

Laporan Kerja Individu

Final Project - Data Analytics

Program Zenius Studi Independen Bersertifikat - Angkatan 4

Nama Lengkap: Ardia Fatma Sari

Nomor ID live class: 129

Nomor Kelompok: 12

Mentor: Muhammad Verly

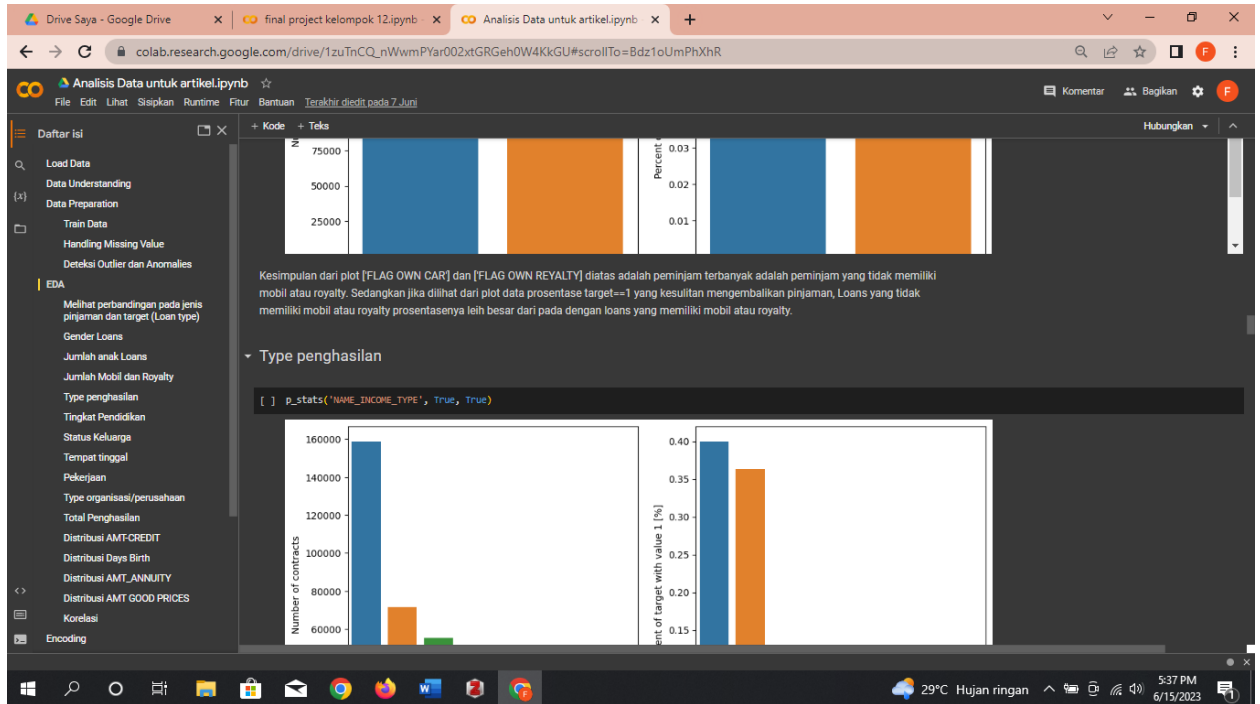
Deskripsi peran

- Melakukan sebagian EDA pada data train berupa pembuatan tabel count, barplot, dan heatmap.
- Membantu dalam Handling Missing Value.
- Melakukan Encoding, Feature Selection, Imbalance Data, Modelling dan Evaluation.
- Membuat bagian dari dashboard berupa barplot yang menampilkan fitur yang korelasi tinggi terhadap target.
- Mengerjakan PPT bagian Modelling dan sebagian EDA.
- Membuat link Looker Studio 1.
- Membantu dalam pembuatan dashboard terkait apa yang harus disertakan.

