

KELOMPOK 3

Prediksi Pengeluaran Customer with Teckitchen



ANGGOTA TIM



ROBBY
DATA ENGINEER



SAIFUL ISLAM
MANAGER



FERY
DATA SCIENCE



NABILLA
BUSINESS DEVELOPMENT



DEA ASHARI
PRODUCT DEVELOPMENT

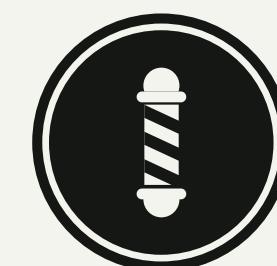
TOPIC PROJECT



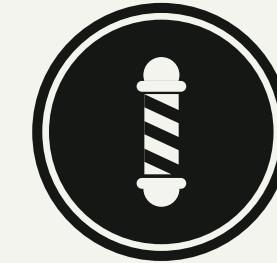
INTRODUCTION



GOALS



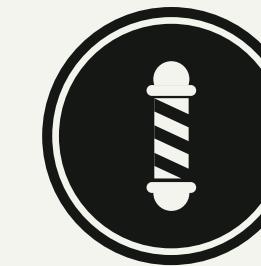
PROBLEM SCOPING



DATA ACQUISITION



DATA EXPLORATION



MODELLING



EVALUATION



DEPLOYMENT

INTRODUCTION

KAMI ADALAH...

The TecKitchen adalah perusahaan yang bergerak dibidang Fintech. Pada kesempatan ini kami ingin mempresentasikan projek kami yaitu melakukan prediksi pengeluaran berdasarkan data.

GOALS

TUJUAN KAMI

Kami memiliki tujuan untuk membantu menemukan solusi dari sebuah permasalahan berdasarkan data yang diberikan dengan menggunakan teknologi kecerdasan buatan.

Dalam masalah ini kami ingin membantu prediksi pengeluaran seorang user berdasarkan data.

