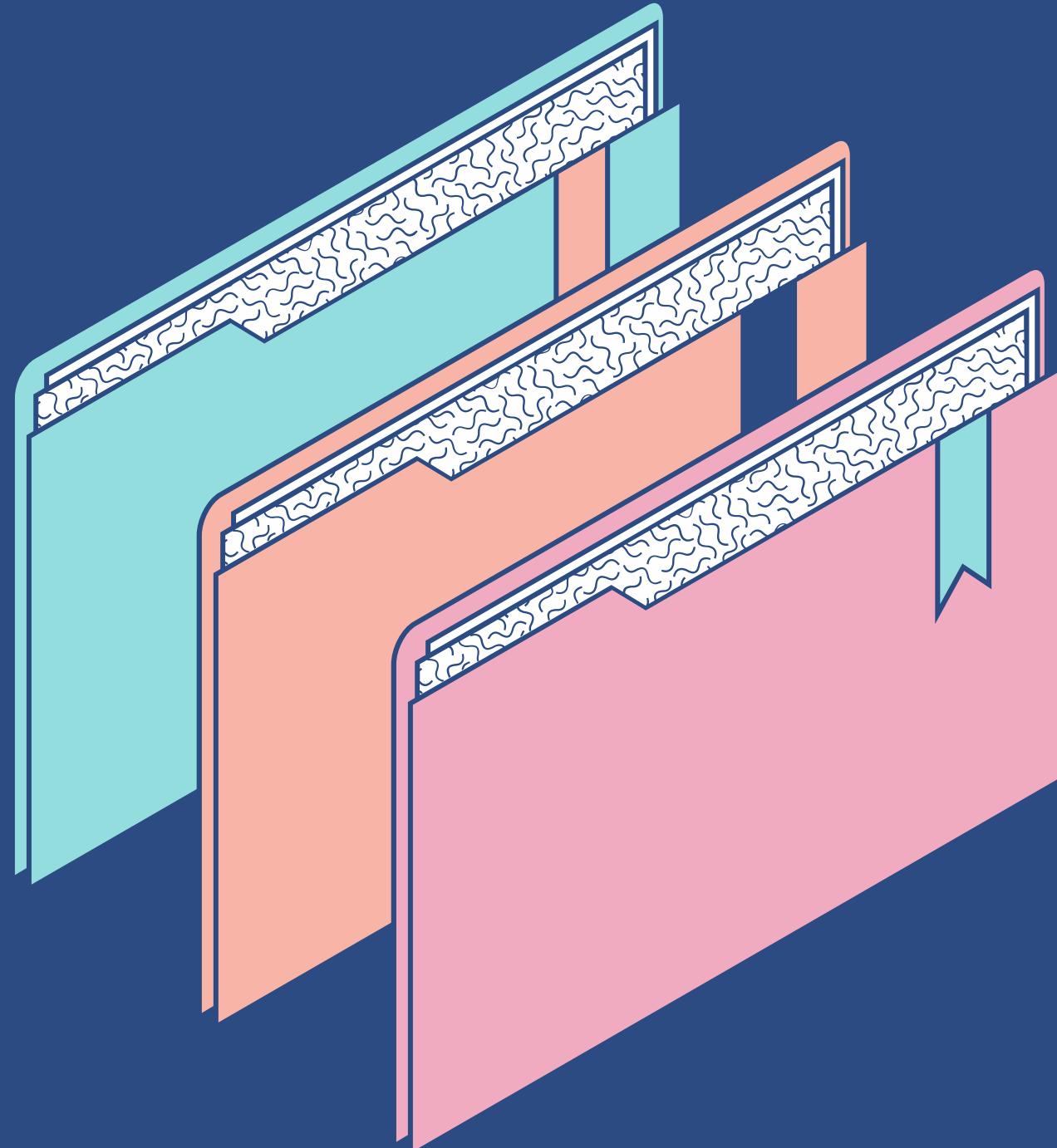




Prediksi Credit Risk menggunakan Data Pinjaman

Oleh Mahmuda

Study Case

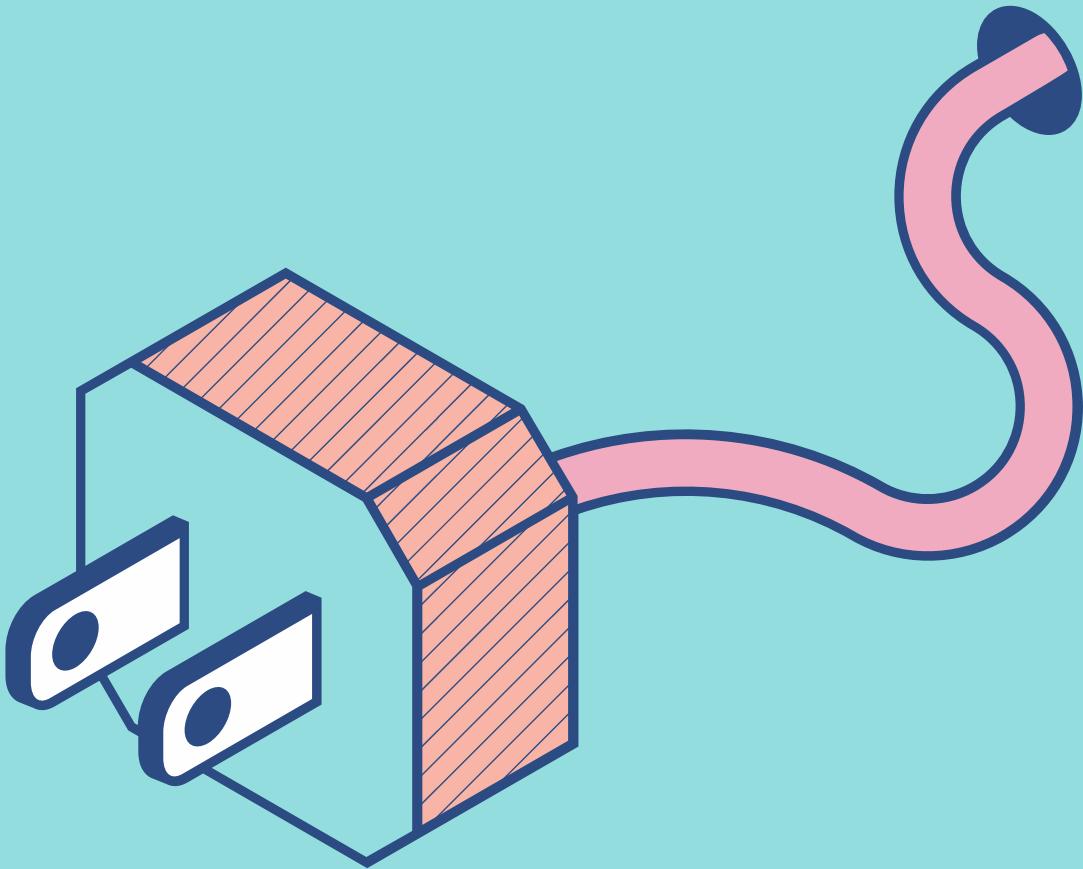


PENDAHULUAN

Sebagai seorang Data Scientist di ID/X Partners, proyek kali ini melibatkan pembuatan model untuk memprediksi risiko kredit pada sebuah lending company. Tujuan utama dari proyek ini adalah memberikan solusi teknologi yang dapat membantu perusahaan dalam mengambil keputusan terkait pemberian pinjaman.

Deskripsi Dataset

Dataset yang diberikan oleh perusahaan terdiri dari data pinjaman yang diterima dan ditolak. Data ini mencakup berbagai variabel yang relevan untuk analisis risiko kredit, seperti jumlah pinjaman, tingkat suku bunga, pendapatan tahunan, jumlah tunggakan, dan lainnya. dalam study case ini menggunakan ‘loan_status’ sebagai variabel target.



