Rakamin Academy

JUPYTHON

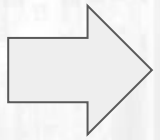
Dokumen Laporan Final Project

STAGE 2 (PREPARATION)

10 April – 16 APRIL 2023

Data Cleansing

LoanNr_ChkDgt	0
Name	14
City	30
State	14
Zip	0
Bank	1559
BankState	1566
NAICS	0
ApprovalDate	0
ApprovalFY	0
Term	0
NoEmp	0
NewExist	136
CreateJob	0
RetainedJob	0
FranchiseCode	0
UrbanRural	0
RevLineCr	4528
LowDoc	2582
DisbursementDate	2368
DisbursementGross	0
BalanceGross	0
MIS_Status	1997
ChgOffPrinGr	0
GrAppv	0
SBA_Appv	0



LoanNr_ChkDgt	0
Name	0
City	0
State	0
Zip	0
Bank	0
BankState	0
NAICS	0
ApprovalDate	0
ApprovalFY	0
Term	0
NoEmp	0
NewExist	0
CreateJob	0
RetainedJob	0
FranchiseCode	0
UrbanRural	0
RevLineCr	0
LowDoc	0
DisbursementDate	0
DisbursementGross	0
BalanceGross	0
MIS_Status	0
ChgOffPrinGr	0
GrAppv	0
SBA_Appv	0

dtype: int64

Handle Duplicated Data

```
## Memeriksa data duplicated
df.duplicated().any()
```