WHO

User

WHAT

Prediksi pengeluaran user

PROBLEM SCOPING

WHERE

JAWA TIMUR

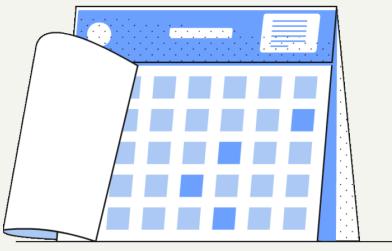
WHY

**Untuk membantu
memprediksi pengeluaran
yang dimiliki oleh User**

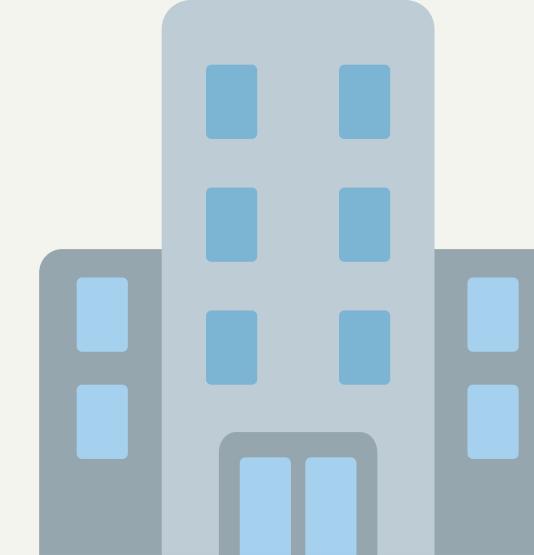
DATA ACQUISITION



Tipe motor



Sales date



District



Gender/jenis kelamin



Pemasukan



Pekerjaan



DATA EXPLORATION

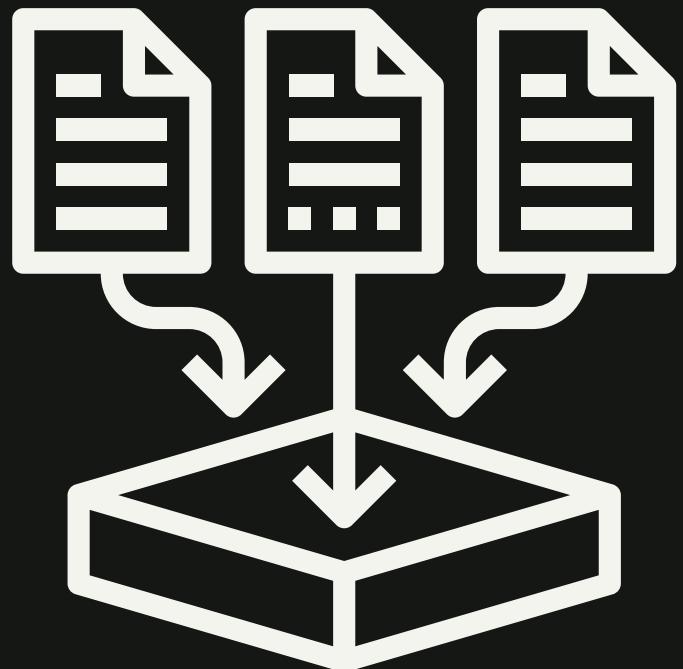
```
1 df.info()
```

```
↳ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 18616 entries, 0 to 18615
Data columns (total 21 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   TYPE MOTOR      18616 non-null   object  
 1   COLOR           18616 non-null   object  
 2   KODE CUSTOMER  18616 non-null   object  
 3   JENIS KELAMIN  18616 non-null   object  
 4   KECAMATAN SURAT 18616 non-null   object  
 5   KOTA SURAT      18616 non-null   object  
 6   PROPINSI        18616 non-null   object  
 7   STATUS RUMAH    18616 non-null   object  
 8   JENIS PENJUALAN STNK 18616 non-null   object  
 9   JENIS PENJUALAN SSU 18616 non-null   object  
 10  AGAMA           18616 non-null   object  
 11  PEKERJAAN       18616 non-null   object  
 12  PENGELUARAN     18616 non-null   object  
 13  PENDIDIKAN      18616 non-null   object  
 14  KEBERSEDIAAN DIHUBUNGI 18616 non-null   object  
 15  MERK MOTOR SBLMNYA 18616 non-null   object  
 16  TYPE MOTOR SBLMNYA 18616 non-null   object  
 17  SMH DIGUNAKAN UNTUK 18616 non-null   object  
 18  YG MENGGUNAKAN SMH 18616 non-null   object  
 19  HOBI            18616 non-null   object  
 20  KEWARGANEGARAAN 18610 non-null   object  
dtypes: object(21)
memory usage: 3.0+ MB
```

```
1 df.info()
```

```
↳ <class 'pandas.core.frame.DataFrame'>
Int64Index: 18610 entries, 0 to 18615
Data columns (total 21 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   TYPE MOTOR      18610 non-null   int64  
 1   COLOR           18610 non-null   int64  
 2   KODE CUSTOMER  18610 non-null   int64  
 3   JENIS KELAMIN  18610 non-null   int64  
 4   KECAMATAN SURAT 18610 non-null   int64  
 5   KOTA SURAT      18610 non-null   int64  
 6   PROPINSI        18610 non-null   int64  
 7   STATUS RUMAH    18610 non-null   int64  
 8   JENIS PENJUALAN STNK 18610 non-null   int64  
 9   JENIS PENJUALAN SSU 18610 non-null   int64  
 10  AGAMA           18610 non-null   int64  
 11  PEKERJAAN       18610 non-null   int64  
 12  PENGELUARAN     18610 non-null   int64  
 13  PENDIDIKAN      18610 non-null   int64  
 14  KEBERSEDIAAN DIHUBUNGI 18610 non-null   int64  
 15  MERK MOTOR SBLMNYA 18610 non-null   int64  
 16  TYPE MOTOR SBLMNYA 18610 non-null   int64  
 17  SMH DIGUNAKAN UNTUK 18610 non-null   int64  
 18  YG MENGGUNAKAN SMH 18610 non-null   int64  
 19  HOBI            18610 non-null   int64  
 20  KEWARGANEGARAAN 18610 non-null   int64  
dtypes: int64(21)
memory usage: 3.1 MB
```