Langkah-Langkah Penggerjaan

1 ————— 2 ————— 3 ————— 4 ————— 5

Eksplorasi Data
(Exploratory Data
Analysis - EDA)

Preprocessing
Data

Pembuatan
Model

Visualisasi Hasil

Evaluasi Hasil



Persiapan Library dan Dataset Yang Digunakan

```
# Impor pustaka-pustaka yang diperlukan
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, confusion_matrix, roc_curve, auc
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import cross_val_score, KFold
from sklearn.impute import SimpleImputer
```

```
# Baca dataset
data = pd.read_csv("drive/MyDrive/loan_data_2007_2014.csv")
```

Pemahaman Data

- Dataset ini terdiri dari 74 kolom dengan 466.285 baris, setiap kolom menampilkan berbagai tipe data.

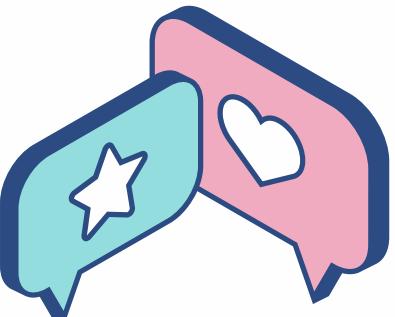
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 466285 entries, 0 to 466284
Data columns (total 75 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Unnamed: 0       466285 non-null   int64  
 1   id               466285 non-null   int64  
 2   member_id        466285 non-null   int64  
 3   loan_amnt        466285 non-null   int64  
 4   funded_amnt      466285 non-null   int64  
 5   funded_amnt_inv  466285 non-null   float64
 6   term              466285 non-null   object 
 7   int_rate          466285 non-null   float64
 8   installment       466285 non-null   float64
 9   grade             466285 non-null   object 
 10  sub_grade         466285 non-null   object 
 11  emp_title         438697 non-null   object 
 12  emp_length        445277 non-null   object 
 13  home_ownership    466285 non-null   object 
 14  annual_inc        466281 non-null   float64
 15  verification_status 466285 non-null   object 
 16  issue_d           466285 non-null   object 
 17  loan_status        466285 non-null   object 
 18  pymnt_plan        466285 non-null   object 

 19  url               466285 non-null   object 
 20  desc              125983 non-null   object 
 21  purpose            466285 non-null   object 
 22  title              466265 non-null   object 
 23  zip_code           466285 non-null   object 
 24  addr_state         466285 non-null   object 
 25  dti                466285 non-null   float64
 26  delinq_2yrs        466256 non-null   float64
 27  earliest_cr_line   466256 non-null   object 
 28  inq_last_6mths     466256 non-null   float64
 29  mths_since_last_delinq 215934 non-null   float64
 30  mths_since_last_record 62638 non-null   float64
 31  open_acc           466256 non-null   float64
 32  pub_rec            466256 non-null   float64
 33  revol_bal          466285 non-null   int64  
 34  revol_util         465945 non-null   float64
 35  total_acc          466256 non-null   float64
 36  initial_list_status 466285 non-null   object 
 37  out_prncp          466285 non-null   float64
 38  out_prncp_inv      466285 non-null   float64
 39  total_pymnt        466285 non-null   float64
 40  total_pymnt_inv    466285 non-null   float64
 41  total_rec_prncp    466285 non-null   float64
 42  total_rec_int       466285 non-null   float64

 43  total_rec_late_fee  466285 non-null   float64
 44  recoveries          466285 non-null   float64
 45  collection_recovery_fee 466285 non-null   float64
 46  last_pymnt_d        465909 non-null   object 
 47  last_pymnt_amnt     466285 non-null   float64
 48  next_pymnt_d        239071 non-null   object 
 49  last_credit_pull_d   466243 non-null   object 
 50  collections_12_mths_ex_med 466140 non-null   float64
 51  mths_since_last_major_derog 98974 non-null   float64
 52  policy_code          466285 non-null   int64  
 53  application_type     466285 non-null   object 
 54  annual_inc_joint     0 non-null    float64
 55  dti_joint            0 non-null    float64
 56  verification_status_joint 0 non-null   float64
 57  acc_now_delinq       466256 non-null   float64
 58  tot_coll_amt         396009 non-null   float64
 59  tot_cur_bal          396009 non-null   float64
 60  open_acc_6m          0 non-null    float64
 61  open_il_6m           0 non-null    float64
 62  open_il_12m          0 non-null    float64
 63  open_il_24m          0 non-null    float64
 64  mths_since_rcnt_il   0 non-null    float64
 65  total_bal_il         0 non-null    float64
 66  il_util              0 non-null    float64