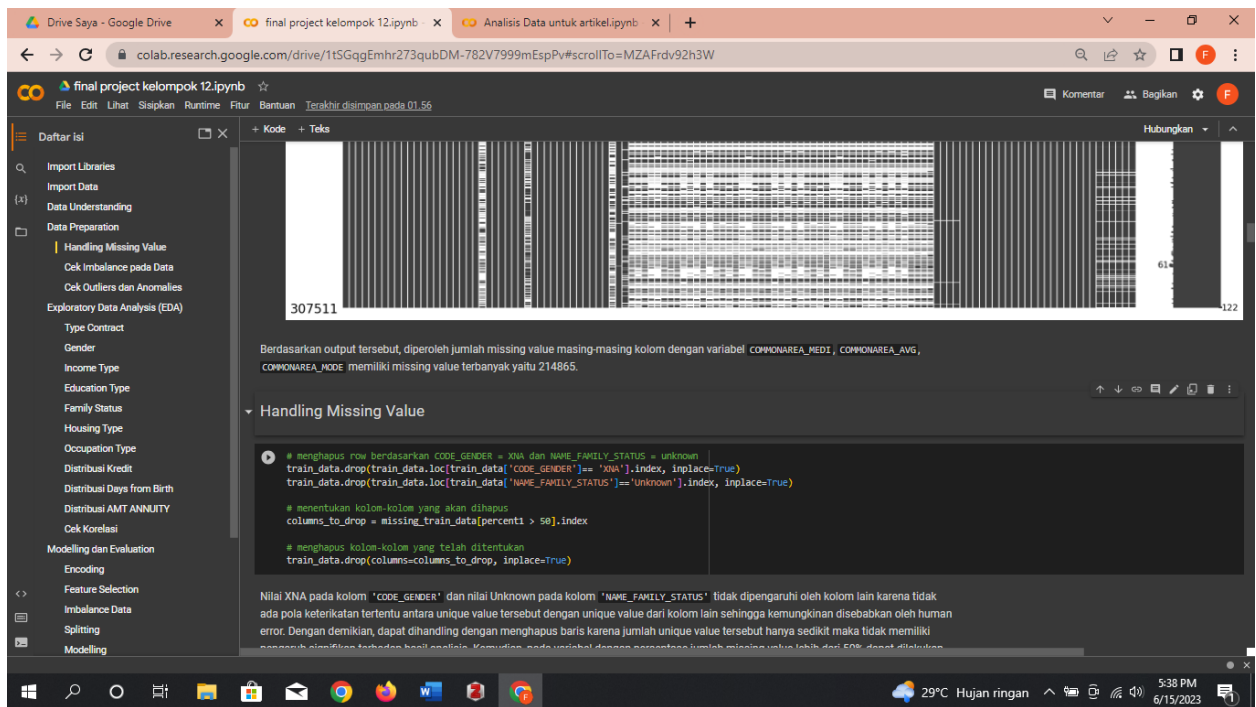
Lampiran Hasil Kerja

- Lampiran Melakukan sebagian EDA pada Data train berupa pembuatan tabel count, barplot, dan heatmap

