## Laporan Kerja Individu

# Final Project - Data Analytics Program Zenius Studi Independen Bersertifikat - Angkatan 4

Nama Lengkap: Zulfa Nabilah Nurvitasari

Nomor ID live class: 099 Nomor Kelompok: 30 Mentor: Erwin Fernanda

#### Deskripsi peran

- Melakukan cleansing dan preparation pada dataset credit card balance dan previous application
- Melakukan cleansing dan preparation pada dataset application train (Semua anggota melakukan cleansing dan preparation. Oleh karena itu, dari semua bagian cleansing dan preparation data yang telah saya lakukan, yang digunakan ialah bagian data encoding.
- Melakukan pemodelan dengan menggunakan *Supervised Learning* menggunakan KNN, *logistic regression*, dan *random forest*
- Membuat dashboard dengan Looker

#### Lampiran Hasil Kerja

- Cleansing dan Preparation pada dataset credit card balance dan previous application: <a href="https://colab.research.google.com/drive/1wyx2ypaaOnYMmLx5s3dTkE0JF76isL6y?u">https://colab.research.google.com/drive/1wyx2ypaaOnYMmLx5s3dTkE0JF76isL6y?u</a> sp=sharing
- Cleansing, Preparation, dan modeling pada dataset application train:
   https://colab.research.google.com/drive/1ki0chl6qm01JWFk0\_sUftZ\_e372jSsmc?us
   p=sharing
- Dashboard: <a href="https://lookerstudio.google.com/reporting/b9b058f8-df07-4f8f-ac71-63d985fb2455">https://lookerstudio.google.com/reporting/b9b058f8-df07-4f8f-ac71-63d985fb2455</a>

# **Proses Kerja**

No	Hari, Tanggal	Pekerjaan	Keterangan		
1	Rabu, 31 Mei	Cleansing dan preparation	Saya melakukan data understanding		
		pada dataset <i>credit card</i>	seperti melihat jumlah kolom dan		
		balance.	baris pada data, melihat kolom pada		
			dataset, melihat tipe data, melihat		
			jumlah missing value di setiap		
			kolom, dan melihat frekuensi setiap		
			unique values pada data.		
2	Kamis, 1 Juni	Cleansing dan preparation	Saya melakukan <i>missing value</i>		
		pada dataset <i>credit card</i>	handling pada data bertipe numerik		
		balance.	dengan mengisi <i>missing value</i>		
			menggunakan nilai median karena		
			tidak berdistribusi normal. Selain itu,		
			saya juga melakukan <i>data encoding</i>		
			dengan menggunakan frequency		
			encoding.		
3	Jumat, 2 Juni	Cleansing dan preparation	Saya melakukan data understanding		
		pada dataset <i>previous</i>	seperti melihat jumlah kolom dan		
		application.	baris pada data, melihat kolom pada		
			dataset, melihat tipe data, melihat		
			jumlah missing value di setiap		
			kolom, dan melihat frekuensi setiap		
			unique values pada data.		
4	Sabtu, 3 Juni	Cleansing dan preparation	Saya melakukan <i>missing value</i>		
		pada dataset <i>previous</i>	handling pada data bertipe numerik		
		application.	dan kategorik. Saya menghapus		
			baris dengan kolom yang memiliki		
			missing value kurang dari 10 karena		
			jumlah tidak terlalu besar. Selain itu,		
			saya juga menghapus kolom dengan		
			missing value lebih dari 1000000.		
			Untuk data bertipe numerik yang		
			memiliki <i>missing value</i> di bawah		
			1000000 saya isi dengan mean untuk		
			kolom yang memiliki distribusi		
			normal. Sedangkan, missing value		
			pada kolom-kolom yang tidak		
			memiliki distribusi normal, saya isi		
			dengan median. Untuk data bertipe		

