



Credition by TecKitchen

OUR ION TEAM



Mellisa

Data Scientist



Banyu

Project Manager



Prissy

Data Analyst



Arifa

Data Architect



Taufiq

Data Engineer

DISCUSSION POINTS

01

Introduction

02

Goals

03

Problem Scoping

04

Data Acquisition

05

Data Exploration

06

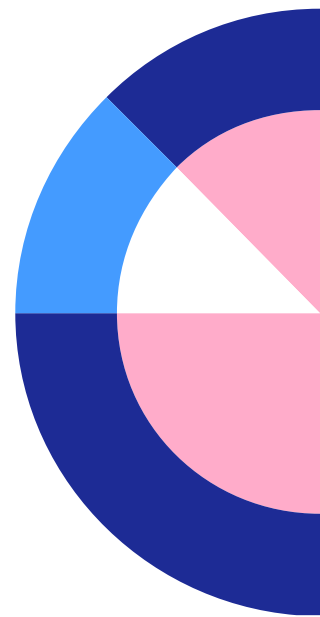
Modeling

07

Evaluation

08

Deployment



INTRODUCTION

Credition merupakan sebuah sistem yang digunakan dalam hal prediksi dan pengelompokan pembayaran yang dilakukan customer. Dilatarbelakangi dengan melonjaknya penjualan motor menjadi salah satu target TecKitchen untuk membantu dealer motor dalam mengetahui pembayaran yang dilakukan calon pembeli dan meng-clustering.



GOALS

Untuk memprediksi customer yang akan melakukan pembayaran dan mengelompokkan customer yang melakukan pembayaran secara credit dalam transaksi pembelian motor tersebut.





• **AI Project Cycle**

PROBLEM SCOPING

WHO

Dealer Motor

WHAT

Customer yang banyak

WHERE

Kab. Gresik, Jawa Timur

WHY

Untuk mengetahui
pembayaran yang akan
dilakukan customer

DATA ACQUISITION

Tim kami menggunakan dataset yang berasal dari salah satu Data Master penjualan cabang Dealer Motor, dan berikut merupakan beberapa data yang tercantum.

	SALES DATE	TYPE MOTOR	COLOR	KODE CUSTOMER	JENIS KELAMIN	TANGGAL LAHIR	KECAMATAN SURAT	KOTA SURAT	PROPINSI	STATUS RUMAH	JENIS PENJUALAN STNK	JENIS PENJUALAN SSU	BESAR DP
0	08/08/2019	REVO FIT FI	HITAM	INDIVIDUAL	LAKI-LAKI	25/08/1970	Manyar	Kab. Gresik	Jawa Timur	RUMAH SENDIRI	CASH	CASH	NaN

Dataset Attributes

- SALES DATE
- TYPE MOTOR
- COLOR
- KODE CUSTOMER
- JENIS KELAMIN
- TANGGAL LAHIR
- KECAMATAN SURAT
- KOTA SURAT
- PROPINSI
- STATUS RUMAH
- JENIS PENJUALAN STNK
- JENIS PENJUALAN SSU
- BESAR DP
- BESAR CICILAN
- LAMA CICILAN
- AGAMA
- PEKERJAAN
- PENGELUARAN
- PENDIDIKAN
- KEBERSEDIAAN DIHUBUNGI
- MERK MOTOR
- SBLMNYA
- TYPE MOTOR
- SBLMNYA
- SMH DIGUNAKAN UNTUK
- YG MENGGUNAKAN SMH
- HOBI
- KETERANGAN
- KEWARGANEGARAAN

