



BISA.AI



ANALISIS FITUR MENGGUNAKAN MODEL RANDOM FOREST

Mengembangkan model prediksi yang akurat untuk membantu manajemen dalam menentukan strategi penjualan dan distribusi yang optimal.

Nurico Aditya Nugroho

LATAR BELAKANG

01

LATAR BELAKANG

Perusahaan memiliki berbagai jenis usaha dan cabang di seluruh dunia, membutuhkan analisis prediktif untuk penjualan di outlet mereka.

02

TUJUAN

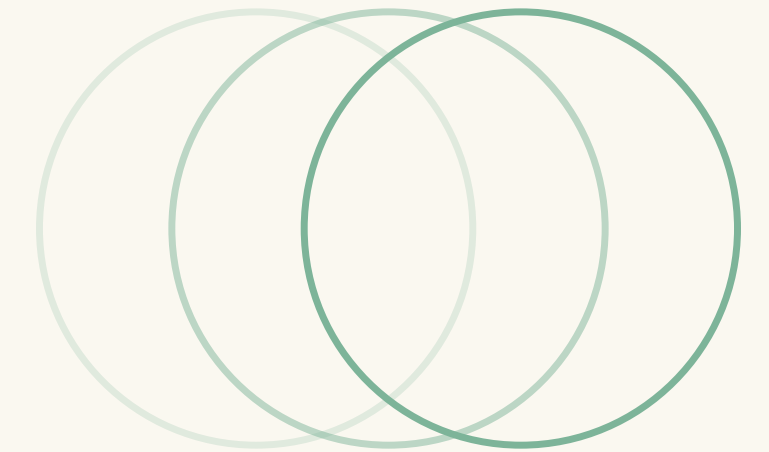
Mengembangkan model prediksi yang akurat untuk membantu manajemen dalam menentukan strategi penjualan dan distribusi yang optimal.

03

METODOLOGI

Data preprocessing, Exploratory Data Analysis (EDA), Feature Engineering, Model Building, Model Evaluation, dan Interpretasi Hasil.

DATA COLLECTION AND PREPROCESSING



DATA COLLECTION

Upload dataset dan mengimport semua library yang diperlukan dalam analisis data

Load Dataset dan Import Library yang Diperlukan

```
from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

[40] # Import Library yang Diperlukan
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import RandomizedSearchCV

# Load datasets
data = pd.read_csv('/content/drive/MyDrive/dataset/Train.csv')
```

PREPROCESSING DATA

Melakukan proses preprocessing sederhana yaitu mengisi nilai yang hilang

Preprocessing Data

```
[41] # Mengecek nilai yang hilang
print(data.isnull().sum())

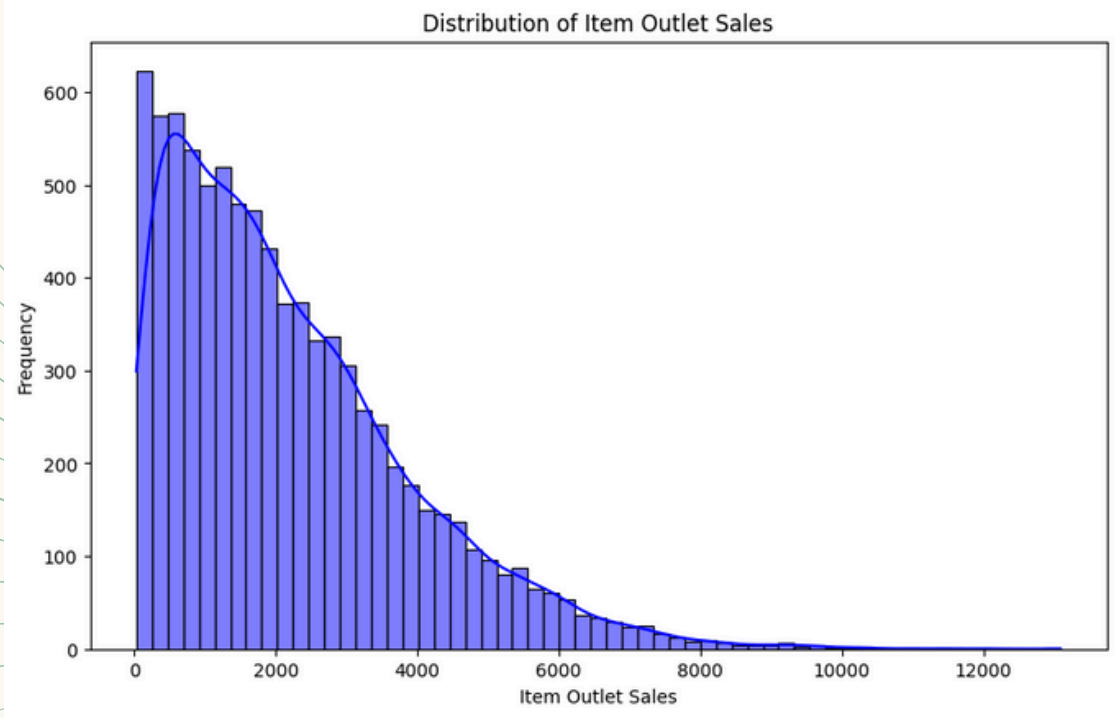
Item_Identifier      0
Item_Weight          1463
Item_Fat_Content      0
Item_Visibility       0
Item_Type            0
Item_MRP             0
Outlet_Identifier     0
Outlet_Establishment_Year  0
Outlet_Size          2410
Outlet_Location_Type  0
Outlet_Type           0
Item_Outlet_Sales     0
dtype: int64

[42] # Mengisi nilai yang hilang
data['Item_Weight'].fillna(data['Item_Weight'].mean(), inplace=True)
data['Outlet_Size'].fillna(data['Outlet_Size'].mode()[0], inplace=True)
print(data.isnull().sum())
```