Screenshoot Kegiatan

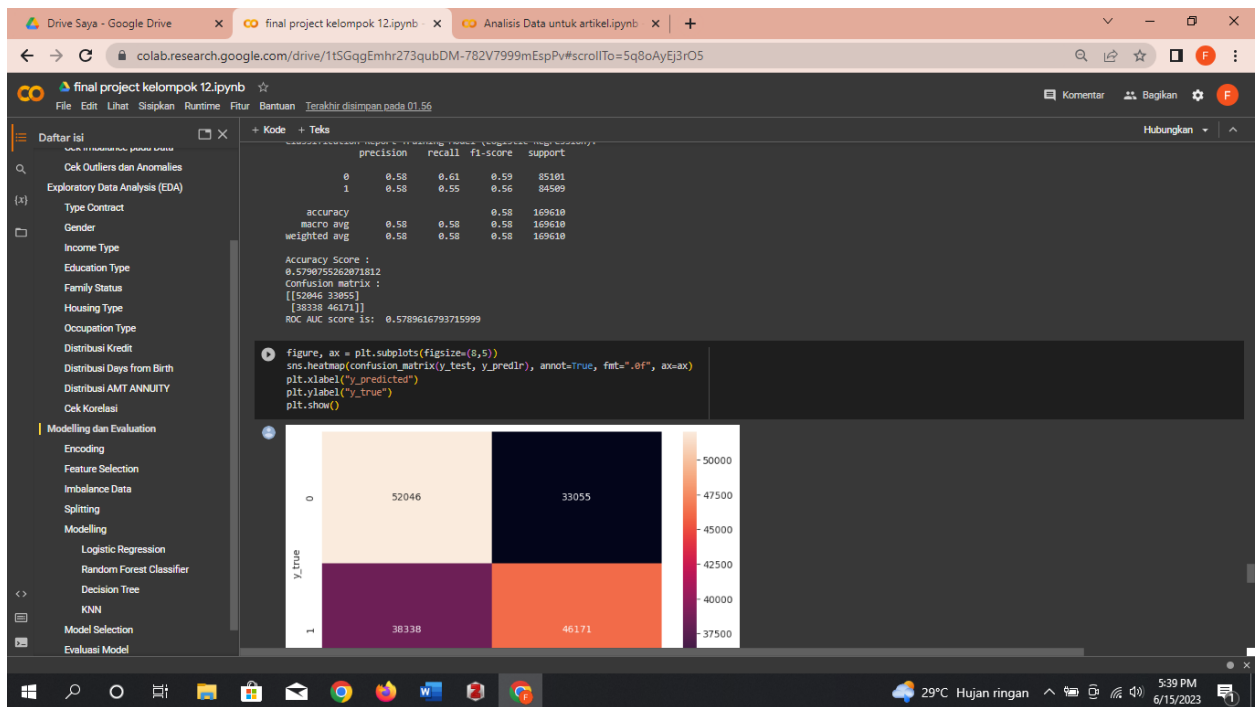


- Lampiran Membantu dalam Handling Missing Value
- ### Hasil Screenshot



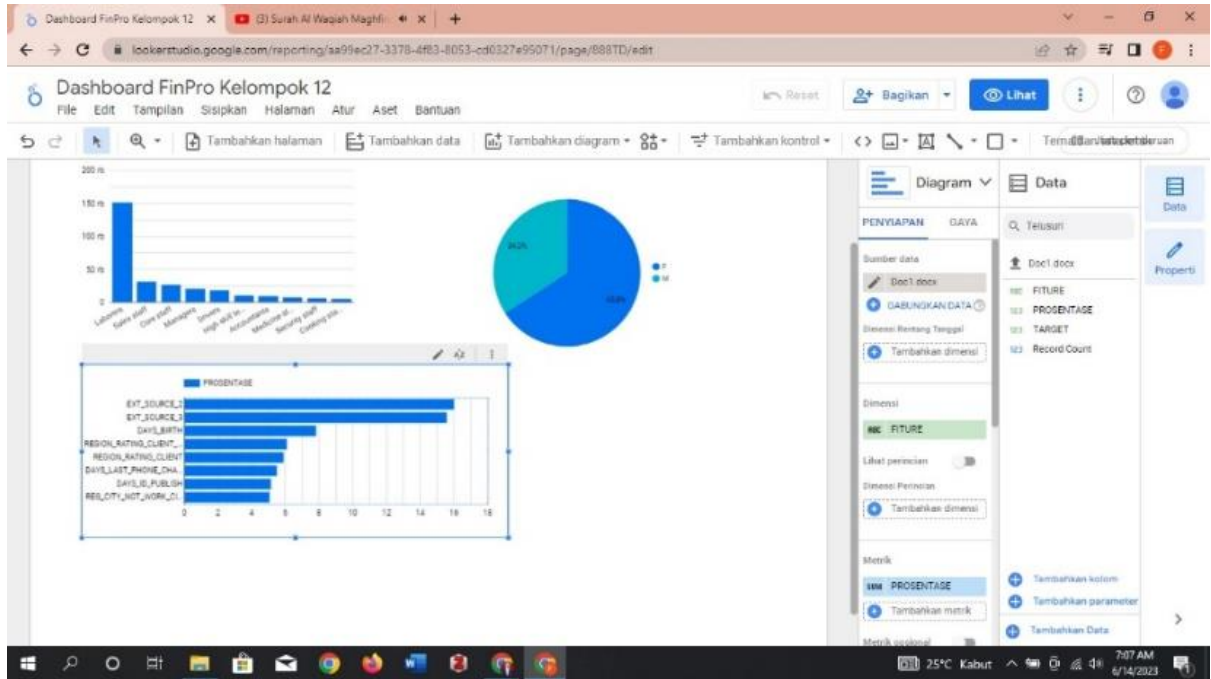
- Lampiran melakukan Encoding, Feature Selection, Imbalance data, modelling, dan evaluation