 67  open_rv_12m          0 non-null    float64
 68  open_rv_24m          0 non-null    float64
 69  max_bal_bc           0 non-null    float64
 70  all_util             0 non-null    float64
 71  total_rev_hi_lim     396009 non-null   float64
 72  inq_fi               0 non-null    float64
 73  total_cu_tl           0 non-null    float64
 74  inq_last_12m          0 non-null    float64

dtypes: float64(46), int64(7), object(22)
memory usage: 266.8+ MB
None
```



Statistik Deskriptif

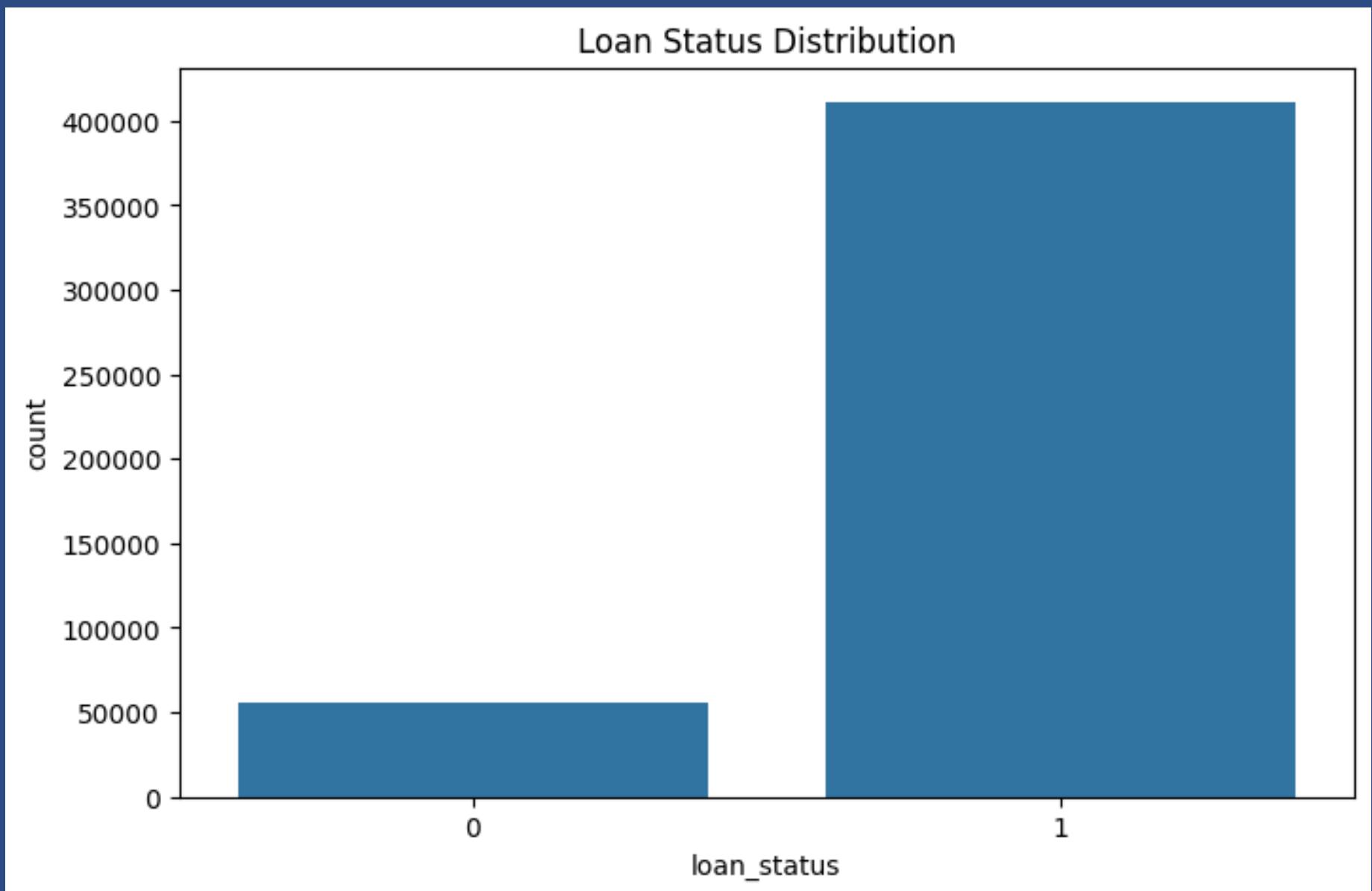
	Unnamed: 0	id	member_id	loan_amnt	annual_inc	dti	...	total_bal_il	il_util	open_rv_12m
