



CONSUMPTION ENERGY



Konsumsi energi adalah jumlah energi yang digunakan oleh individu, rumah tangga, industri, atau negara untuk memenuhi suatu kebutuhan dalam melakukan berbagai aktivitas. Kemampuan memprediksi konsumsi tentunya membantu sekali dalam rangka penghematan energi untuk tetap menjaga bumi. Dengan menggunakan algoritma yang paling umum yaitu regresi.

SOLUTION STATEMENTS

Untuk eksplorasi fitur dilakukan

- Analisis Univariat (data kategorik dan numerik)
- Analisis Multivariat.

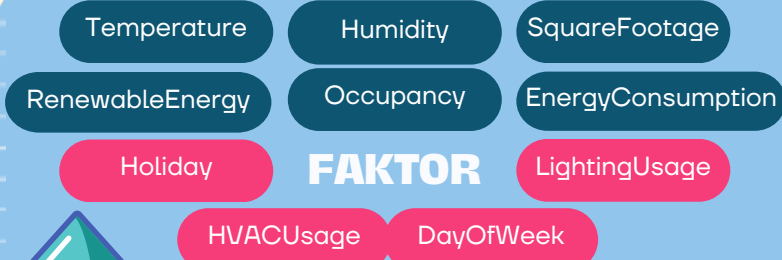
Teknik yang digunakan

- Catplot, pairplot, dan heatmap untuk melihat Correlation Matrix dari fitur-fitur yang dimiliki.

Agar didapatkan model prediksi yang baik maka dilakukan

- Proses Data Wrangling yang meliputi

Data Gathering, Data Assessing, dan Data Cleaning. Untuk mengetahui performa model dilakukan pengecekan performa dengan metrik evaluasi.



Energy-consumption-prediction

Predict the Energy Consumption

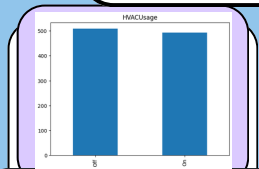
[kaggle.com](https://www.kaggle.com)

- Dataset berupa CSV
- Dataset terdiri dari 1000 records dengan 10 buah fitur yang diukur.
- Dataset terdiri dari 4 data kategori dan 6 data numerik.
- Dataset memiliki missing value sejumlah 0 records

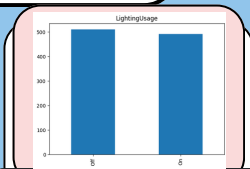
1. Data Analysis

A. ANALISIS UNIVARIATE

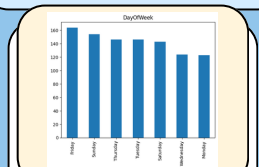
Data Kategori



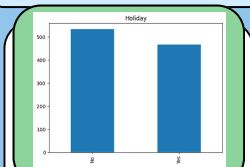
Jumlah sampel 508
Off 492
On 16



Jumlah sampel 509
Off 491
On 18



Jumlah sampel	Persentase
Friday	16.4
Sunday	15.4
Thursday	14.6
Tuesday	14.3
Saturday	12.4
Wednesday	12.3
Monday	12.3



Jumlah sampel 533
No 467
Yes 66

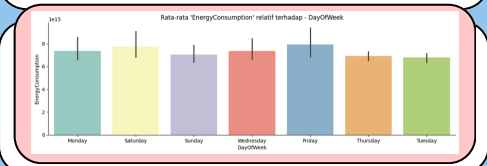
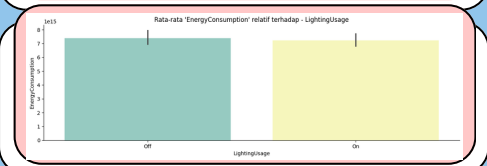
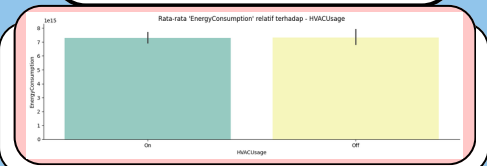
Data Numerik



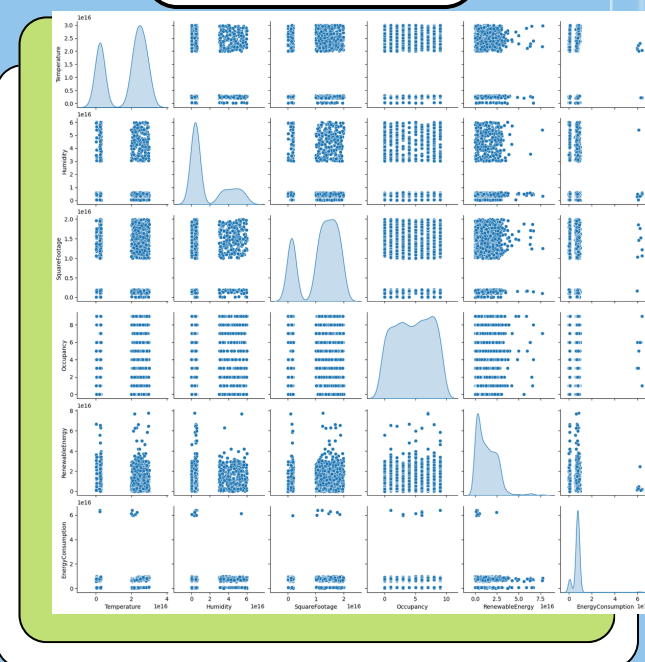
- Temperatur untuk konsumsi energi mayoritas berada pada 25 derajat celsius dan 25 derajat celsius.
- Humidity untuk konsumsi energi mayoritas berada pada 0.4% dan 0.3%.
- median dari RenewableEnergy banyak terdistribusi pada rentang persentase 0%-3%, namun nilai terbanyak terdapat pada nilai 0.2%.
- rata-rata terbanyak untuk data total SquareFootage yaitu di antara angka 1%-2%.
- rata-rata terbanyak untuk data Occupancy berada di antara angka 6-8 orang.
- distribusi EnergyConsumption miring ke kanan (right-skewed). Hal ini akan berimplikasi pada model.

