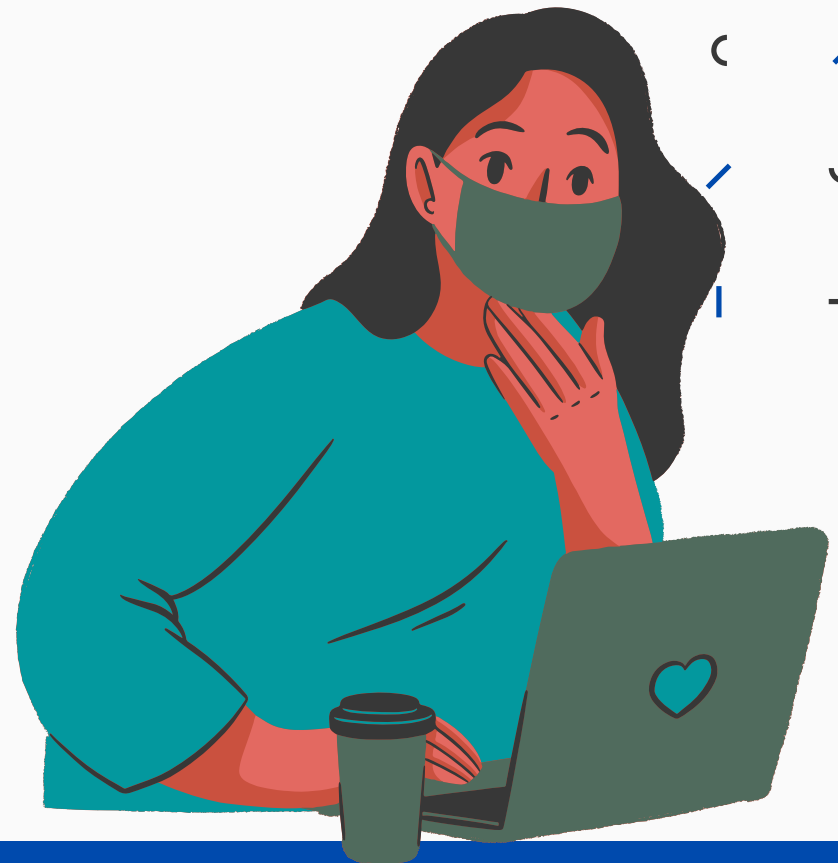


Credit Loan Predictions

MAYA MARYANAH



Virtual Internship Experience ID/X Partners

Junior Data Scientist

MAYA MARYANAH

Data Science Trainee at Jakarta Smart City



**PROJECT
OVERVIEW**

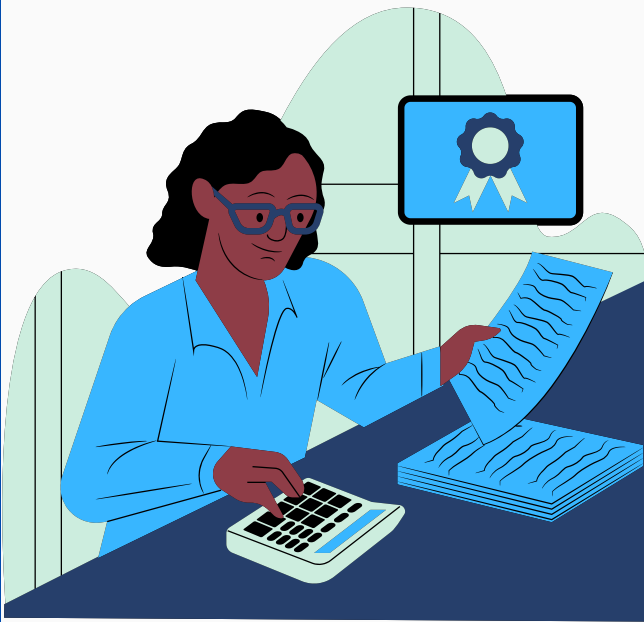
STEP 1
**Exploratory
Data
Analysis**

STEP 2
**Data
Preprocessing**

STEP 3
**Modeling
&
Evaluation**

**Explain Data
Model**

Contents

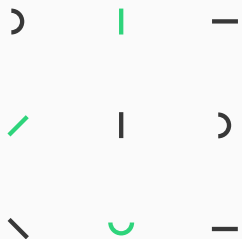




OVERVIEW

Build model which can predict credit risk based on previous data.