count	466285.000000	4.662850e+05	4.662850e+05	466285.000000	4.662810e+05	466285.000000	...	0.0	0.0	0.0
mean	233142.000000	1.307973e+07	1.459766e+07	14317.277577	7.327738e+04	17.218758	...	NaN	NaN	NaN
std	134605.029472	1.089371e+07	1.168237e+07	8286.509164	5.496357e+04	7.851121	...	NaN	NaN	NaN
min	0.000000	5.473400e+04	7.047300e+04	500.000000	1.896000e+03	0.000000	...	NaN	NaN	NaN
25%	116571.000000	3.639987e+06	4.379705e+06	8000.000000	4.500000e+04	11.360000	...	NaN	NaN	NaN
50%	233142.000000	1.010790e+07	1.194108e+07	12000.000000	6.300000e+04	16.870000	...	NaN	NaN	NaN
75%	349713.000000	2.073121e+07	2.300154e+07	20000.000000	8.896000e+04	22.780000	...	NaN	NaN	NaN
max	466284.000000	3.809811e+07	4.086083e+07	35000.000000	7.500000e+06	39.990000	...	NaN	NaN	NaN

	funded_amnt	funded_amnt_inv	int_rate	installment	open_rv_24m	max_bal_bc	all_util	total_rev_hi_lim	inq_fi
count	466285.000000	466285.000000	466285.000000	466285.000000	0.0	0.0	0.0	3.960090e+05	0.0
mean	14291.801044	14222.329888	13.829236	432.061201	NaN	NaN	NaN	3.037909e+04	NaN
std	8274.371300	8297.637788	4.357587	243.485550	NaN	NaN	NaN	3.724713e+04	NaN
min	500.000000	0.000000	5.420000	15.670000	NaN	NaN	NaN	0.000000e+00	NaN
