

STAGE 0



Kelompok: 2 (Synergies)

Stage: 0

Mentor: Faalih M Yusa

Pukul/ Tanggal: 18.30/5 Januari 2024

Pembagian tugas di stage ini:

Nama: Yusuf (PIC pengarahan dan schedule)

Nama: Mellia dan David (PIC notulen)

Nama: Alfath dan Agung (PIC laporan project)

Nama: Zaima (PIC slide presentasi)

Nama: Wulan (PIC sumber referensi)

Poin pembahasan:

1. Problem statement
2. Role
3. Goal
4. Objective
5. Business metrics

Hasil Diskusi:

1. Problem statement → diberi tambahan Bank RBC tidak mempunyai model untuk memprediksi nasabah mana yang akan melakukan churn.
2. Role → diberi tambahan sebagai team DS akan melakukan analisa data dan membuat model Machine Learning untuk mendeteksi nasabah mana yang akan melakukan churn.
3. Goal → membuat sebuah model Machine Learning yang dapat menunjukkan nasabah mana yang berpotensi churn sehingga dapat memberikan rekomendasi bisnis.
4. Objective → diberi tambahan melakukan optimisasi model sehingga dapat memberikan akurasi yang baik.
5. Business metrics → membahas tentang metrics sekunder yaitu akurasi dari model.

Tindak Lanjut:

Mencari sumber referensi toleransi maksimal nasabah churn di bank pada umumnya..

Mempersiapkan untuk stage 1 (berdasarkan arahan singkat dari mentor).



LAMPIRAN :

Daftar Kehadiran

- ☒ Faalih M Yusa (MENTOR)
- ☒ Burhanuddin Yusuf Robbani
- ☒ Mellia Anggreani
- ☒ David Melanius Nai
- ☒ Alfath Arrahman
- ☒ Moch Agung Laksono
- ☒ Dzul Wulan Ningtyas
- ☒ Zaima Syarifa Asshafa

STAGE 1



Kelompok: 2 (Synergies)

Stage: 1

Mentor: Faalih M Yusa

Pukul/ Tanggal: 18.30/17 Januari 2024

Pembagian tugas di stage ini:

Nama: Yusuf (PIC source code)

Nama: Mellia dan David (PIC notulen)

Nama: Alfath dan Agung (PIC laporan project)

Nama: Zaima (PIC slide presentasi)

Nama: Wulan (PIC Github)

Nama: Alfath, Mellia, David, Zaima (PIC insight)

Poin pembahasan:

1. EDA
2. Business insight
3. Rekomendasi bisnis
4. Github

Hasil Diskusi:

1. Cek ulang tulisan, pastikan tidak ada yang typo dan gambar dibuat per slide.
2. Akurasi machine learning yang ideal di atas 70% → ditambahkan di goal.
3. Korelasi paling tinggi antara exited dan umur → cari insight dari korelasi. Lakukan juga klasifikasi pada balance, bisa ditambahkan insight dari klasifikasi tersebut.
4. Github dibuat lebih detail, dengan perincian tiap stage.

Tindak Lanjut:

1. Perapian laporan final project.
2. Menambahkan insight berdasarkan arahan dari mentor.
3. Perapian github.
4. Persiapan stage berikutnya dengan memperhatikan hal-hal yang sudah dibahas (metode undersampling, age tidak perlu digrouping, menghilangkan kolom-kolom yang tidak penting, penentuan persentase data tes dan data train)



LAMPIRAN :

Daftar Kehadiran

- ☒ Faalih M Yusa (MENTOR)
- ☒ Burhanuddin Yusuf Robbani
- ☒ Mellia Anggreani
- ☒ David Melanius Nai
- ☒ Alfath Arrahman
- ☒ Moch Agung Laksono
- ☒ Dzul Wulan Ningtyas
- ☒ Zaima Syarifa Asshafa

STAGE 2



Kelompok: 2 (Synergies)

Stage: 2

Mentor: Faalih M Yusa

Pukul/ Tanggal: 18.40/1 Februari 2024

Pembagian tugas di stage ini:

Nama: Yusuf (PIC source code)

Nama: Mellia dan David (PIC notulen)

Nama: Alfath dan Agung (PIC laporan project)

Nama: Zaima (PIC slide presentasi)

Nama: Wulan (PIC Github)

Poin pembahasan:

1. Data cleansing (handle missing values, duplicated data, outliers, feature transformation, feature encoding dan handle class imbalance)
2. Feature engineering

Hasil Diskusi:

- 1.. Mengubah metrik sekunder dari akurasi menjadi F1.
2. Handling missing value → jangan hanya dicek nilai null-nya saja tetapi dicek apa ada data yang tidak sesuai, misalnya ada data "unknown" di kolom gender.
3. Mengganti metode label encoding pada kolom gender menjadi One Hot Encoding.
4. Mencoba grouping kolom geografi menjadi isGermany/NotGermany atau isFrance/NotFrance. Jika nilai VIF masih tinggi lakukan feature engineering dan handling outliers.
5. Proporsi yang ideal untuk data imbalance adalah 70:30, tetapi bisa dieksplorasi dengan mencoba proporsi 80:20.
6. Disarankan model menggunakan KNN karena sudah ada scaling, atau bisa juga SVM.

