

# Laporan Kerja Individu

## Final Project - Data Analytics

### Program Zenius Studi Independen Bersertifikat - Angkatan 4

**Nama Lengkap:** [Raihanah Assa'adah]

**Nomor ID live class:** [142]

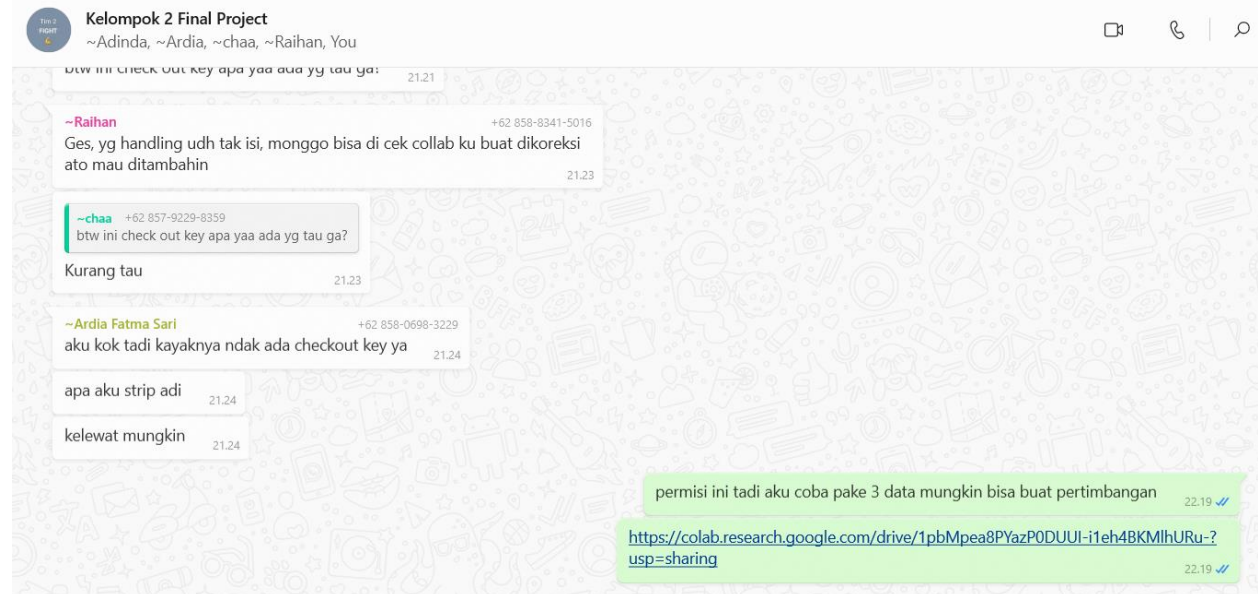
**Nomor Kelompok:** [12]

**Mentor:** [M. Verly]

## Deskripsi peran

- Berdiskusi dengan kelompok menentukan data yang akan dipakai
- Mencoba memilih data yang akan digunakan
- Mendownload data yang diolah di colab
- Melakukan EDA pada data train
- Membuat link looker studio
- Membuat Dashboard

# Lampiran Hasil Kerja



## Menggabungkan Dataset Previous Application dan Bureau

Jumlah peminjaman Home Credit sebelumnya pada tiap pelanggan

```
[ ] # Membuat variabel total previous
prev = previous_application.groupby(['SK_ID_CURR'])['SK_ID_CURR'].agg(['co
prev.columns = ['SK_ID_CURR', 'TOTAL_PREV']
prev.head()
```

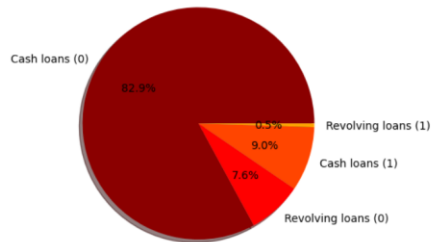
	SK_ID_CURR	TOTAL_PREV
0	100001	1
1	100002	1

```
[ ] mylabels = ["Cash loans (0)", "Revolving loans (0)", "Cash loans (1)", "Revolving loans (1)"]
myexplode = [0, 0, 0, 0]
colors = ( "darkred", "red", "orangered", "orange")
```

```
[ ] def func(pct):
    return "{:1.1f}%".format(pct)
```

```
plt.title("Kemampuan Membayar Loans Berdasarkan Tipe Kontrak", fontsize=12, weight="bold")
plt.pie(ax['SK_ID_CURR'], labels = mylabels, autopct=lambda pct: func(pct),
        explode = myexplode, colors=colors, shadow=True)
labels = ["0 = ", "1="]
plt.show()
```