25%	8000.000000	8000.000000	10.990000	256.690000	NaN	NaN	NaN	1.350000e+04	NaN
50%	12000.000000	12000.000000	13.660000	379.890000	NaN	NaN	NaN	2.280000e+04	NaN
75%	20000.000000	19950.000000	16.490000	566.580000	NaN	NaN	NaN	3.790000e+04	NaN
max	35000.000000	35000.000000	26.060000	1409.990000	NaN	NaN	NaN	9.999999e+06	NaN

	total_cu_tl	inq_last_12m
count	0.0	0.0
mean	NaN	NaN
std	NaN	NaN
min	NaN	NaN
25%	NaN	NaN
50%	NaN	NaN
75%	NaN	NaN
max	NaN	NaN

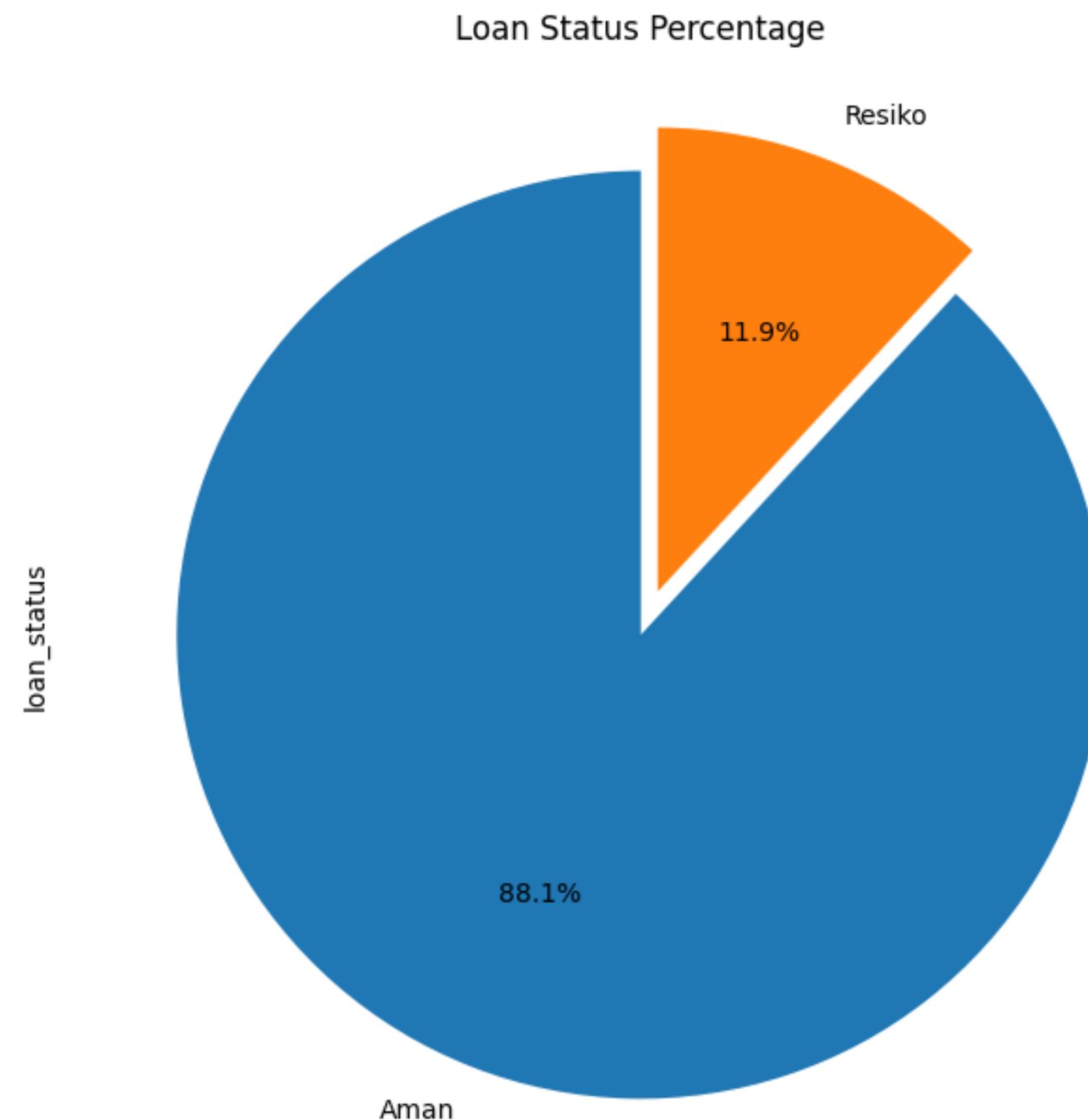
Eksplorasi Data (Exploratory Data Analysis - EDA)