This dataset contains data since 2014 until 2017.



Exploratory Data Analysis

01

04

› | –
/ | ›
\ ∪ –



ITEM 1

Presentations are communication tools.

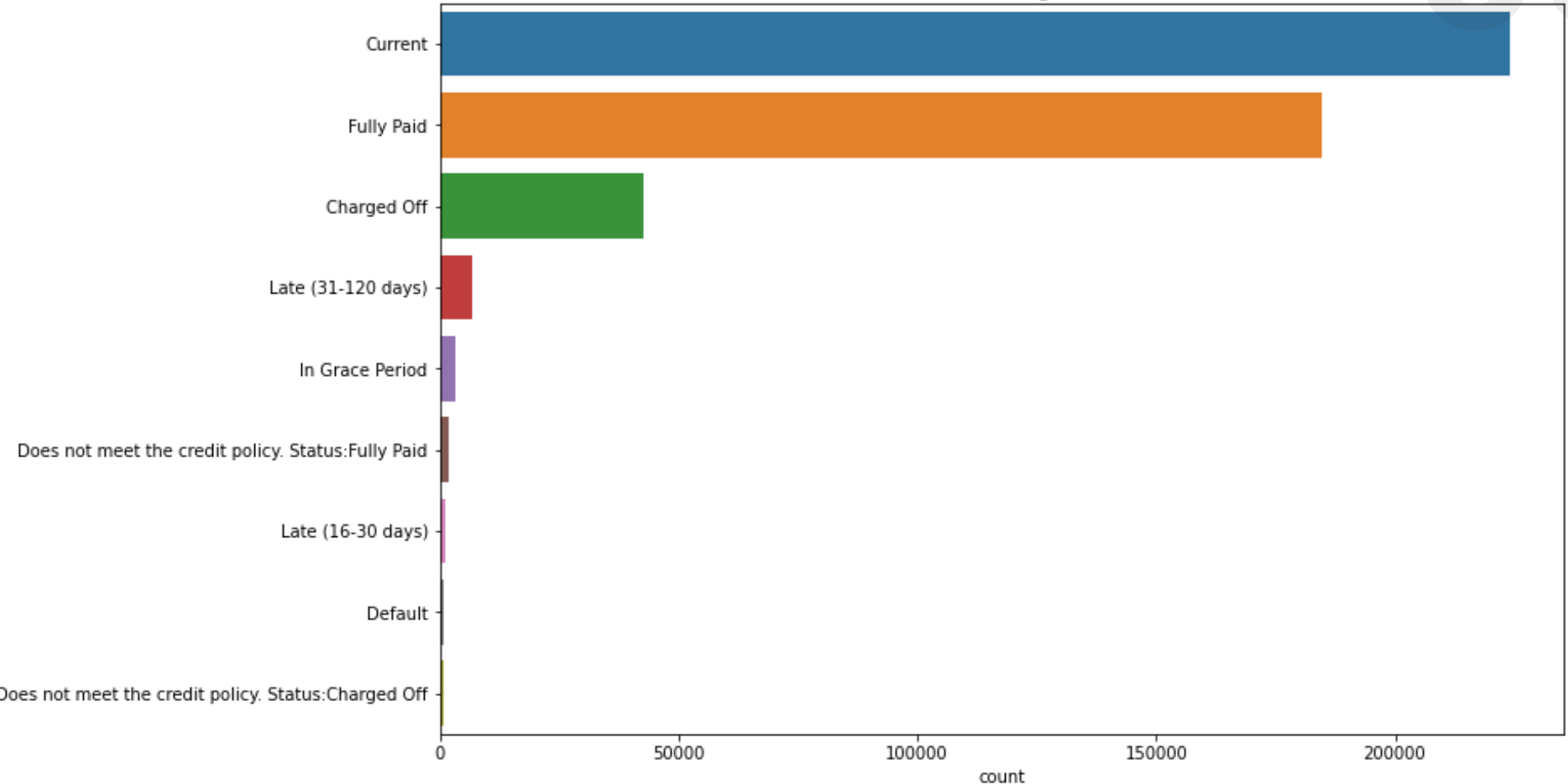