Tindak Lanjut:

1. Perapian laporan final project, diberi keterangan yang lengkap, misalnya jika terjadi underfitting maka akan dilakukan metode apa.
2. Mencoba menggunakan One Hot Encoding pada kolom gender.
3. Mencoba grouping di kolom geografi.
4. Mencoba mencari variabel baru lagi berdasarkan feature-feature yang sudah ada.
5. Persiapan untuk stage 3 dengan memperhatikan hal-hal seperti penentuan nilai k pada cross validation dan model apa yang digunakan.



LAMPIRAN :

Daftar Kehadiran

- ☒ Faalih M Yusa (MENTOR)
- ☒ Burhanuddin Yusuf Robbani
- ☒ Mellia Anggreani
- ☒ David Melanius Nai
- ☒ Alfath Arrahman
- ☒ Moch Agung Laksono
- ☒ Dzul Wulan Ningtyas
- ☒ Zaima Syarifa Asshafa

STAGE 3



Kelompok: 2 (Synergies)

Stage: 3

Mentor: Faalih M Yusa

Pukul/ Tanggal: 18.30/13 Februari 2024

Pembagian tugas di stage ini:

Nama: Yusuf dan Wulan (PIC source code)
Nama: Mellia dan David (PIC notulen)
Nama: Alfath dan Agung (PIC laporan project)
Nama: Zaima (PIC slide presentasi)
Nama: Wulan (PIC Github)

Poin pembahasan:

1. Algoritma yang dipakai
2. Hyperparameter tuning
3. Feature importance dan business insight

Hasil Diskusi:

- 1.. Menggunakan metode undersample (setelah feature engineering) dan membandingkan hasil antara sebelum dan sesudah hyperparameter tuning.
2. Pilih 3 algoritma berdasarkan F1 score (sementara yang diambil Adaboost, KNN, Logistic Regression). F1 score yang bagus 0,75-0,85.
3. Tampilkan langkah-langkah proses modelling dan alasan pemilihan model. Tampilkan juga confusion matrixnya.
4. Hyperparameter tuning memakai GridSearch, ambil 3 parameter saja
5. Feature importance lebih ke arah bisnisnya saja dan business insight harus berhubungan dengan problem.
6. Kelebihan adaboost : bagus untuk underfitting yang datanya imbalance (walaupun sudah di undersample)

Tindak Lanjut:

1. Melakukan metode undersample dan di hyperparameter tuning, bandingkan F1 scorenya dengan yang tanpa hyperparameter tuning.
2. Pilih 3 algoritma yang performa paling tinggi, menengah dan paling rendah/paling sederhana.
3. Menjelaskan alasan pemilihan model (cari kelebihan dan kekurangan) dan menulis step by step.
4. Mencari business insight → pastikan sesuai dengan problem dan goalnya.
5. Persiapan untuk stage 4, persiapan untuk presentasi.



LAMPIRAN :

Daftar Kehadiran

- ✓ Faalih M Yusa (MENTOR)
- ✓ Burhanuddin Yusuf Robbani
- ✓ Mellia Anggreani
- ✓ David Melanius Nai
- ✓ Alfath Arrahman
- ✓ Moch Agung Laksono
- ✓ Dzul Wulan Ningtyas
- ✓ Zaima Syarifa Asshafa

STAGE 4



Kelompok: 2 (Synergies)

Stage: 4

Mentor: Faalih M Yusa

Pukul/ Tanggal: 18.30/28 Februari 2024

Pembagian tugas di stage ini:

Nama: Yusuf dan Wulan (PIC source code)

Nama: Mellia dan David (PIC notulen)

Nama: Alfath dan Agung (PIC slide presentasi)

Nama: Zaima (PIC presentasi)

Poin pembahasan:

1. Goals
2. Objective
3. Rekomendasi bisnis berdasarkan modeling
4. Isi Keseluruhan PPT (Teknis Persiapan Presentasi)

Hasil Diskusi:

- 1.. Slide PPT di persingkat -> Menggabungkan beberapa slide menjadi 1 slide gabungan.
2. Diberi keterangan berupa poin-poin penting nya saja pada masing-masing slide.
3. Karena parameter yang digunakan hanya F1 Score, Recall, ROC AOC, maka beri kotak merah untuk menghighlight point tersebut.
4. Dari 7 model yang dibuat, hanya 3 model terbaik yang dipresentasikan.
5. Untuk random forest hanya tampilkan 3 teratas

Tindak Lanjut:

1. Business rekomendasi ditambahkan strategi pemasaran.
2. Tambahkan confusion matrix 3 model terbaik.
3. Tambahkan alasan kenapa memilih 3 model tsb.
4. Perkuat rekomendasi bisnisnya



LAMPIRAN :

Daftar Kehadiran

- ✓ Faalih M Yusa (MENTOR)
- ✓ Burhanuddin Yusuf Robbani
- ✓ Mellia Anggreani
- ✓ David Melanius Nai
- ✓ Alfath Arrahman
- ✓ Moch Agung Laksono
- ✓ Dzul Wulan Ningtyas
- ✓ Zaima Syarifa Asshafa