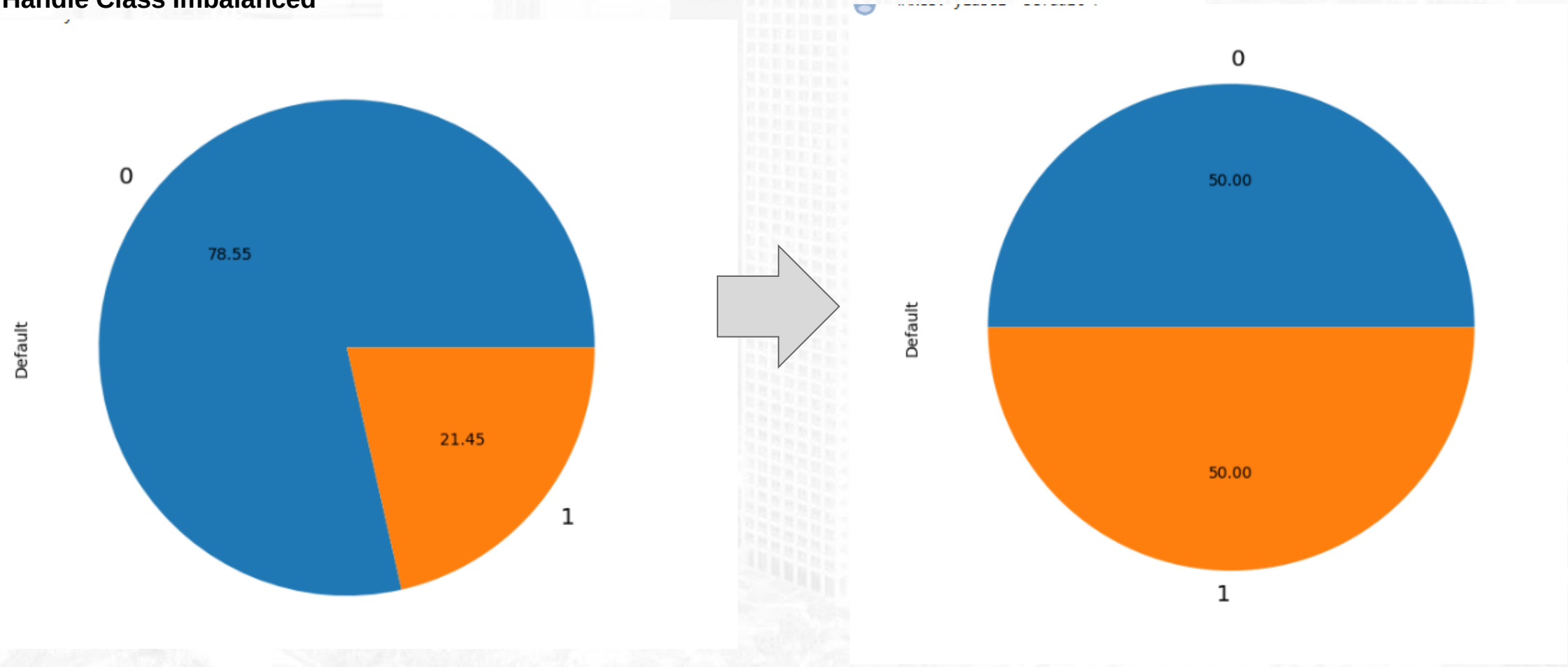
```
False
```

Handle Class Imbalanced

Kami menggunakan random oversampling + undersampling dengan imblearn. Selain untuk menyeimbangkan data kita gunakan random over dan under sampling agar kita bisa meningkatkan sampel kelas minoritas sama dengan kelas mayoritas lain.

Data Cleansing

Handle Class Imbalanced



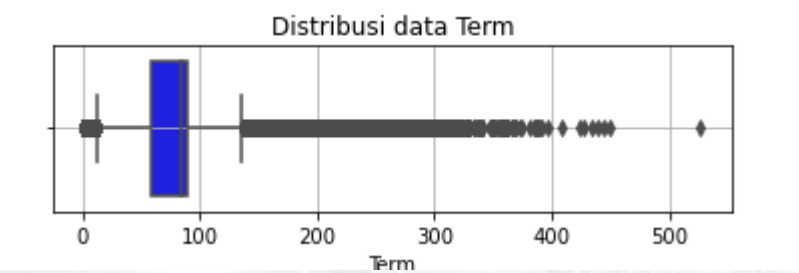
Data Cleansing

Feature Transformation, Feature Encoding, Feature Transformation

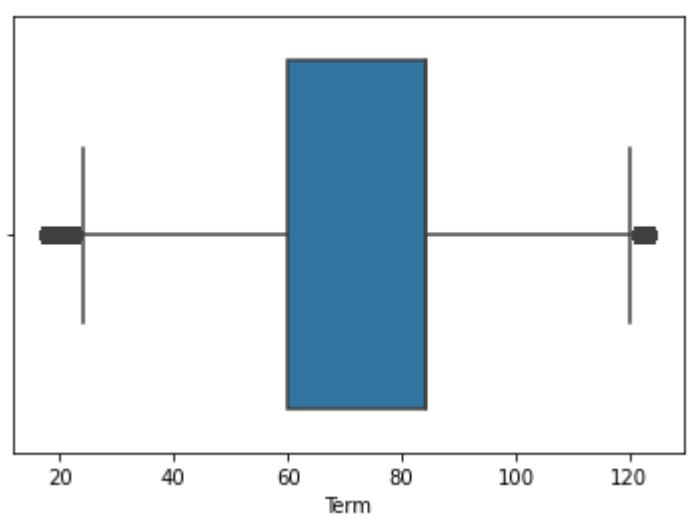
State	Categorical	Feature Selection
BankState	Categorical	Feature Selection
ApprovalFY	Numerical	Handle Outlier, Feature Transformation
Term	Numerical	Handle Outlier, Feature Transformation
NoEmp	Numerical	Handle Outlier, Feature Transformation
CreateJob	Numerical	Handle Outlier, Feature Transformation
RetainedJob	Numerical	Handle Outlier, Feature Transformation
UrbanRural	Categorical	Feature Encoding
RevLineCr	Categorical	Feature Encoding
LowDoc	Categorical	Feature Encoding
DisbursementGross	Numerical	Handle Outlier, Feature Transformation
GrAppv	Numerical	Handle Outlier, Feature Transformation

