></ul>	<ul style="list-style-type: none"><li>- r square score: 0.15</li><li>- MAE: 430345</li><li>- RMSE: 519738</li><li>- Test Train Score Error: -11.8 % (Underfitting)</li></ul>	<ul style="list-style-type: none"><li>- r square score: 0.15</li><li>- MAE: 430006</li><li>- RMSE: 519498</li><li>- Test Train Score Error: -13.5 % (Underfitting)</li></ul>

# Feature Engineering on Linear Regression

**Menggantikan feature Date menjadi Day, Month, dan Year.**

## **Metrics :**

- r square score: 0.16
- MAE: 430241
- RMSE: 517156
- Test Train Score Error: -7.9 % (Underfitting)

# Random Forest Regression

## Metrics :

- r square score: 0.94
- MAE: 72988
- RMSE: 141030
- Test Train Score Error: 5.4% (Baik)

# Hyperparameter Tuning on Random Forest Regression

## Parameter terbaik:

```
{'n_estimators': 733,  
 'min_samples_split': 2,  
 'min_samples_leaf': 2,  
 'max_features': 'auto',  
 'max_depth': 70}
```

## Metrics :

- r square score: 0.94
- MAE: 72032
- RMSE: 139004
- Test Train Score Error: 4.2%  
(Menjadi lebih baik)

Improvement of RMSE: 1.4362279939306402 %

Improvement of MAE: 1.309243511179372 %

# Feature Engineering on Random Forest Regression

Melakukan drop pada feature Date dan Fuel Price dan menggantikannya dengan feature Day, Month, dan Year.

## Metrics :

- r square score: 0.95
- MAE: 66716
- RMSE: 130091
- Test Train Score Error: 3.5% (Menjadi lebih baik)

```
Improvement of Final RMSE: 6.412360127198383 %  
Improvement of Final MAE: 7.379744379321748 %
```

# Conclusion

# Conclusion

Dalam meningkatkan penjualan, perlu diperhatikan beberapa faktor, seperti:

- Hari-hari besar seperti hari thanksgiving, Christmas, dan sebagainya. Selain itu, event libur semester dan masuk semester bagi pelajar juga perlu diperhatikan.
- Consumer Price Index juga dapat menjadi perhatian yang dapat diakibatkan oleh inflasi ataupun penurunan angka Unemployment Rate.
- Metode pemasaran yang dilakukan pada store dengan penjualan terbanyak dapat dijadikan referensi bagi store lain guna meningkatkan penjualan.

**Terima  
kasih!**  
Ada pertanyaan?

**zenius**



**Kampus  
Merdeka**  
INDONESIA JAYA