DATA EXPLORATION



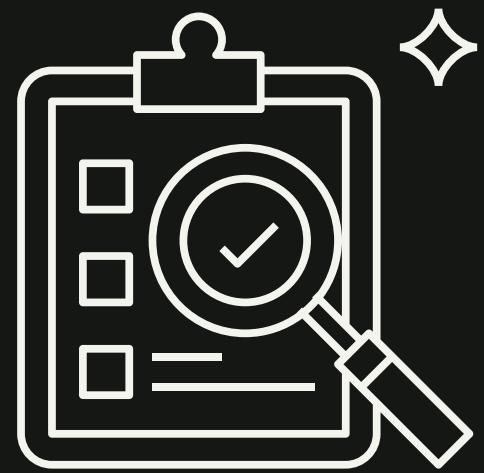
```
y = df[['PENGELUARAN']] # Variabel Terikat
X = df[['STATUS RUMAH','JENIS KELAMIN','TYPE MOTOR','PEKERJAAN','PENDIDIKAN','SMH DIGUNAKAN UNTUK','YG MENGGUNAKAN SMH','HOBI']] # Variabel Bebas

y.info()
X.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 18610 entries, 0 to 18615
Data columns (total 1 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   PENGELUARAN    18610 non-null   int64  
dtypes: int64(1)
memory usage: 290.8 KB
<class 'pandas.core.frame.DataFrame'>
Int64Index: 18610 entries, 0 to 18615
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   STATUS RUMAH    18610 non-null   int64  
 1   JENIS KELAMIN   18610 non-null   int64  
 2   TYPE MOTOR      18610 non-null   int64  
 3   PEKERJAAN       18610 non-null   int64  
 4   PENDIDIKAN      18610 non-null   int64  
 5   SMH DIGUNAKAN UNTUK 18610 non-null   int64  
 6   YG MENGGUNAKAN SMH 18610 non-null   int64  
 7   HOBI            18610 non-null   int64  
dtypes: int64(8)
memory usage: 1.3 MB
```

MODELING

Kami menggunakan Metode supervised learning dan algoritma random forest classifier. Alasan kami menggunakan Algoritma random forest classifier adalah karena adanya 2 atau lebih variabel yang saling terikat satu sama lain.



EVALUATION

model yang dibuat memiliki akurasi 75, 9% untuk memprediksi penggunaan sepeda motor dari variabel : "STATUS RUMAH", "JENIS KELAMIN", "TYPE MOTOR", "COLOR", "PEKERJAAN", "PENDIDIKAN"

Dengan memasukan nilai variabel yang digunakan, maka menghasilkan prediksi penggunaan sepeda motor oleh kostumer.

DEPLOYMENT

Analisis yang kami dapatkan dengan menggunakan metode supervised learning dengan metode algoritma random forest classifier adalah sebagai berikut:

yaitu dengan parameter yang digunakan kita bisa memprediksi

```
[48] y = df[['SMH DIGUNAKAN UNTUK']] # Variabel Terikat
    X = df[['STATUS RUMAH','JENIS KELAMIN','TYPE MOTOR','COLOR','PEKERJAAN','PENDIDIKAN','YG MENGGUNAKAN SMH','HOBI']] # Variabel Bebas
    ↑ ↓
y.info()
X.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 18610 entries, 0 to 18615
Data columns (total 1 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   SMH DIGUNAKAN UNTUK  18610 non-null   int64  
dtypes: int64(1)
memory usage: 290.8 KB
<class 'pandas.core.frame.DataFrame'>
Int64Index: 18610 entries, 0 to 18615
Data columns (total 8 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   STATUS RUMAH      18610 non-null   int64  
 1   JENIS KELAMIN     18610 non-null   int64  
 2   TYPE MOTOR        18610 non-null   int64  
 3   COLOR             18610 non-null   int64  
 4   PEKERJAAN         18610 non-null   int64  
 5   PENDIDIKAN        18610 non-null   int64  
 6   YG MENGGUNAKAN SMH 18610 non-null   int64  
 7   HOBI              18610 non-null   int64  
dtypes: int64(8)
memory usage: 1.3 MB
```

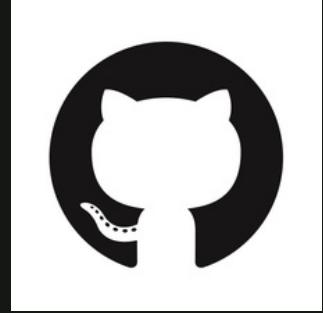
THANK YOU



Reference



https://colab.research.google.com/drive/1igyud_W4_GMwUHeiZTTxOvxu7teGJdk?usp=sharing



1. <https://github.com/Deaashari/OrbitGuru> (Dea Ashari Oktovia)
2. <https://github.com/Snakesystem> (Feri Irawan)
3. <https://github.com/nabillarmdhn> (Nabilla Ramadhani)
4. <https://github.com/Saifulislam365> (Saiful Islam)
5. <https://github.com/robbi621/data-set.git> (Robi Adam)