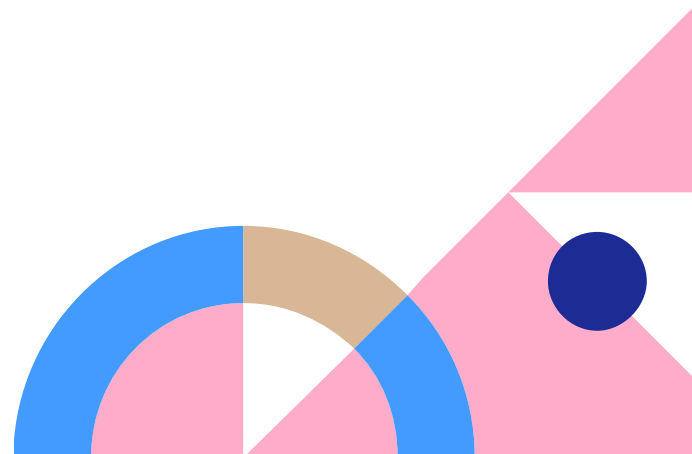
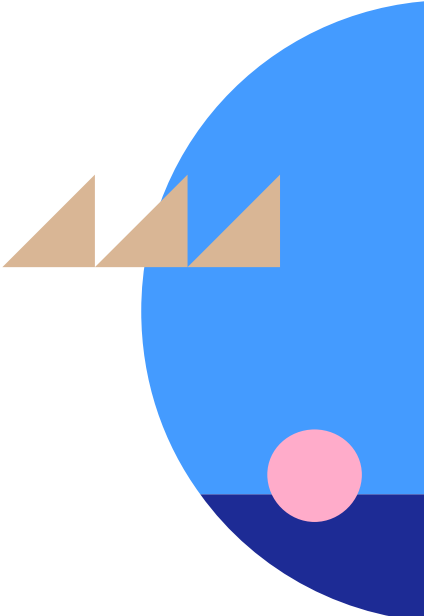
18.616 rows × 27 columns



DATA EXPLORATION

Untuk membangun model, tim kami hanya menggunakan semua kolom kecuali keterangan.

Dan memisahkan “TANGGAL LAHIR”, sehingga dimiliki tanggal, bulan, dan tahun lahir masing-masing dalam bentuk integer pada kolom-kolom terpisah



DATA EXPLORATION

```
[ ] df_motorcredit = df[["TYPE MOTOR", "JENIS PENJUALAN STNK"]][df["JENIS PENJUALAN STNK"]=="CREDIT"]
percent_credit = len(df_motorcredit)/len(df)*100
print(f"Jumlah motor yang dikredit adalah {len(df_motorcredit)} dari {len(df)}")
print(f"yaitu {percent_credit} % dari total transaksi\n")
```

```
Jumlah motor yang dikredit adalah 11266 dari 18616
yaitu 60.51783412118608 % dari total transaksi
```

```
[ ] # Ingin dimiliki variabel usia yang diambil dari tahun lahir
# Variabel "TANGGAL LAHIR" dipisah sehingga dimiliki tanggal, bulan, dan tahun lahir masing-masing dalam bentuk integer pada kolom-kolom terpisah
df[df.columns[5]] = df[df.columns[5]].replace('/', ' ', regex=True).astype(object)
df[['Tanggal lahir', 'Bulan lahir', 'Tahun lahir']] = df["TANGGAL LAHIR"].str.split(' ', 2, expand=True).astype(int)
```

Analisis jumlah motor yang dibeli secara credit oleh pembeli dan
splitting kolom "TANGGAL LAHIR"

Berikut adalah data sebelum dan sesudah di analisa serta di cleansing.

Data columns (total 27 columns):

#	Column	Non-Null Count	Dtype
0	SALES DATE	18616 non-null	object
1	TYPE MOTOR	18616 non-null	object
2	COLOR	18616 non-null	object
3	KODE CUSTOMER	18616 non-null	object
4	JENIS KELAMIN	18616 non-null	object
5	TANGGAL LAHIR	18616 non-null	object
6	KECAMATAN SURAT	18616 non-null	object
7	KOTA SURAT	18616 non-null	object
8	PROPINSI	18616 non-null	object
9	STATUS RUMAH	18616 non-null	object
10	JENIS PENJUALAN STNK	18616 non-null	object
11	JENIS PENJUALAN SSU	18616 non-null	object
12	BESAR DP	10928 non-null	float64
13	BESAR CICILAN	10928 non-null	float64
14	LAMA CICILAN	10928 non-null	float64
15	AGAMA	18616 non-null	object
16	PEKERJAAN	18616 non-null	object
17	PENGELUARAN	18616 non-null	object
18	PENDIDIKAN	18616 non-null	object
19	KEBERSEDIAAN DIHUBUNGI	18616 non-null	object
20	MERK MOTOR SBLMNYA	18616 non-null	object
21	TYPE MOTOR SBLMNYA	18616 non-null	object
22	SMH DIGUNAKAN UNTUK	18616 non-null	object
23	YG MENGGUNAKAN SMH	18616 non-null	object
24	HOBİ	18616 non-null	object
25	KETERANGAN	7262 non-null	object
26	KEWARGANEGARAAN	18610 non-null	object