ITEM 2

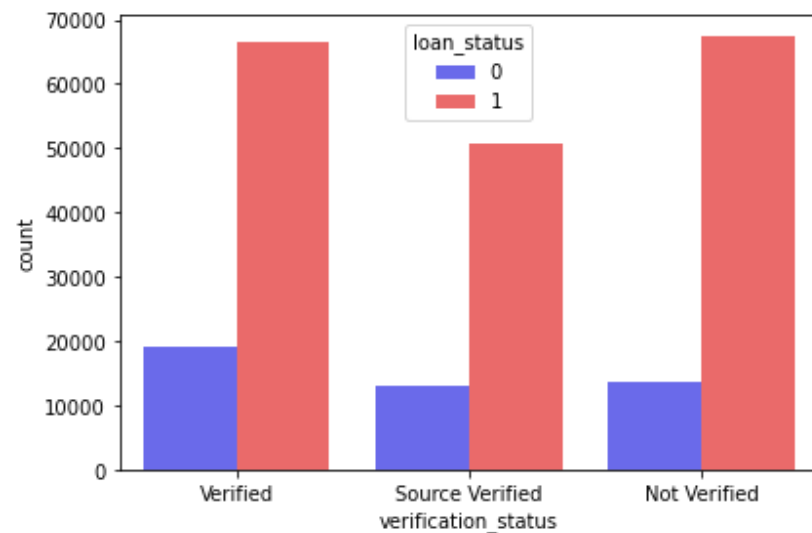
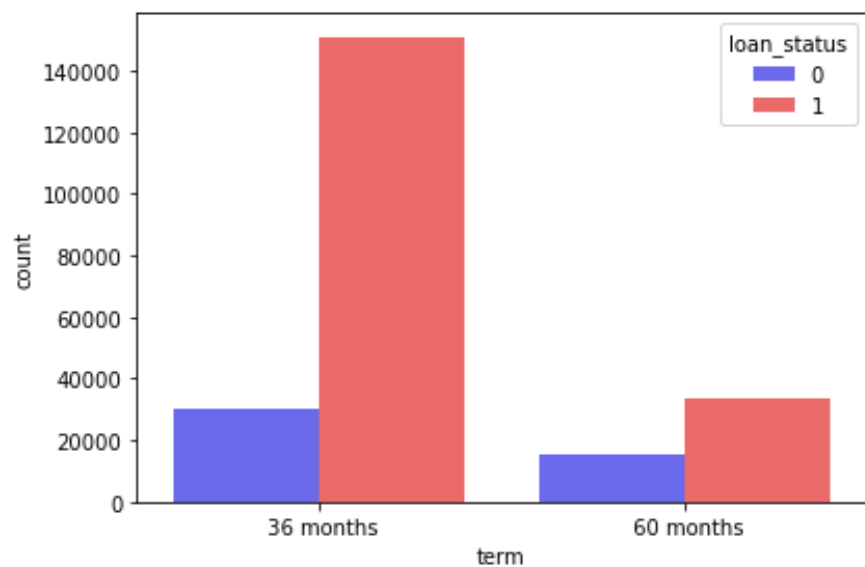
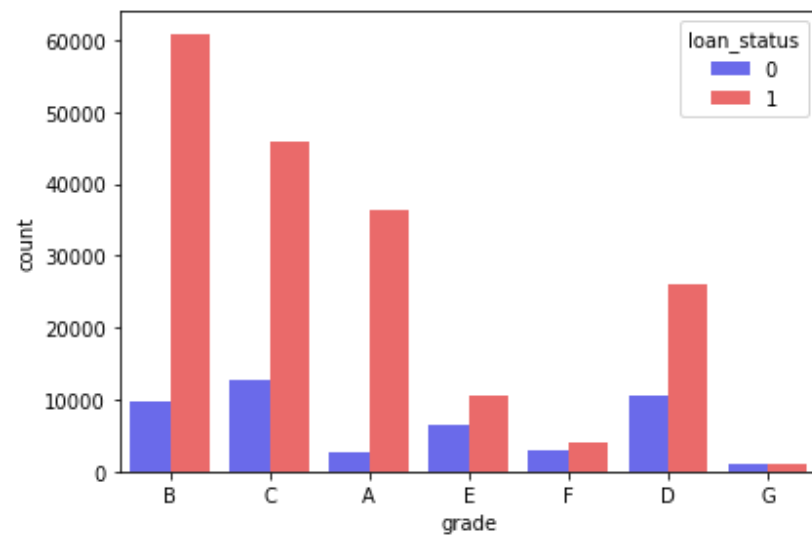
Presentations are communication tools.