B. ANALISIS MULTIVARIATE

Data Kategori



Data Numerik



Linear Regresi

2. MODELING

```
#Selanjutnya, untuk melatih data dengan KNN, tuliskan code berikut.
from sklearn.linear_model import LinearRegression
from sklearn.metrics import accuracy_score, r2_score, mean_absolute_error, mean_squared_error

LR = LinearRegression()
LR.fit(X_train, y_train)
print ('Coefficients: ', LR.coef_)
print ('Intercept: ', LR.intercept_)
models.loc['train_mse', 'Linear Regression'] = mean_squared_error(y_pred = LR.predict(X_train), y_true=y_train)

Coefficients: [-5.06771709e+13  2.11809536e+14 -3.70625255e+14]
Intercept: 7358854493151196.0
```

Ridge Regresi

```
from sklearn.linear_model import Ridge
RR = Ridge()
RR.fit(X_train, y_train)
models.loc['train_mse', 'Ridge Regression'] = mean_squared_error(y_pred = RR.predict(X_train), y_true=y_train)
```

Ridge Regresi

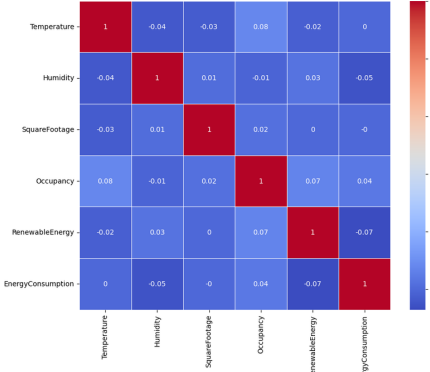
```
#Impor library yang dibutuhkan
from sklearn.ensemble import RandomForestRegressor

#buat model prediksi
RF1 = RandomForestRegressor(random_state=100)
RF1.fit(X_train, y_train)
models.loc['train_mse', 'Random Forest'] = mean_squared_error(y_pred=RF1.predict(X_train), y_true=y_train)
```

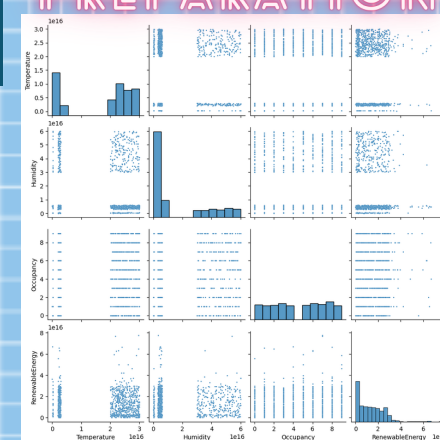
Random Forest dengan Tuning GridSearchCV

```
{'max_depth': 3,
 'min_samples_leaf': 4,
 'min_samples_split': 2,
 'n_estimators': 80}
```

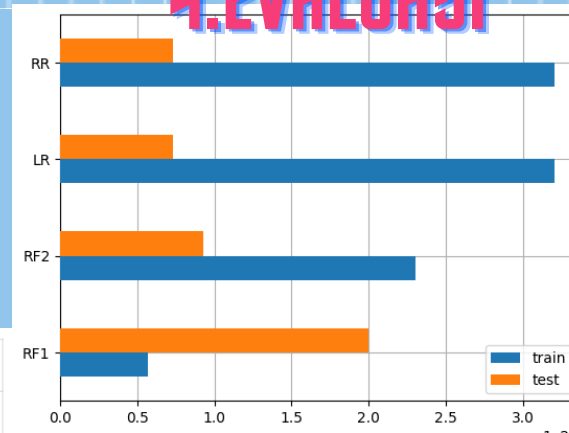
Correlation Matrix untuk Fitur Numerik



3. DATA PREPARATION



4. EVALUASI



Ketahui Lebih lanjut dan Lengkap pada Link!

[MACHINE-LEARNING](#)

BY. FADILA NUR AZIZAH

	y_true	prediksi_LR	prediksi_RR	prediksi_RF1	prediksi_RF2
131	7.142759e+14	7.633840e+15	7.633540e+15	4.995583e+15	5.906515e+15