Data Cleansing

Handle Outlier



handle outlier mengguna kan metode IQR



Feature Engineering

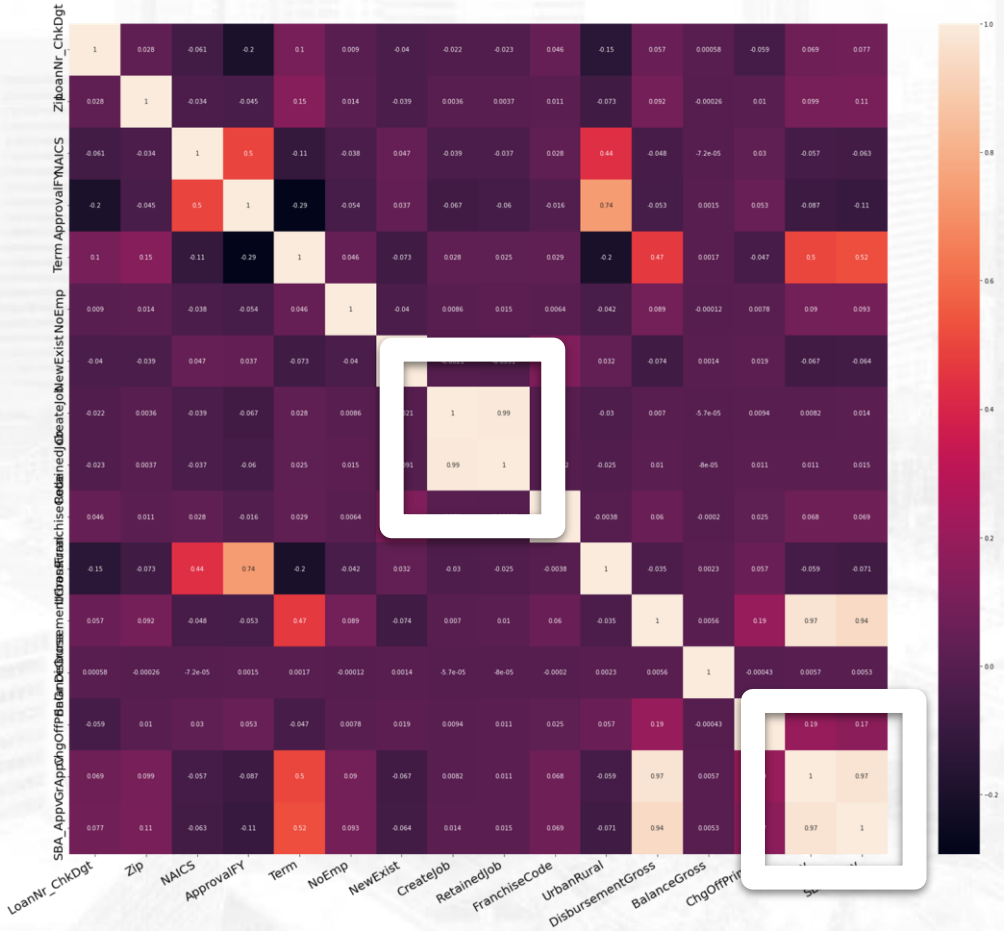
Feature Selection (membuang feature yang kurang relevan atau redundan)

LoanNr_ChkDgt	Categorical	Feature selection - Drop Feature
Name	Categorical	Feature selection - Drop Feature
City	Categorical	Feature selection - Drop Feature
Zip	Categorical	Feature selection - Drop Feature
Bank	Categorical	Feature selection - Drop Feature
NAICS	Categorical	Feature selection - Drop Feature
ApprovalDate	Timestamp	Feature selection - Drop Feature
NewExist	Categorical	Feature selection - Drop Feature
FranchiseCode	Categorical	Feature selection - Drop Feature
DisbursementDate	Timestamp	Feature selection - Drop Feature
BalanceGross	Numerical	Feature selection - Drop Feature
MIS_Status	Categorical	Feature selection - Drop Feature
ChgOffPrinGr	Numerical	Feature selection - Drop Feature
SBA_Appv	Numerical	Feature selection - Drop Feature