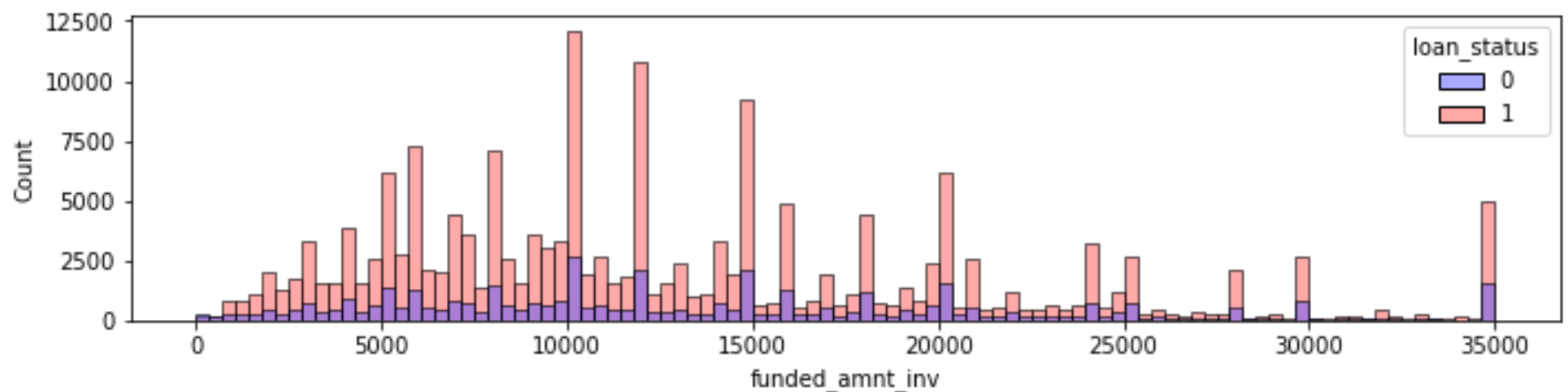
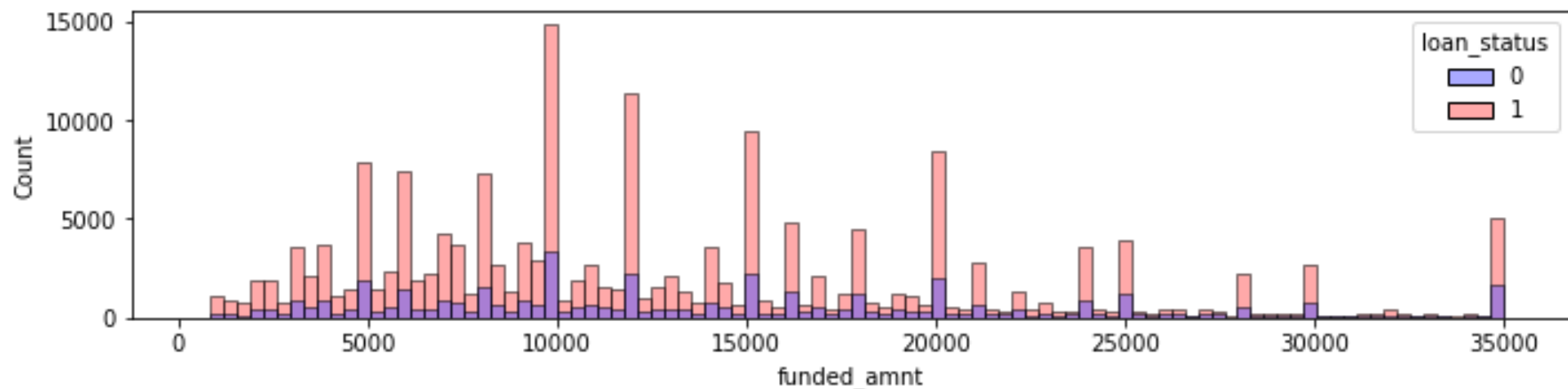
ITEM 3