EXPLORATORY DATA ANALYSIS (EDA)

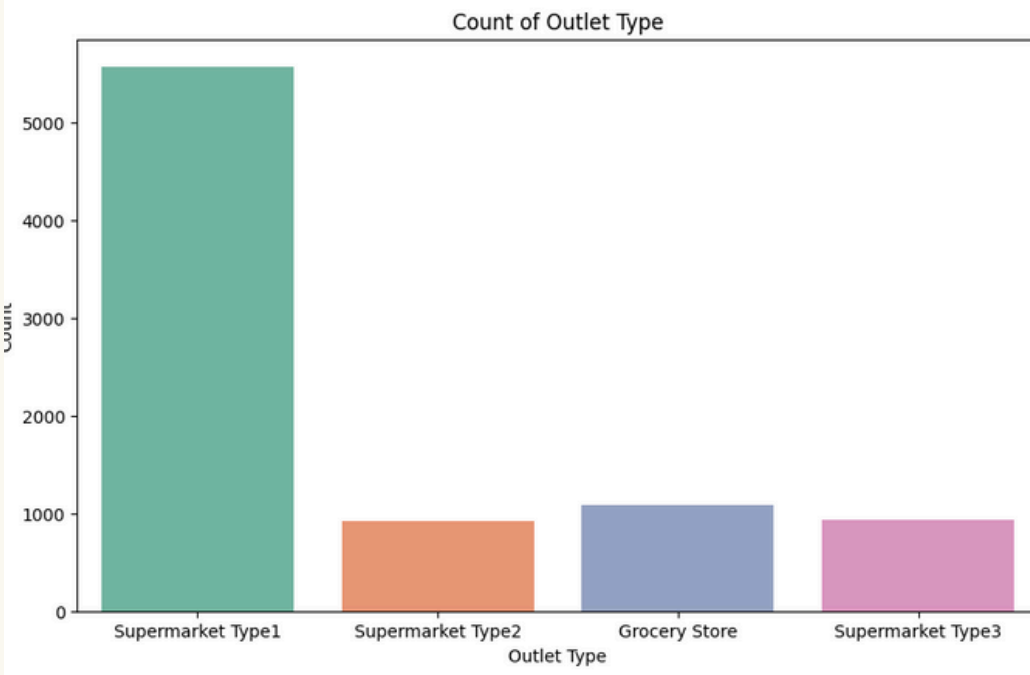
01

DISTRIBUTION OF ITEM OUTLET SALES



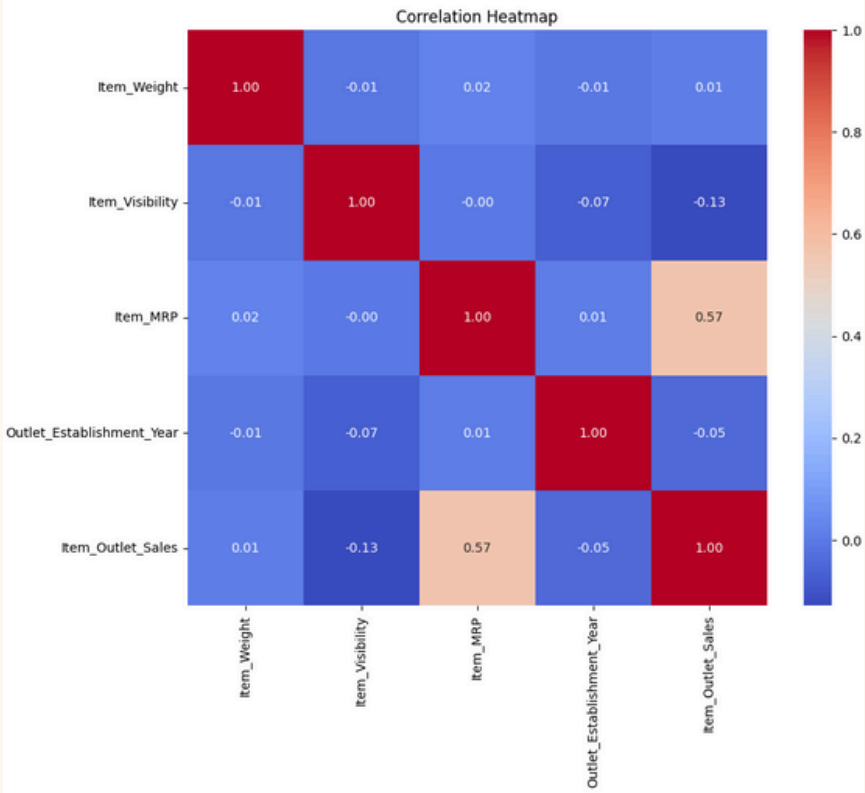
02

COUNT OF OUTLET TYPE



03

CORRELATION HEATMAP



MODEL BUILDING, EVALUATION DAN TUNING

Feature Engineering

```
[47] # Encoding categorical features with one-hot encoding
data_encoded = pd.get_dummies(data, columns=[
    'Item_Fat_Content', 'Item_Type', 'Outlet_Size',
    'Outlet_Location_Type', 'Outlet_Type'
])

# Splitting data into features and target
X = data_encoded.drop(columns=['Item_Outlet_Sales'])
y = data_encoded['Item_Outlet_Sales']
```

Model Building

```
[48] # Splitting the data into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)

# Training a basic Random Forest model
model = RandomForestRegressor(random_state=42)
model.fit(X_train, y_train)

# Predicting on validation set
y_pred = model.predict(X_val)
```

Model Evaluation

```
# Calculating evaluation metrics
mae = mean_absolute_error(y_val, y_pred)
rmse = mean_squared_error(y_val, y_pred, squared=False)
r2 = r2_score(y_val, y_pred)

print(f"Mean Absolute Error (MAE): {mae}")
print(f"Root Mean Squared Error (RMSE): {rmse}")
print(f"R-squared (R2 Score): {r2}")
```

```
➡ Mean Absolute Error (MAE): 761.0150114639296
Root Mean Squared Error (RMSE): 1091.2335991758343
R-squared (R2 Score): 0.5618823096707173
```