Screenshot Kegiatan:



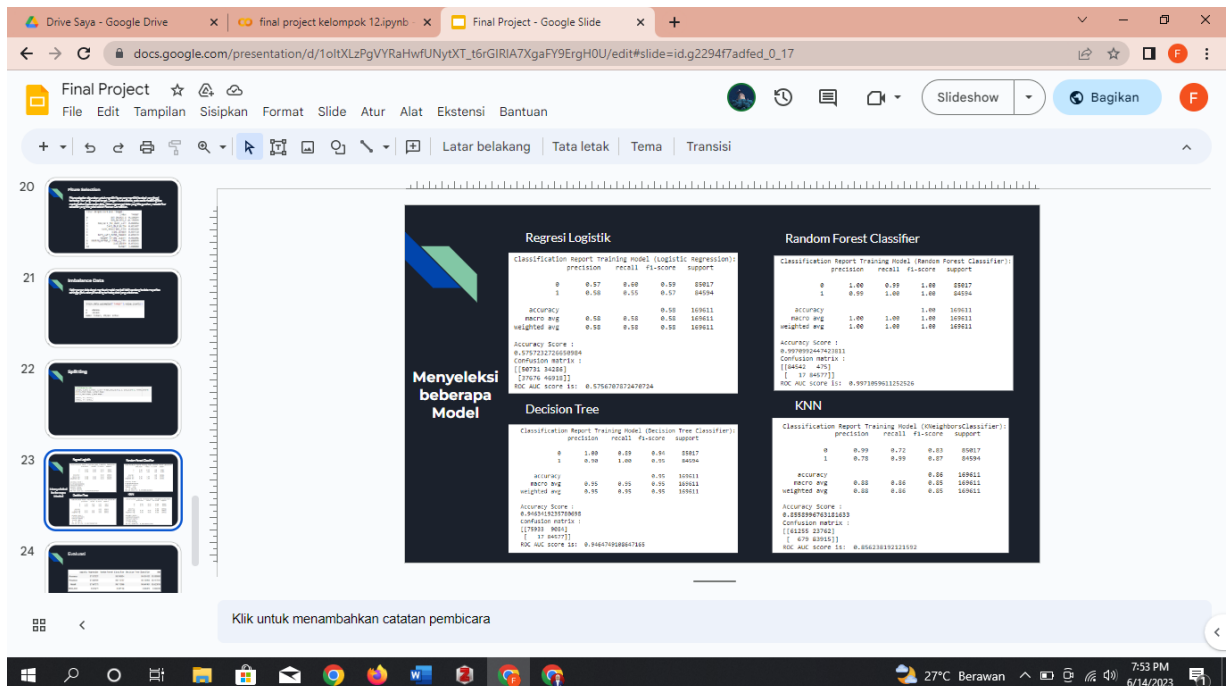
- Lampiran membuat bagian dari Dashboard berupa barplot yang menampilkan fitur yang korelasi tinggi terhadap target