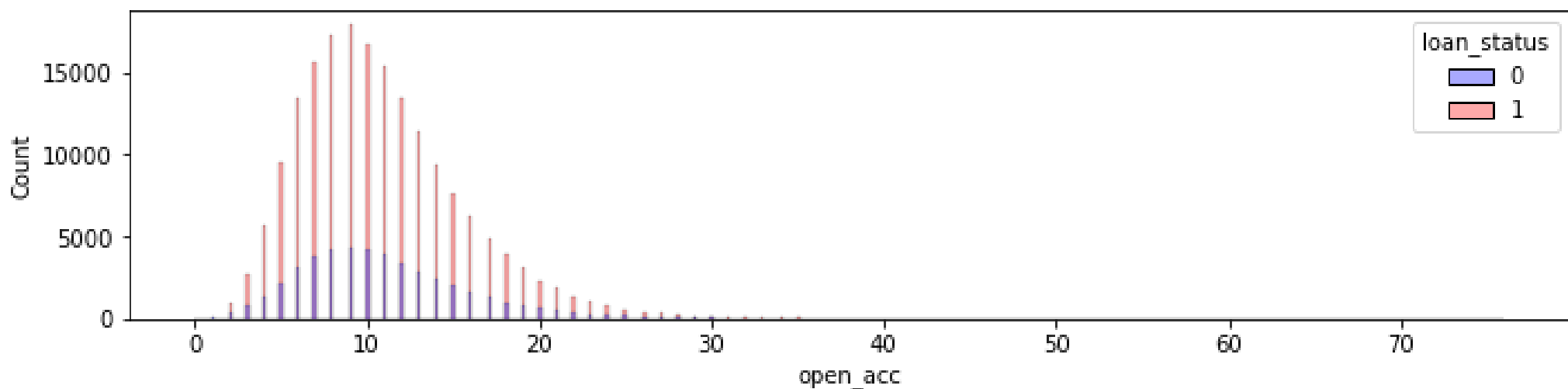
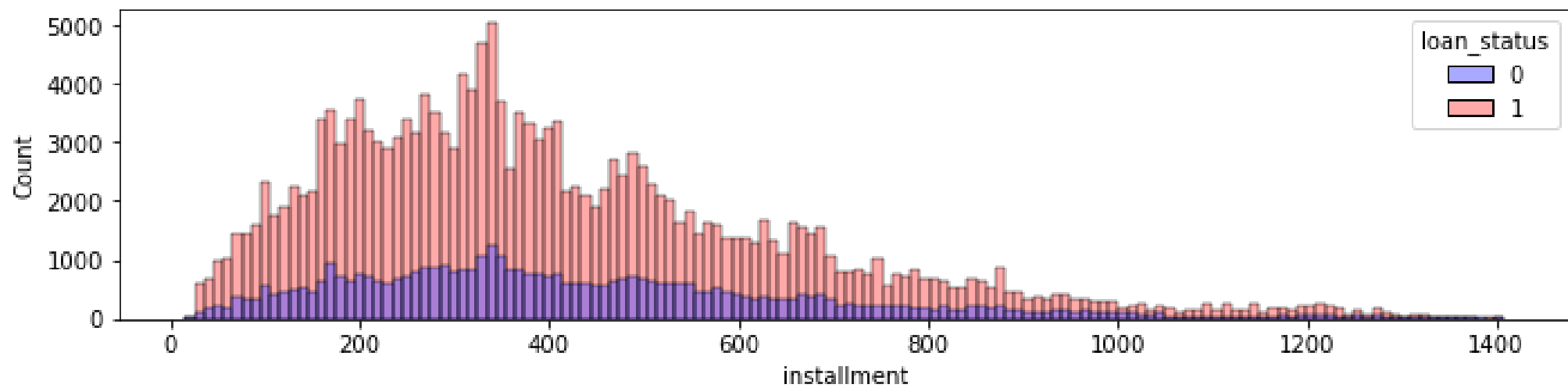
Presentations are communication tools that can be used as demonstrations and more.