		T	kategorik, <i>missing value</i> saya isi		
	DA' A I '	Classica de accessorio	dengan modus.		
5	Minggu, 4 Juni	Cleansing dan preparation	Saya melakukan data encoding pada		
		pada dataset <i>previous</i>	kolom-kolom bertipe data kategori.		
		application.	Untuk kolom-kolom yang memiliki		
			unique value kurang dari 5, saya		
			melakukan <i>label encoding</i> ,		
			sedangkan untuk kolom-kolom yang		
			memiliki <i>unique value</i> lebih dari 5		
			dilakukan frequency encoding.		
6	Senin, 5 Juni	Cleansing dan preparation	Saya melakukan data understanding		
		pada dataset application	seperti melihat jumlah kolom dan		
		train	baris pada data, melihat tipe data,		
			melihat statistik deskriptif pada		
			data, melihat jumlah missing value di		
			setiap kolom, melihat jumlah		
			duplikat, melihat frekuensi setiap		
			unique values pada data, dan juga		
			melakukan pengecekan		
			keseimbangan data. Selain itu, saya		
			juga melakukan unusual value		
			handling dengan mengganti salah		
			satu unique value pada kolom		
			bertipe data kategori dengan		
			modus.		
7	Selasa, 6 Juni	Cleansing dan preparation	Saya melakukan <i>missing value</i>		
′	Sciasa, o Jain	pada dataset application	handling pada data bertipe numerik		
		train	dan kategorik. Saya menghapus		
		truin	, , ,		
			baris dengan kolom yang memiliki		
			missing value kurang dari 1000		
			karena jumlah tidak terlalu besar.		
			Selain itu, saya juga menghapus		
			kolom dengan missing value lebih		
			dari 70%. Untuk data bertipe		
			numerik yang memiliki <i>missing value</i>		
			di bawah 70% saya isi dengan mean		
			untuk kolom-kolom yang memiliki		
			distribusi normal. Sedangkan,		
			<i>missing value</i> pada kolom-kolom		
			yang tidak memiliki distribusi		
			normal, saya isi dengan median.		

memiliki missing value di atas 50% saya isi dengan 'unk', sedangkan untuk kolom yang memiliki missing value di bawah 50% saya isi dengan modus.  8 Rabu, 7 Juni Cleansing dan preparation pada dataset application train Saya melakukan data encoding pada kolom-kolom bertipe data kategori. Untuk kolom-kolom yang memiliki unique value kurang dari 5, saya melakukan label encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  9 Jumat, 9 Juni Modeling Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= 0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan spilit train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan nenggunakan nenggunakan undersompling. Setelah melakukan undersompling.				Untuk data bertipe kategorik yang
untuk kolom yang memiliki missing value di bawah 50% saya isi dengan modus.  8 Rabu, 7 Juni  Cleansing dan preparation pada dataset application train  Saya melakukan data encoding pada kolom-kolom bertipe data kategori. Untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  9 Jumat, 9 Juni  Modeling  Untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding. Sebelum melakukan modeling, saya mengambil kolom-kolom yang memiliki inilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya memiliki inilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X train dan Y train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan fandom Forest dan KNN pada data yang telah di sampling dengan menggunakan				memiliki <i>missing value</i> di atas 50%
Rabu, 7 Juni  Cleansing dan preparation pada dataset application train  Cleansing dan preparation pada kolom-kolom bertipe data kategori. Untuk kolom-kolom yang memiliki unique value kurang dari 5, saya melakukan label encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  Sebelum melakukan modeling, saya mengam kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahulu performa pada model.  Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan monggunakan menggunakan menggunaka				-
Rabu, 7 Juni  Cleansing dan preparation pada dataset application train  Untuk kolom-kolom bertipe data kategori. Untuk kolom-kolom yang memiliki unique value kurang dari 5, saya melakukan label encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  Jumat, 9 Juni  Modeling  Modeling  Modeling  Modeling  Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengiliki nilai korelasi dengan TARGET sebesar >= 0.01 an sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X tarena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X train dan Y train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  Sabtu, 10 Juni  Modeling  Modeling  Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				untuk kolom yang memiliki <i>missing</i>
Rabu, 7 Juni  Cleansing dan preparation pada dataset application train  Cleansing dan preparation pada kolom-kolom bertipe data kategori. Untuk kolom-kolom yang memiliki unique value kurang dari 5, saya melakukan label encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  Jumat, 9 Juni  Modeling  Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambili kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  Modeling  Modeling  Sabtu, 10 Juni  Modeling  Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan men				value di bawah 50% saya isi dengan
pada dataset application train  kolom-kolom bertipe data kategori. Untuk kolom-kolom yang memiliki unique value kurang dari 5, saya melakukan label encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  9 Jumat, 9 Juni  Modeling  Modeling  Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan men				modus.
train  Untuk kolom-kolom yang memiliki unique value kurang dari 5, saya melakukan label encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan	8	Rabu, 7 Juni	Cleansing dan preparation	Saya melakukan data encoding pada
unique value kurang dari 5, saya melakukan label encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  9 Jumat, 9 Juni Modeling Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunaka			pada dataset application	kolom-kolom bertipe data kategori.
melakukan label encoding, sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  9 Jumat, 9 Juni Modeling Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antat tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan			train	Untuk kolom-kolom yang memiliki
sedangkan untuk kolom-kolom yang memiliki unique value lebih dari 5 dilakukan frequency encoding.  9 Jumat, 9 Juni Modeling Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				unique value kurang dari 5, saya
memiliki unique value lebih dari 5 dilakukan frequency encoding.  9 Jumat, 9 Juni Modeling Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling dengan menggunakan mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				melakukan label encoding,
memiliki unique value lebih dari 5 dilakukan frequency encoding.  9 Jumat, 9 Juni Modeling Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling dengan menggunakan mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				sedangkan untuk kolom-kolom yang
dilakukan frequency encoding.  9 Jumat, 9 Juni Modeling Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar > 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
9 Jumat, 9 Juni  Modeling  Sebelum melakukan modeling, saya mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling  Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				•
mencari nilai korelasi antar semua kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling  Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan	9	lumat, 9 luni	Modelina	
kolom dengan kolom TARGET terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan		samae, s sam	- modemig	
terlebih dahulu. Setelah itu, saya mengambil kolom-kolom yang memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan				kolom dengan kolom TARGET
memiliki nilai korelasi dengan TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya menghitung metric evaluasi untuk mengetahui performa pada model. Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan menggunakan				_
TARGET sebesar >= 0.01 dan sebesar <= -0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling  Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan menggunakan				mengambil kolom-kolom yang
<=-0.01. Selanjutnya, saya membagi data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model. Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan				memiliki nilai korelasi dengan
data menjadi X (Independen) dan Y (Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan menggunakan				
(Dependen) dan melakukan standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan				
standarisasi pada data X karena setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan				
setiap kolomnya memiliki satuan yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
yang berbeda. Setelah itu, saya melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan menggunakan				
melakukan split train test pada data. Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
Tahap selanjutnya, yaitu melakukan undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				, .
undersampling dan oversampling pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
pada data X_train dan Y_train karena ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
ada ketidakseimbangan antar data tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
tersebut. Setelah itu, saya melakukan modeling dengan menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
menggunakan logistic regression. Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
Setelah melakukan modeling, saya menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
menghitung metric evaluasi untuk mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				menggunakan logistic regression.
mengetahui performa pada model.  10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				Setelah melakukan modeling, saya
10 Sabtu, 10 Juni Modeling Saya melakukan modeling dengan menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
menggunakan Random Forest dan KNN pada data yang telah di sampling dengan menggunakan				
KNN pada data yang telah di sampling dengan menggunakan	10	Sabtu, 10 Juni	Modeling	
sampling dengan menggunakan				
				KNN pada data yang telah di
undersampling. Setelah melakukan				sampling dengan menggunakan
,				undersampling. Setelah melakukan
modeling, saya menghitung metric				modeling, saya menghitung metric