dtypes: float64(3), object(24)

memory usage: 3.8+ MB

Data columns (total 29 columns):

#	Column	Non-Null Count	Dtype
0	SALES DATE	18616 non-null	object
1	TYPE MOTOR	18616 non-null	object
2	COLOR	18616 non-null	object
3	KODE CUSTOMER	18616 non-null	object
4	JENIS KELAMIN	18616 non-null	object
5	TANGGAL LAHIR	18616 non-null	object
6	KECAMATAN SURAT	18616 non-null	object
7	KOTA SURAT	18616 non-null	object
8	PROPINSI	18616 non-null	object
9	STATUS RUMAH	18616 non-null	object
10	JENIS PENJUALAN STNK	18616 non-null	object
11	JENIS PENJUALAN SSU	18616 non-null	object
12	BESAR DP	10928 non-null	float64
13	BESAR CICILAN	10928 non-null	float64
14	LAMA CICILAN	10928 non-null	float64
15	AGAMA	18616 non-null	object
16	PEKERJAAN	18616 non-null	object
17	PENGELUARAN	18616 non-null	object
18	PENDIDIKAN	18616 non-null	object
19	KEBERSEDIAAN DIHUBUNGI	18616 non-null	object
20	MERK MOTOR SBLMNYA	18616 non-null	object
21	TYPE MOTOR SBLMNYA	18616 non-null	object
22	SMH DIGUNAKAN UNTUK	18616 non-null	object
23	YG MENGGUNAKAN SMH	18616 non-null	object
24	HOBİ	18616 non-null	object
25	KEWARGANEGARAAN	18610 non-null	object
26	Tanggal lahir	18616 non-null	int64
27	Bulan lahir	18616 non-null	int64
28	Tahun lahir	18616 non-null	int64

dtypes: float64(3), int64(3), object(23)

memory usage: 4.1+ MB

MODELING

Metode

Supervised Learning

Algoritma

Random Forest Classifier

```
# import algoritma Random Forest
from sklearn.ensemble import RandomForestClassifier

rnd_clf = RandomForestClassifier(n_estimators=1000, max_leaf_nodes=5000, n_jobs=-1)
rnd_clf.fit(X_train, y_train)
y_pred_rf = rnd_clf.predict(X_test)

from sklearn.metrics import accuracy_score

print(f"Akurasi Random Forrest: {accuracy_score(y_test, y_pred_rf)}")

Akurasi Random Forrest: 0.827245380318006
```

EVALUATION

Model yang dibuat memiliki akurasi 82.7% untuk memprediksi apakah pembeli akan menggunakan pembayaran CREDIT untuk membeli motor dari variabel: "TYPE MOTOR", "COLOR", "JENIS KELAMIN", "Tahun lahir", "STATUS RUMAH", "PEKERJAAN", "PENGELUARAN", "MERK MOTOR SBLMNYA", "TYPE MOTOR SBLMNYA", "SMH DIGUNAKAN UNTUK", "YG MENGGUNAKAN SMH", dan "HOBI".

Maka, bisa dilakukan pengambilan data customer baru terkait variabel-variabel tersebut dan kemudian model dapat memprediksi apakah customer baru tersebut memiliki kecenderungan untuk membayar dengan CREDIT.

DEPLOYMENT

Rencananya model akan di-deploy menjadi sebuah sistem berbasis web yang mampu memprediksi seperti apakah kriteria customer yang berpeluang besar melakukan pembelian motor secara kredit.