Status Pinjaman







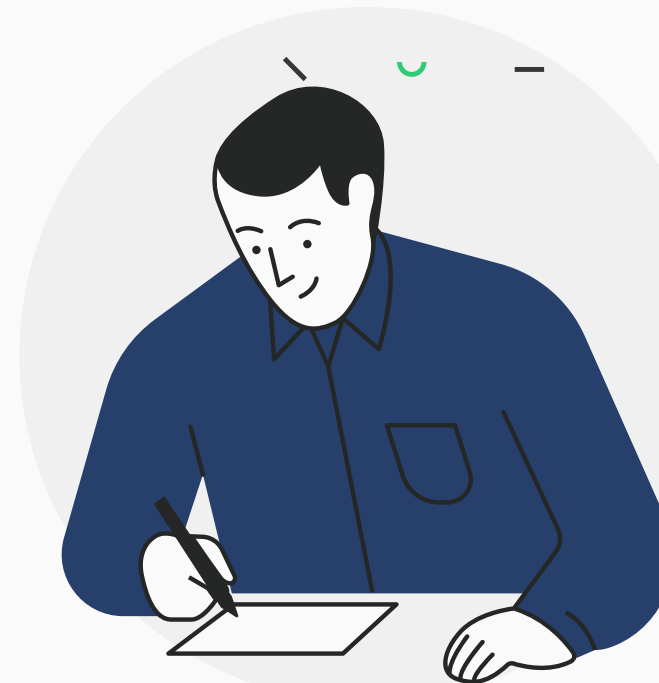


Data Preprocessing

1. Feature modification

melakukan modifikasi pada kategorik agar menjadi numerik : emp_length, term, earliest_cr_line, issue_d, last_payment_d, next_payment_d, last_credit_pull_d.

ex: 36 months menjadi 36



Drop Redundant Feature

Feature-feature yang memiliki korelasi lebih dari 0.7 akan didrop dari data.

✓ | >
✗ | <

'months_since_next_payment_d',
'revol_bal',
'funded_amnt',
'mths_since_last_major_derog',
'funded_amnt_inv',
'recoveries',
'total_rec_late_fee',
'installment',
'total_pymnt_inv',
'total_rec_prncp',
'total_pymnt',
'mths_since_issue_d'



Drop Feature

1. Drop Feature dengan jumlah unique banyak :
'emp_title', 'title', 'application_type'
2. Drop Feature yang hanya memiliki satu unique :
application_type

Handling Missing Values

1. Feature-feature kategorik yang memiliki missing value dilakukan imputasi oleh nilai modus
2. feature-fetaure numerik yang memilki nilai missing values dilakukan imputasi dengan nilai media



Feature Encoding

1. Label Encoding : grade, initial_list_status, pymnt_plan, verification_status.
2. One Hot Encoding home_ownership, verification_status, purpose'

Feature Scaling

Menggunakan minmax scaler



Split Dataset

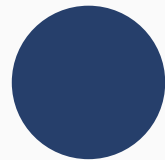
```
x = data.drop('loan_status', axis=1)
```

```
y = data['loan_status']
```

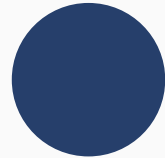
```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,  
random_state=0)
```



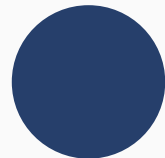
Modeling And Evaluation



LOGISTIC REGRESSION



RANDOM FOREST



XGBOOST CLASSIFIER

TABLE SCORE OF MODELING

Evaluation Score	Logistics Regression	Random Forest	XGBoost
Accuracy	91	98	98
Precisio	97	98	99
Recall	93	99	99