			evaluasi untuk mengetahui		
			performa pada model.		
11	Minggu, 11 Juni	Modeling	Saya melakukan modeling dengan		
			menggunakan Random Forest dan		
			KNN pada data yang telah di		
			sampling dengan menggunakan		
			oversampling. Setelah melakukan		
			modeling, saya menghitung metric		
			evaluasi untuk mengetahui		
			performa pada model.		
12	Rabu, 14 Juni	Membuat Dashboard	Membuat pie chart client by target,		
			score card jumlah client dengan		
			target 1 dan 0, membuat dua tabel		
			jumlah client dengan target 1 dan 0		
			berdasarkan <i>education type</i> dan		
			organization type, membuat dua		
			barchart jumlah client dengan target		
			1 dan 0 berdasarkan <i>contract type</i>		
			dan gender.		
13	Kamis, 15 Juni	Membuat Dashboard	Menampilkan tabel metric		
			evaluation yang terdiri dari		
			precision, recall, f1-score, dan		
			support. Selain itu, saya juga		
			menampilakan kurva ROC dan nilai		
			ROC AUC pada model <i>logistic</i>		
			regression tanpa sampling, dengan		
			SMOTETomek, dan over sampling		
			serta pada model random forest		
			tanpa <i>sampling,</i> dengan		
			SMOTETomek, dan over sampling.		

### Persetujuan Anggota

Laporan	ini ta	lah	dicati	iirii	Alah.
Laboran	1111111	าเสเเ	msen	11111	OICH.

- ☑ Christ Jordan Baeha
- ☑ Zulfa Nabilah Nurvitasari
- ☑ Dinny Meilinda Sari
- ☐ Sania Salsabila Agustin
- ☐ Fiqih Imanul Haq