#### Kemampuan Membayar Loans Berdasarkan Tipe Kontrak

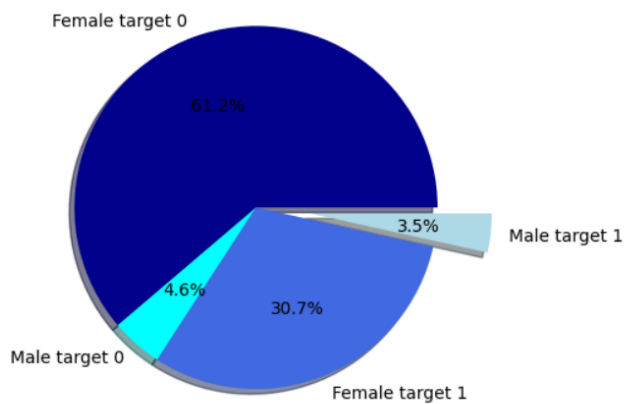


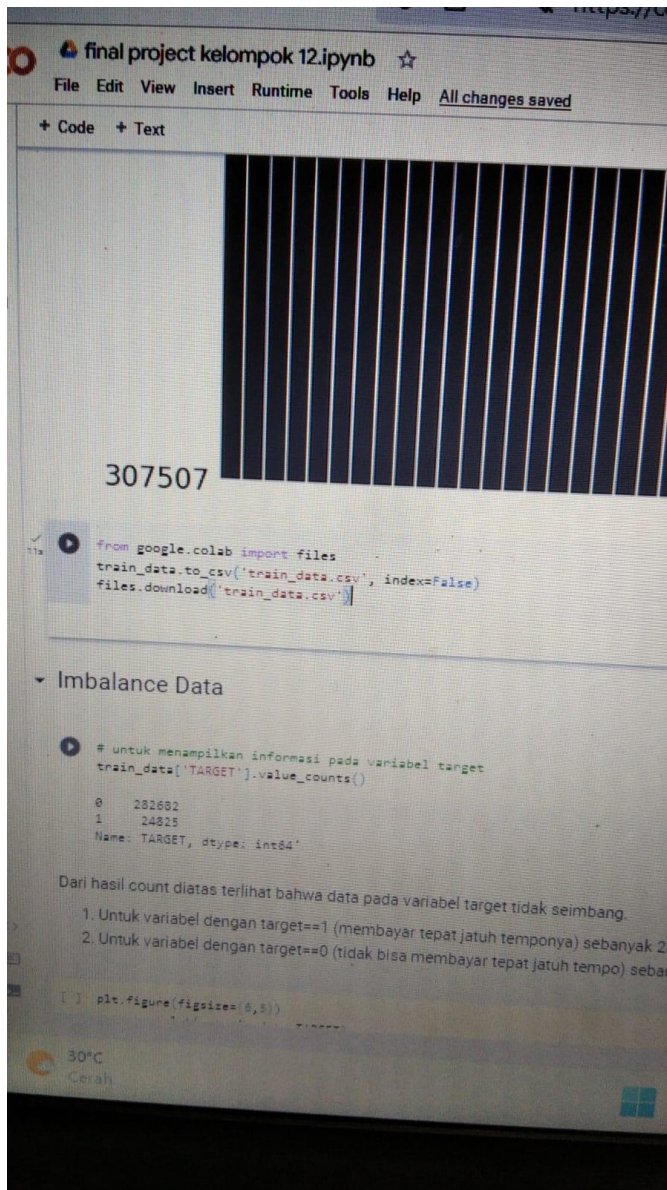
```
[ ] mylabels = ["Female target 0", "Male target 0", "Female target 1", "Male target 1"]
myexplode = [0, 0, 0, 0.3]
colors = ( "darkblue", "cyan", "royalblue", "lightblue")
```

```
[ ] def func(pct):
    return "{:1.1f}%".format(pct)
```

```
plt.title("Kemampuan Membayar Loans berdasarkan Gender", fontsize=12, weight="bold")
plt.pie(ax['SK_ID_CURR'], labels = mylabels, autopct=lambda pct: func(pct),
        explode = myexplode, colors=colors, shadow=True)
plt.show()
```