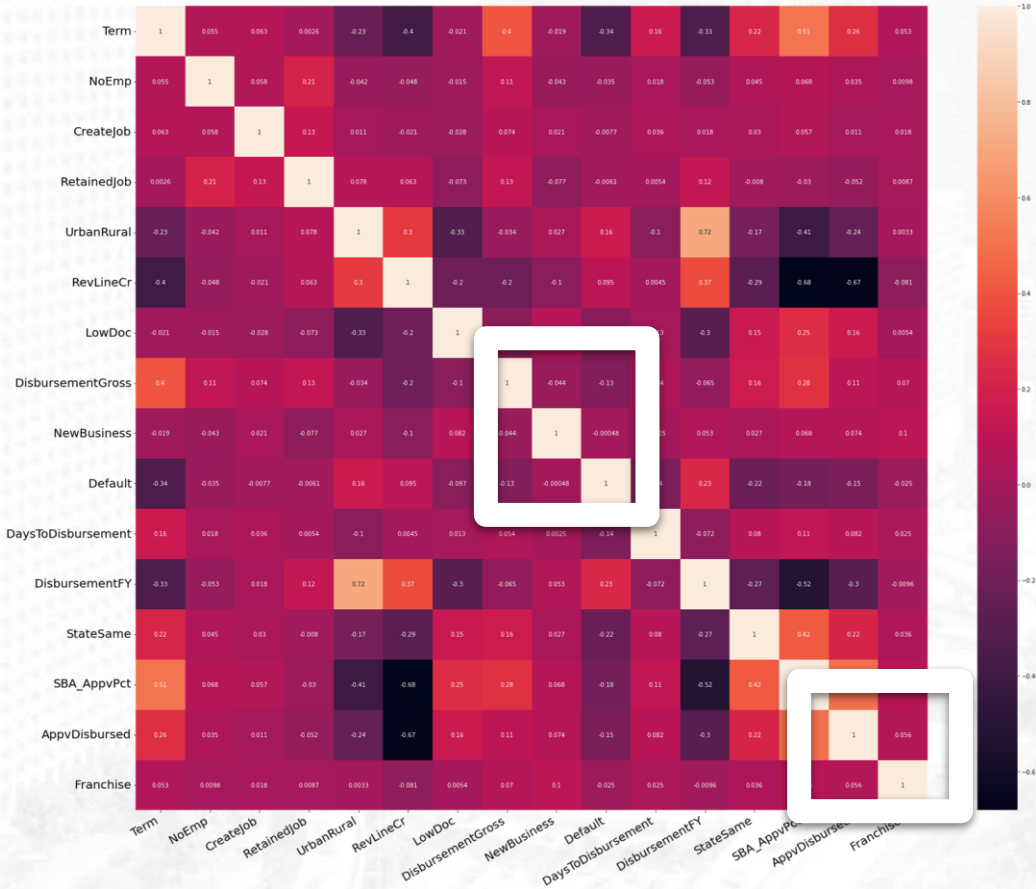
NewBusiness	Categorical	Feature Extraction : untuk mengetahui apakah nasabah peminjam merupakan bisnis baru atau bisnis lama.
Default	Categorical	Feature Extraction : persentase jumlah yang di jamin oleh SBA
DaysToDisbursement	Numerical	Feature Extraction : untuk mengetahui berapa lama pencairan setelah disetujui (dalam hari)
DisbursementFY	Timestamp	Feature Extraction : untuk mengatahui tahun pencairan
StateSame	Categorical	Feature Extraction : untuk mengetahui lokasi nasabah dengan bank apakah berada di satu lokasi atau tidak
SBA_AppvPct	Numerical	Feature Extraction : persentase jumlah yang di jamin oleh SBA
AppvDisbursed	Categorical	Feature Extraction : untuk mengetahui persentasi pinjaman yang cair , jika 100% = 1, jika tidak =0
Franchise	Categorical	Feature Extraction : binari 0, 1 (0 = Bukan Franchise, 1 = Franchise)
industri	Categorical	Feature Extraction : dari feature NAICS untuk mengeluarkan klasifikasi jenis industri dari yang sebelumnya menggunakan kode

Feature Engineering

Feature Selection (membuang feature yang kurang relevan atau redundan)



Feature Selection, dropping features with high correlation



Link Git Hub

https://github.com/gustiayuseptiandani/Homework-Jupython.git

Link Google Colab

https://colab.research.google.com/drive/1Mip8OxNY6VkmifQHrHoGGjwBaunCn1n9?usp=sharing#scrollTo=w_xs9LvcNKwa