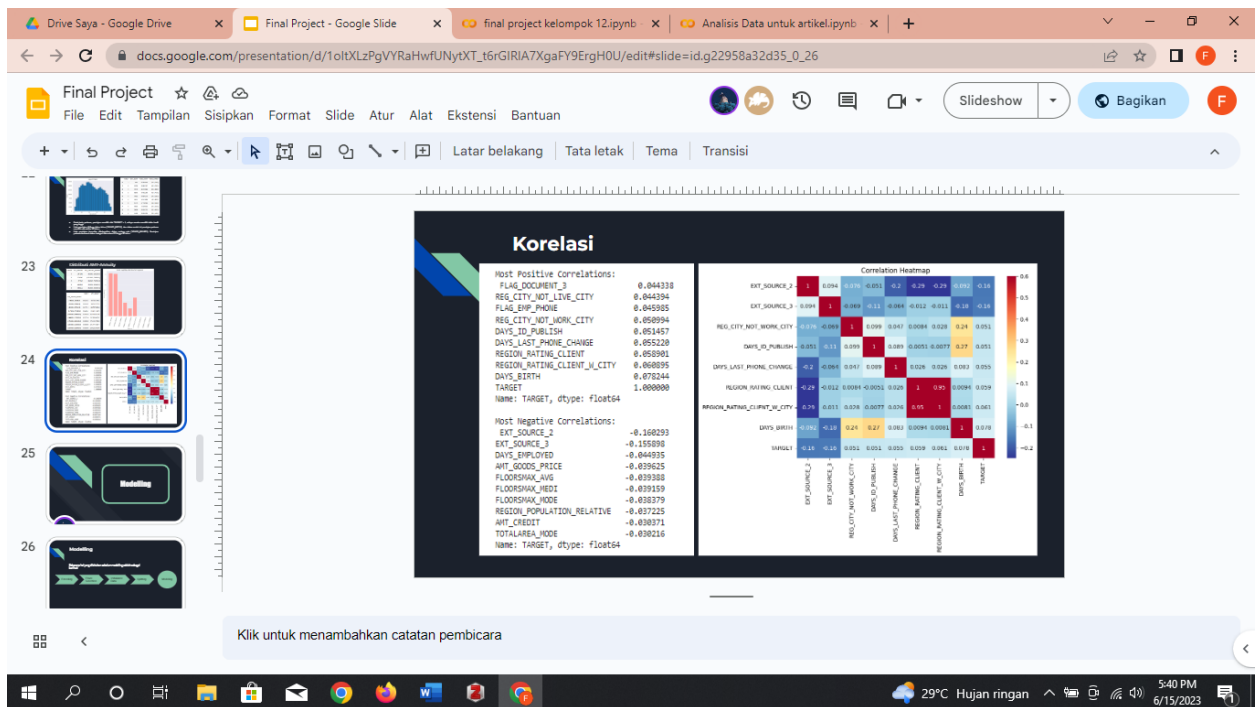
Screenshoot Kegiatan



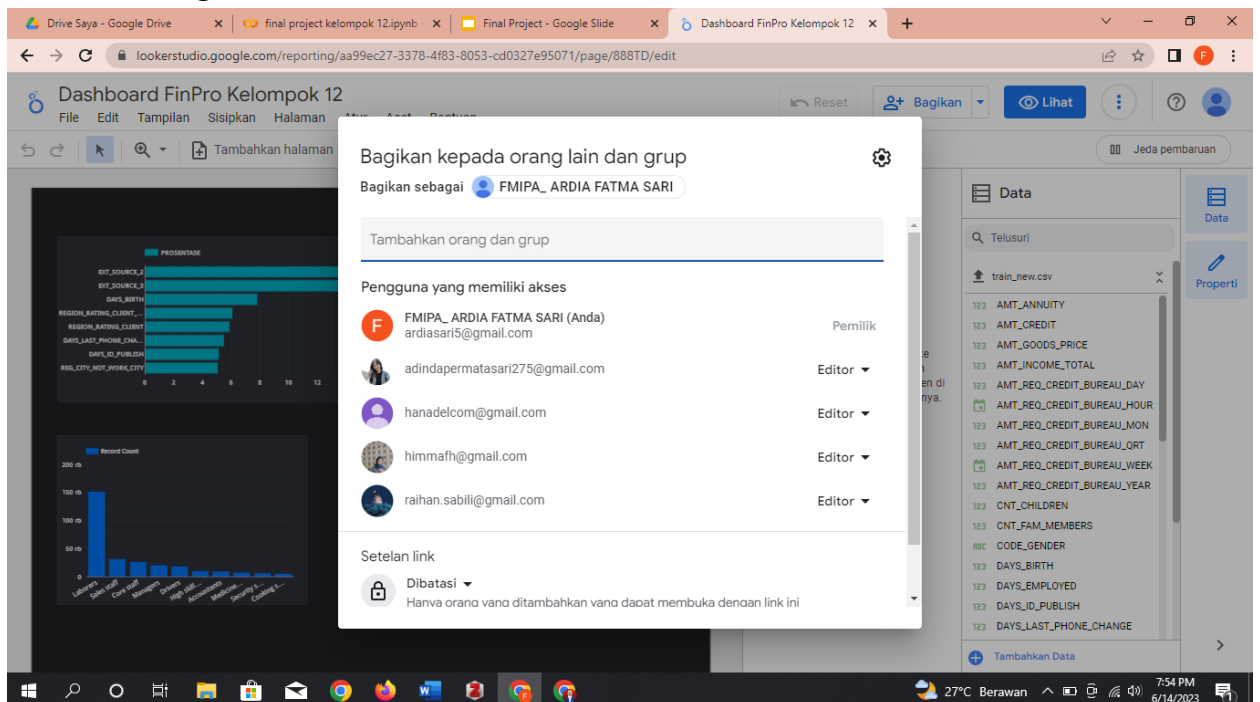
- Lampiran mengerjakan PPT bagian Modelling

Screenshoot Kegiatan:

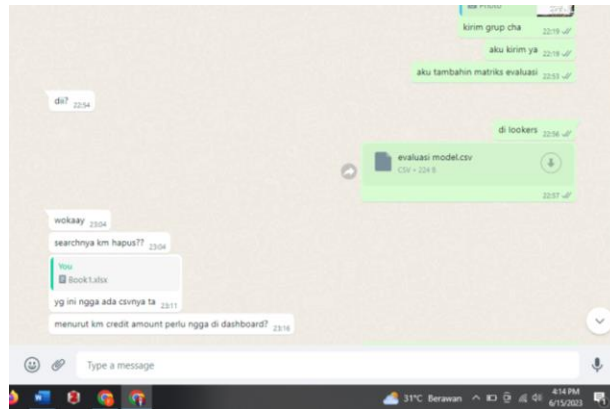




- Lampiran membuat link Looker Studio 1
- Screenshoot Kegiatan:**



- Membantu dalam pembuatan dashboard terkait apa yang harus disertakan
- Screenshoot Kegiatan:**



Proses Kerja

Dalam studi Independen PZSIB ini terdapat final project yang dikerjakan secara kelompok. Dalam final project ini, proses saya bekerja adalah sebagai berikut.

1. Saya sering meminta kelompok saya untuk melakukan diskusi terkait progress final project.
2. Sebelum melakukan tugas saya yaitu modelling, saya membantu kelompok dalam melakukan Exploratory Data Analysis (EDA). Proses ini dilakukan untuk memvisualisasikan beberapa data kategorik dan numerik. EDA digunakan untuk mengetahui seberapa data menyebar, dan mengetahui korelasi dari fitur yang ada. Dalam melakukan EDA, saya mencari beberapa informasi dari data application train dengan membentuk tabel dimana tabel tersebut berisi count dari target berdasarkan beberapa fitur yang ada. Selain itu, saya juga membuat barplot untuk melihat prosentase dan jumlah target==0 (nasabah yang tidak susah dalam mengembalikan pinjaman) dan target==1 (nasabah yang susah mengembalikan pinjaman) berdasarkan beberapa fitur yang saya tentukan. Saya juga membuat plot distribusi terkait days birth, income, dan annuity. Dari plot tersebut ada beberapa hal yang dapat diambil kesimpulan yang berkaitan dengan pemberian pinjaman kepada Loan.
3. Saya membantu dalam handling missing value. Ada beberapa hal yang saya lakukan yaitu melakukan checking prosentase missing value yang ada disetiap fitur. Untuk fitur yang prosentase missingnya >50% maka akan di drop. Setelah itu memisahkan antara variabel kategorik dan numerik Dalam pengisian data yang prosentase missing valuenya<50% maka ada beberapa ketentuan yang digunakan yaitu:
 - a. jika datanya kategorik maka akan diisi modus,
 - b. jika datanya numerik maka akan diisi dengan nilai median. Missing value pada data numerik diisi dengan nilai median karena datanya tidak berdistribusi normal, banyak diantara mereka yang right skew.
4. Saya melakukan modelling. Sebelum modelling ada beberapa hal yang dilakukan yaitu:

- a. imbalance data karena terdapat ketimpangan data,
- b. Encoding, karena dalam pemodelan dengan machine learning hanya dapat membaca dan mengolah data numerik,
- c. Feature selection, hal ini saya lakukan dengan menyeleksi feature mana saja yang korelasinya tinggi terhadap target,
- d. Splitting, dilakukan untuk memisahkan antara data train dan data test,
- e. Mencoba beberapa model diantaranya adalah Regresi Logistik, Random Forest Classifier, Decision Tree, dan KNN.
- f. Membandingkan keakuratan, nilai recall, f-score, dan roc-auc. Setelah membandingkannya dapat disimpulkan satu model terbaik yang digunakan dalam prediksi yaitu model random forest classifier dimana model tersebut memiliki keakuratan yang hampir sempurna dan nilai ROC-AUC nya paling tinggi dibandingkan model yang lain. Selain itu, dalam pemilihan model saya juga membandingkan dari hasil confusion matrix yang saya dapatkan dalam tahap modeling. Dari confusion matrix beberapa model yang saya seleksi, hasil FP dan FN dari model Random Forest Classifier menunjukkan nilai yang paling sedikit sehingga model tersebut dimungkinkan adalah model terbaik dari beberapa model yang diseleksi.

Dalam mengerjakan modeling saya mencari beberapa referensi dari github dan diskusi dengan beberapa teman saya.

5. Membantu dalam pembuatan dashboard. Saya membuat link Looker Studio 1 untuk membuat dashboard yang dikerjakan bersama. Awalnya link looker sudah dibuat namun ada kesalahan upload data kemudian saya mencoba membuat sendiri linknya dan mencoba upload datanya, data hasil download dari googlecolab dimana data tersebut adalah data bersih setelah handling missing value. Setelah itu, saya berhasil upload data kemudian share dan mendiskusikan dengan kelompok. Selain itu, saya membuat barplot dimana memberikan informasi terkait feature mana saja yang korelasinya tinggi terhadap target. Selain membuat barplot, saya juga membuat matrik evaluasi dalam dashboard. Sebelum membuat matrik evaluasi saya download hasil evaluasi dari beberapa model kemudian memasukkannya ke lookers studio. Matrik evaluasi yang ditampilkan adalah nilai maksimal dari nilai yang ada pada masing-masing kategori.
6. Selanjutnya adalah pembuatan PPT, yang saya lakukan adalah membuat PPT dengan memasukkan sebagian EDA dan materi - materi yang saya kerjakan dalam google colab, diantaranya adalah Encoding, Feature Selection, Imbalance Data, Splitting, Modelling, dan Evaluation.
7. Selain itu, di sela-sela waktu diskusi dan pengerjaan final project terdapat sesi LiveClass, dimana setiap LiveClass terdapat beberapa revisi terkait apa yang sudah dikerjakan. Saya sering sekali melakukan revisi beberapa hal dengan menambahkan penjelasan ataupun mengganti hasil pekerjaan.

Persetujuan Anggota

Laporan ini telah disetujui oleh:

- ☒ Adinda Permata Sari
- ☒ Himma Faicha Hubbiya
- ☒ Raihanah Assa'adah
- ☒ Raihan Tsabita Sabil