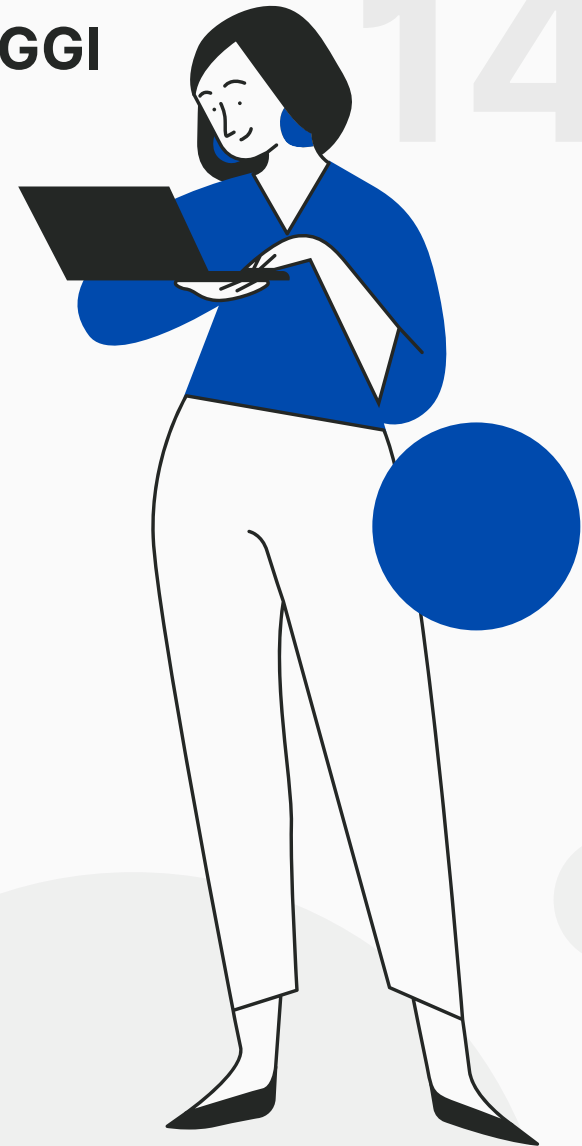


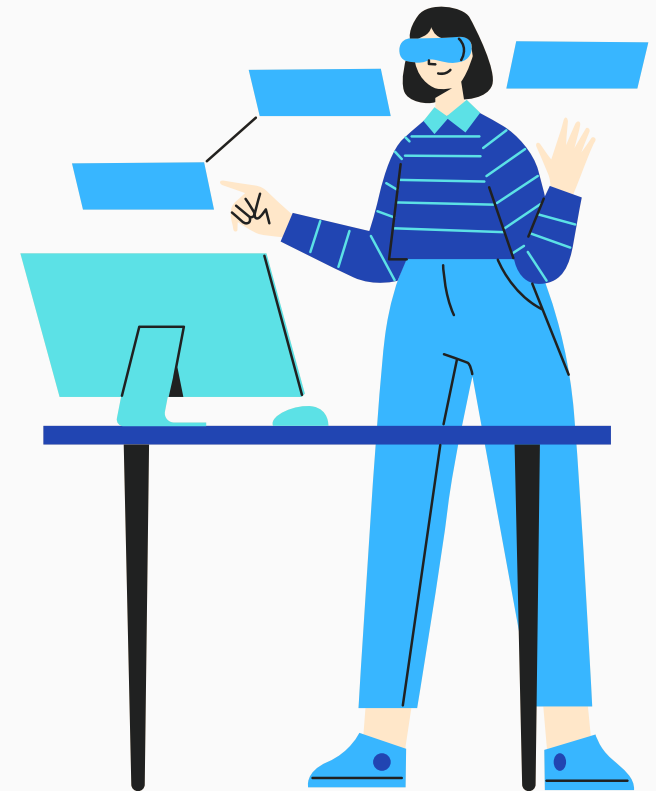
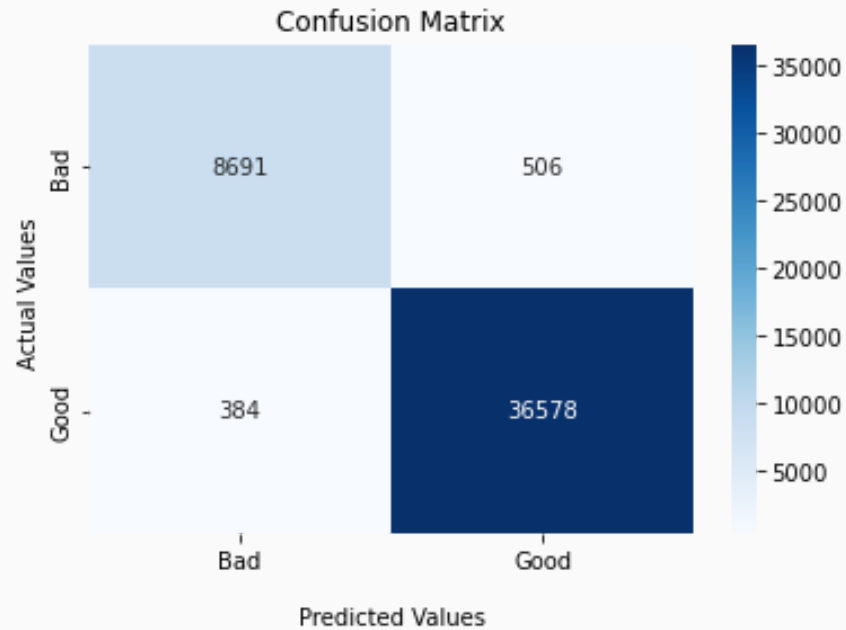
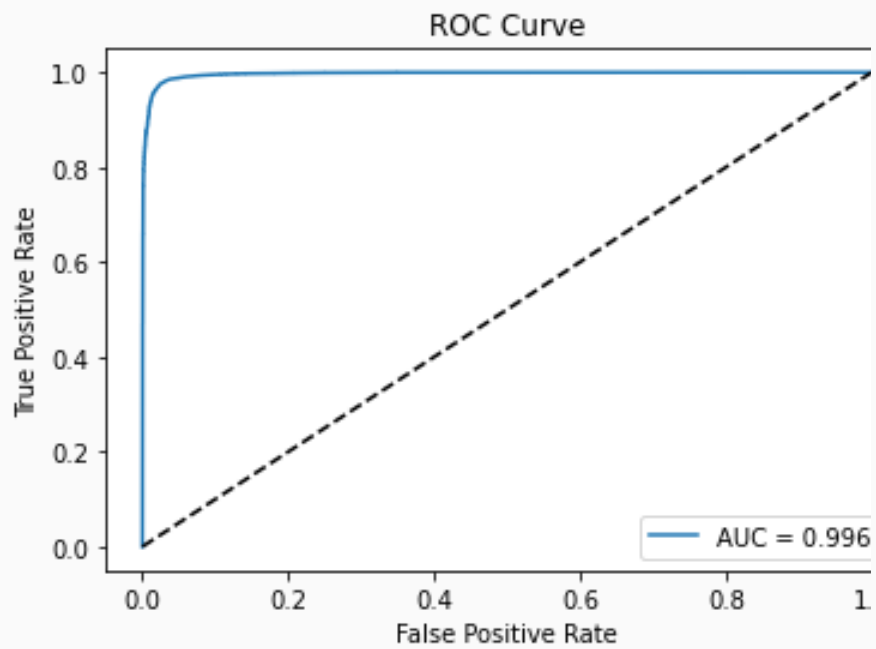
Feature Importance

10 FEATURE DENGAN KORELASI TERTINGGI

1. **collection_recovery_fee**
2. **months_since_last_pull_d_date**
3. **initial_list_status**
4. **last_pymnt_amnt**
5. **term_int**
6. **tot_cur_bal**
7. **grade**
8. **months_since_last_payment_d**
9. **out_prncp**
10. **inq_last_6months**

› | –
✓ | ›
\ ✓ –





Contact me



mayamaryanah4@gmail.com



linkedin.com/in/mayamaryanah



github.com/mayamaryanah