#### Kemampuan Membayar Loans berdasarkan Gender





## Kelompok 2 Final Project

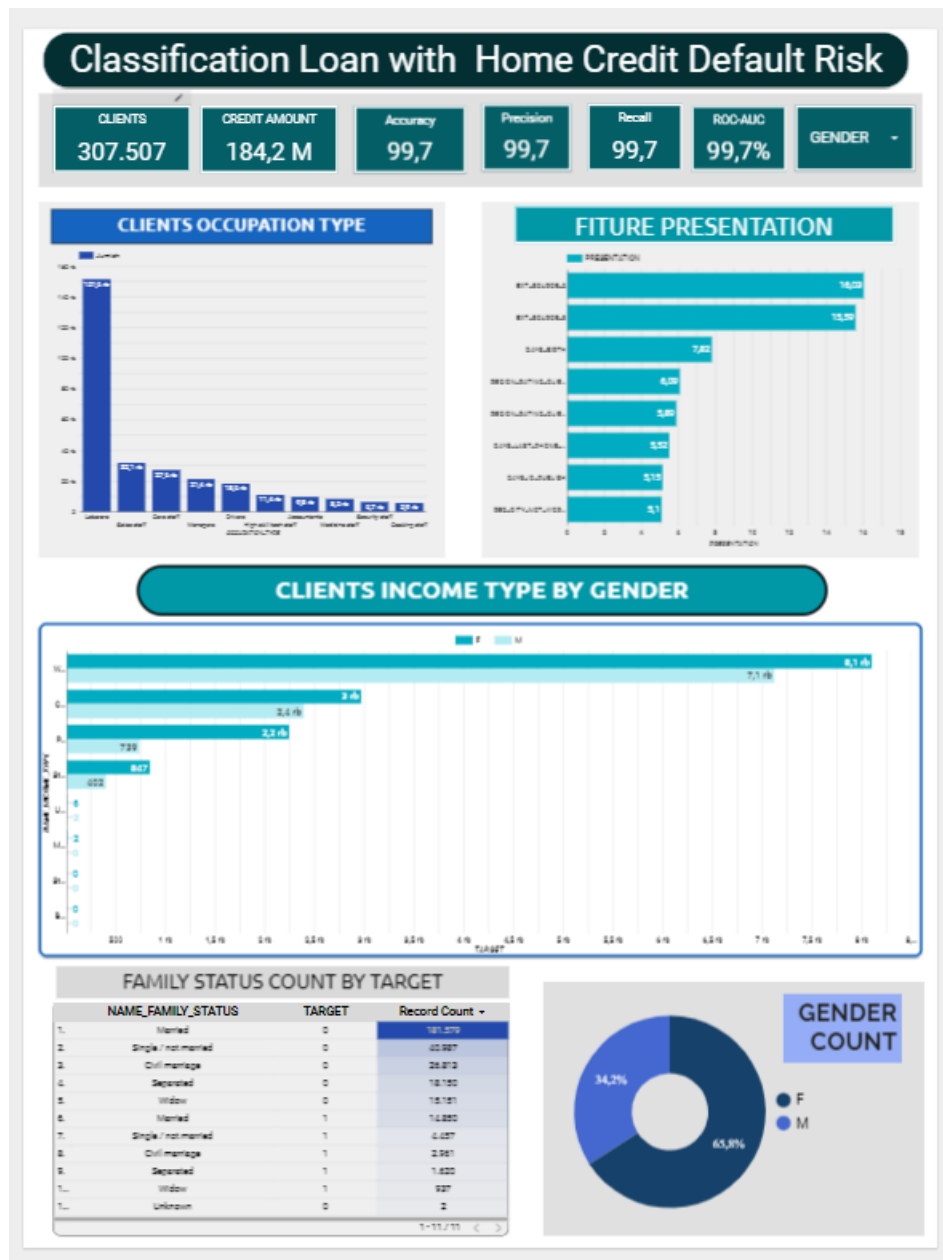
~Adinda, ~Ardia, ~chaa, ~Raihan, You

10.18

<https://lookerstudio.google.com/reporting/91bade96-2953-4a0a-9c32-72920589e100>

Ini linknya dari raihanah

10.18



## Proses Kerja

Proses kerja saya selama final project adalah sebagai berikut.