Model Tuning

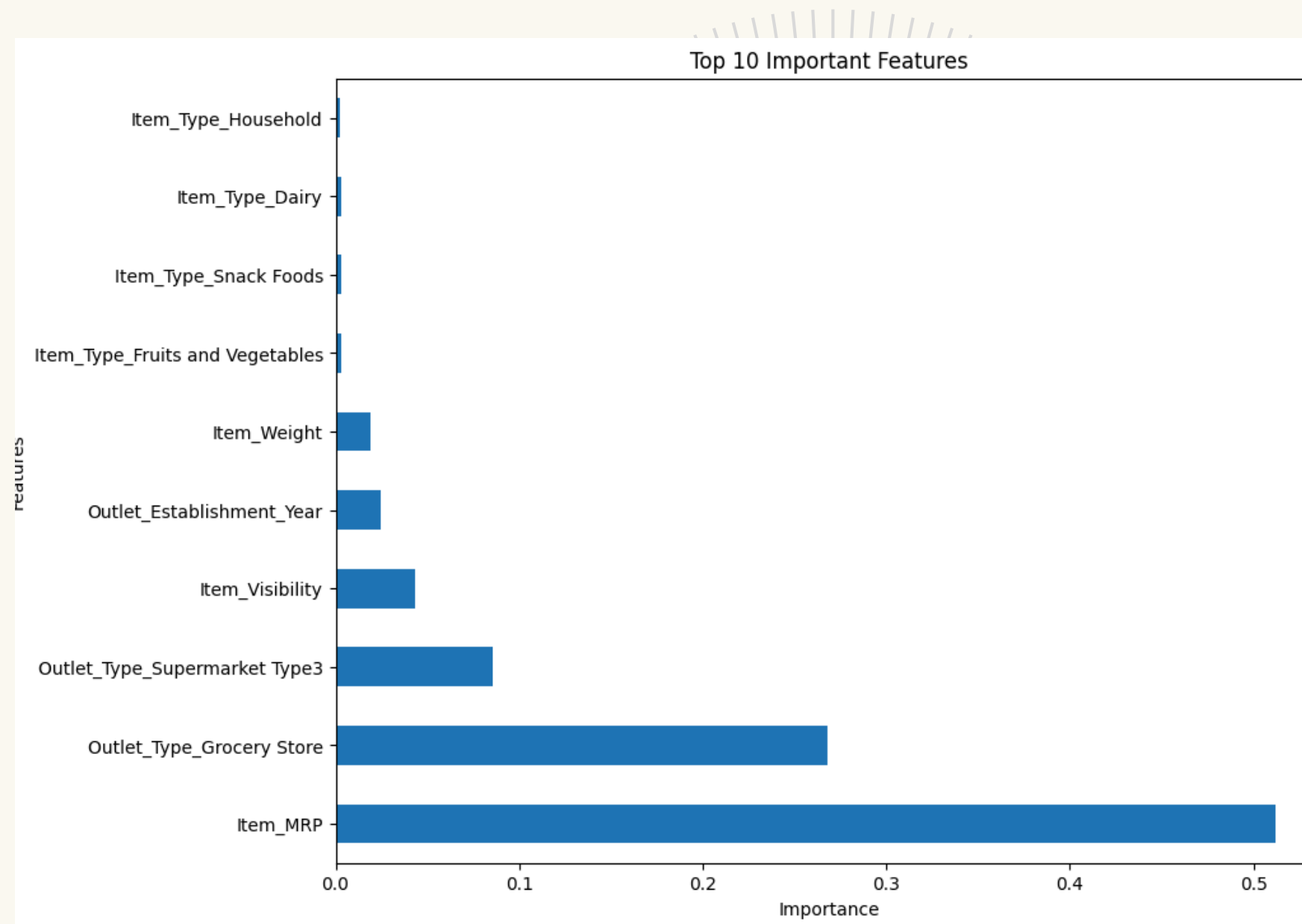
```
# Defining parameter grid for tuning
param_dist = {
    'n_estimators': [50, 100],
    'max_depth': [None, 10],
    'min_samples_split': [2, 5]
}

# Running RandomizedSearchCV
random_search = RandomizedSearchCV(estimator=model, param_distributions=param_dist, n_iter=10, cv=2)
random_search.fit(X_train, y_train)

# Best parameters and score
print("Best Parameters:", random_search.best_params_)
print("Best Score:", -random_search.best_score_)
```

```
➡ /usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_search.py:320: UserWarning: The to
warnings.warn(
Best Parameters: {'n_estimators': 100, 'min_samples_split': 5, 'max_depth': 10}
Best Score: 791.7153092812055
```

INTERPRETASI HASIL



Fitur Penting dalam Strategi Penjualan

01 ITEM_MRP

Fitur ini memiliki nilai kepentingan tertinggi dan berkontribusi paling besar terhadap prediksi penjualan. Ini masuk akal karena harga ritel maksimum (MRP) biasanya sangat memengaruhi pendapatan dari penjualan suatu produk.

02 OUTLET_TYPE_GROCERY STORE

tipe outlet tempat produk dijual berpengaruh pada penjualan. Ini mungkin karena tipe toko ini memiliki pola penjualan yang berbeda dibandingkan supermarket, baik dalam volume penjualan atau jenis produk yang dijual.

KESIMPULAN

Insight Utama dari perusahaan terletak pada harga ritel maksimum produk dan jenis outlet untuk tempat pemasaran. Kedua hal tersebut merupakan hal yang paling mempengaruhi dalam penjualan produk.

STRATEGI UNTUK PERUSAHAAN

01 MENYESUAIKAN HARGA RITEL

**02 OPTIMALISASI PEMILIHAN JENIS
OUTLET**

<https://colab.research.google.com/drive/1mTXhyPSZLKvjp1rDhW3QWilj3ymFADK8?usp=sharing>