The background features various geometric shapes in blue, pink, and tan. In the top left, there's a blue triangle and a tan circle. In the top right, there's a pink triangle and a blue circle. In the bottom left, there's a blue triangle and a pink circle. In the bottom right, there's a tan arc. The text is centered in the middle of the slide.

Other WHY

**Mengelompokan Pelanggan yang melakukan
pembayaran credit**

MODELING

Metode

Unsupervised Learning

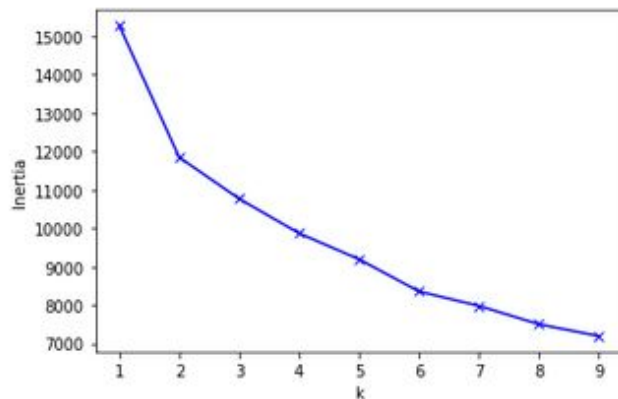
Algoritma

K-Means

```
[24] from sklearn.preprocessing import MinMaxScaler  
from sklearn.cluster import KMeans  
mmscaler = MinMaxScaler()  
features_scaled = mmscaler.fit_transform(df_credit_encode)
```

```
inertia = []  
K = range(1,10)  
for k in K:  
    kmeanModel = KMeans(n_clusters=k).fit(features_scaled)  
    kmeanModel.fit(features_scaled)  
    inertia.append(kmeanModel.inertia_)
```

```
[31] plt.plot(K, inertia, 'bx-')  
plt.xlabel('k')  
plt.ylabel('Inertia')  
plt.show()
```



```
[32] kmeans = KMeans(n_clusters=4).fit(features_scaled)
```

EVALUATION

Model K-Means yang dibuat menggunakan cluster sebanyak 4, berdasarkan grafik L-bow. Setelah dilakukan pemodelan pada data yang telah diolah diketahui bahwa dari 6549 data pelanggan terdapat 959 data pelanggan termasuk kategori 1, 1260 data pelanggan termasuk kategori 2, 2837 data pelanggan termasuk kategori 3, dan 1493 data pelanggan termasuk kategori 4. Kategori tersebut mengarah pada tingkat kredibilitas pelanggan. Kategori 1 termasuk pelanggan yang memiliki tingkat kredibilitas paling tinggi, dan kategori 4 kredibilitas paling rendah.

DEPLOYMENT

Menambahkan fitur pada sistem web untuk mengelompokkan pelanggan yang melakukan credit, sehingga setiap pelanggan mendapatkan perlakuan dan kebijakan yang sesuai dengan kategori yang ditentukan, untuk menghindari kredit macet yang berkelanjutan.

Deployment

Berdasarkan hasil pemodelan, pihak dealer dapat memberikan pelayanan yang sesuai dengan kondisi customer, untuk menjaga hubungan dengan customer. Adapun saran yang bisa diberikan adalah :

1. Pembelian cash diberikan souvenir tambahan ketika pembelian dilakukan, karena transaksi dan komunikasi normalnya hanya terjadi 1x.
2. Pembelian Kredit diberikan promo berkala, misal setelah memenuhi setengah dari total pembayaran diberikan hadiah atau promo untuk menjaga komunikasi terus berlangsung dengan lancar.
3. Khusus Pembelian Kredit dengan kategori 4, dealer perlu menyiapkan strategi khusus, karena pembelian dengan kategori 4 sangat beresiko customer tidak bisa menyelesaikan credit.

References



Our Github :

- <https://github.com/Kemi41>
- <https://github.com/prissyyy>
- <https://github.com/arifanf>
- <https://github.com/mellisairawan>
- <https://github.com/TaufiqHidayatullah528>

Our Google Colab :





Thanks!

Do you have any questions?

And please give us suggestions :)