Setelah ditentukan kelompok, saya dan teman memulai diskusi. Pertama-tama kami berdiskusi dan menyatukan suara terkait apa yang akan kami kerjakan. Kami berdiskusi mengenai apa sebenarnya perintah dan tujuan dari Final Project yang kami kerjakan. Kemudian kami berdiskusi tentang data yang akan digunakan untuk memenuhi tujuan dari pekerjaan kami. Dari dataset yang ada perusahaan memiliki latar belakang penyedia layanan kredit dimana kita diminta untuk memprediksi kemungkinan klien gagal membayar berdasarkan fitur yang ada, contohnya pada

fitur target dengan target==0 (nasabah yang tidak susah dalam mengembalikan pinjaman) dan target==1 (nasabah yang susah mengembalikan pinjaman).

Kemudian, melakukan business understanding dengan pengolahan data di google colab. Pertama dilakukan input data train dan sejumlah library yang digunakan pada analisis ini. Selanjutnya dilakukan data understanding seperti cek kolom, value data, dan info data. Selain itu ditampilkan tipe tipe data yang digunakan.

Pada tahap data preparation dilakukan cek missing value dan ditampilkan persentase data yang missing. Selanjutnya, dilakukan handle missing value dengan menghapus kolom yang sudah dipertimbangkan untuk dihapus, Mengisi missing value pada kolom data kategorik dan numerik). Setelah variabel dengan persentase jumlah missing value lebih dari 50% dihapus, dapat dilakukan metode penanganan Missing Value untuk variabel yang persentase jumlah missing valuenya kurang dari 50% dengan metode fillna. Pada variabel yang memiliki persentase jumlah missing value lebih dari 50%, untuk variabel kategorik diisi dengan Nilai Modus dan variabel numerik diisi dengan Nilai Median.

Setelah dilakukan handle missing value, kemudian dilakukan pengecekan pada imbalance data, bisa dengan melihat visual barplot. Contohnya pada data target terjadi ketimpangan antara kode 1 dan kode 0 di mana kode 0 frekuensinya 282680 dibandingkan dengan kode 1 yang hanya 24825. Kemudian juga dilakukan pengecekan data outliers dan anomalies

Selanjutnya dilakukan EDA (Exploratory Data Analisis) dengan membuat berbagai macam plot sesuai dengan jenis data dan kesesuaian data dengan plot yang akan ditampilkan.

Kemudian pada pemodelan dan evaluasi dilakukan hal-hal sebagai berikut.

Encoding (melakukan label encoding pada dataset categorical/data yang memiliki tipe data "object" dengan menggunakan label encoding). Feature selection dilakukan untuk memilih fitur terbaik dari banyak kumpulan fitur pada data. Dengan melakukan feature selection, pada saat modelling dapat meningkatkan akurasi model. Fitur yang akan digunakan adalah Pearson Correlation. Juga dilakukan splitting pada data train dan test.

Kemudian dilakukan modelling dengan cek confusion matrix dan output model untuk menentukan model mana yang terbaik. Beberapa model yang dicek yaitu

Logistic Regression

Random Forest Classifier

Decision Tree

KN

Dan didapatkan model Random Forest Classifier sebagai model terbaik

## **Persetujuan Anggota**

*Minta anggota kelompok untuk membaca laporan kamu. Jika mereka setuju dengan isi laporan kamu, minta mereka centang checkbox.*

Laporan ini telah disetujui oleh:

✓ Adinda Permata Sari

✓ Himma Faicha Hubbiya

✓ Raihan Tsabita Sabil

✓ Ardia Fatma Sari