Gambar tersebut Menunjukkan jumlah peminjam berdasarkan status pinjaman. Status pinjaman yang ditunjukkan adalah resiko (0), dan aman (1).



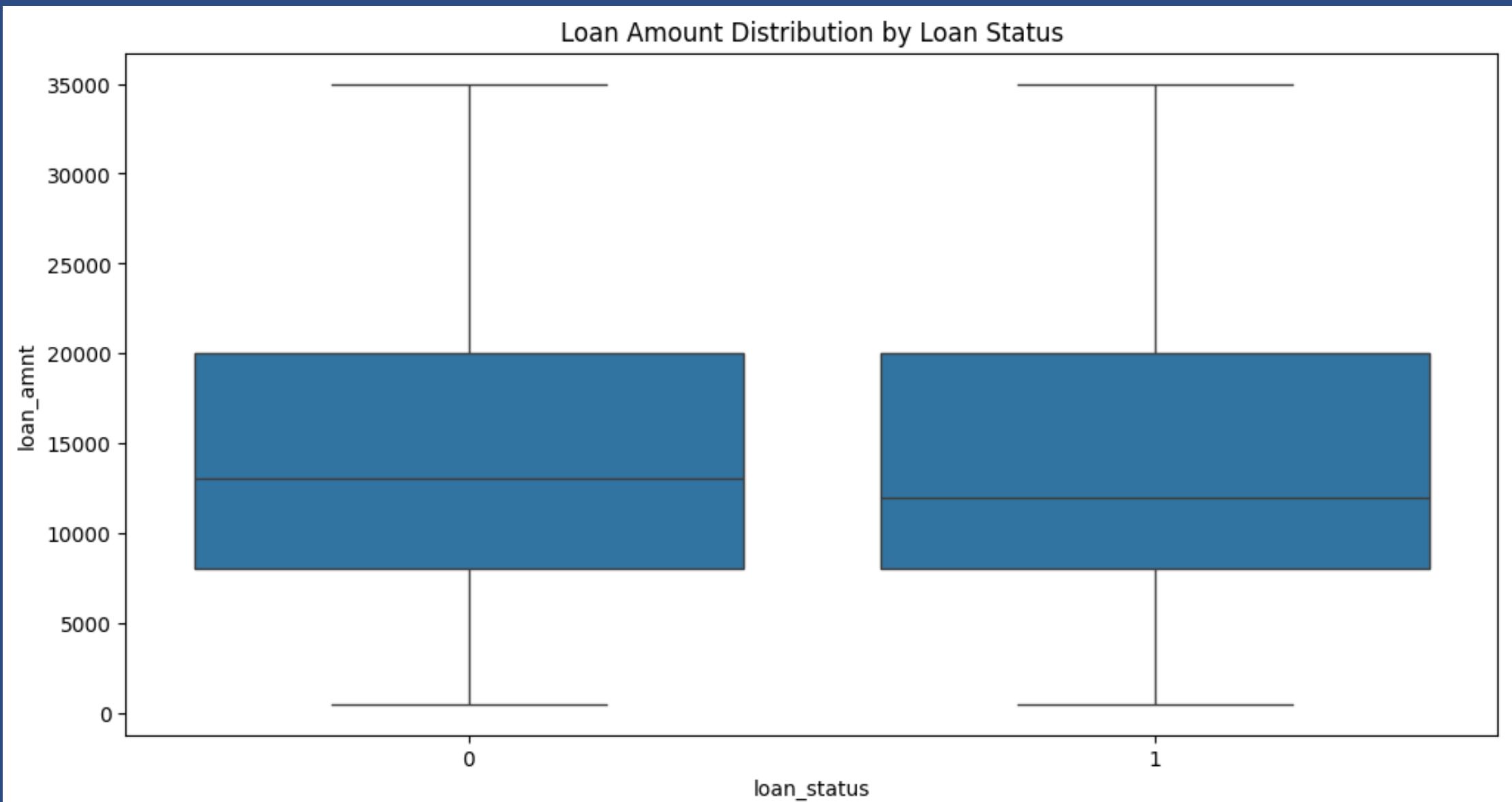
Eksplorasi Data (Exploratory Data Analysis - EDA)

Gambar tersebut menunjukkan persentase peminjam berdasarkan status pinjaman. Status pinjaman yang ditunjukkan adalah Aman (88.1%) dan Resiko (11.9%).



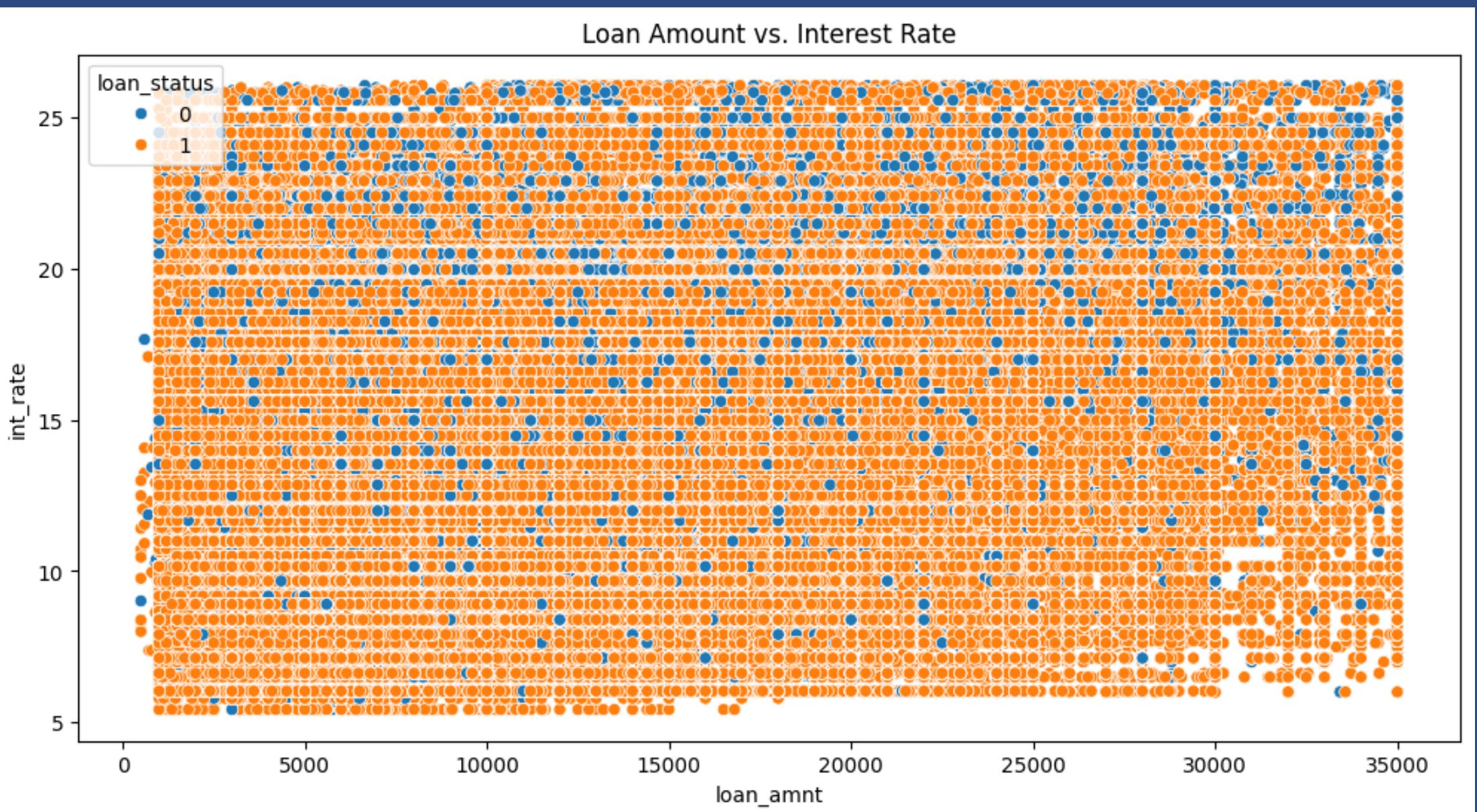
Eksplorasi Data (Exploratory Data Analysis - EDA)

Gambar tersebut menunjukkan jumlah peminjam berdasarkan status pinjaman dan jumlah pinjaman. Jumlah pinjaman yang ditunjukkan adalah dari 0 hingga 35.000 dengan interval 5.000.



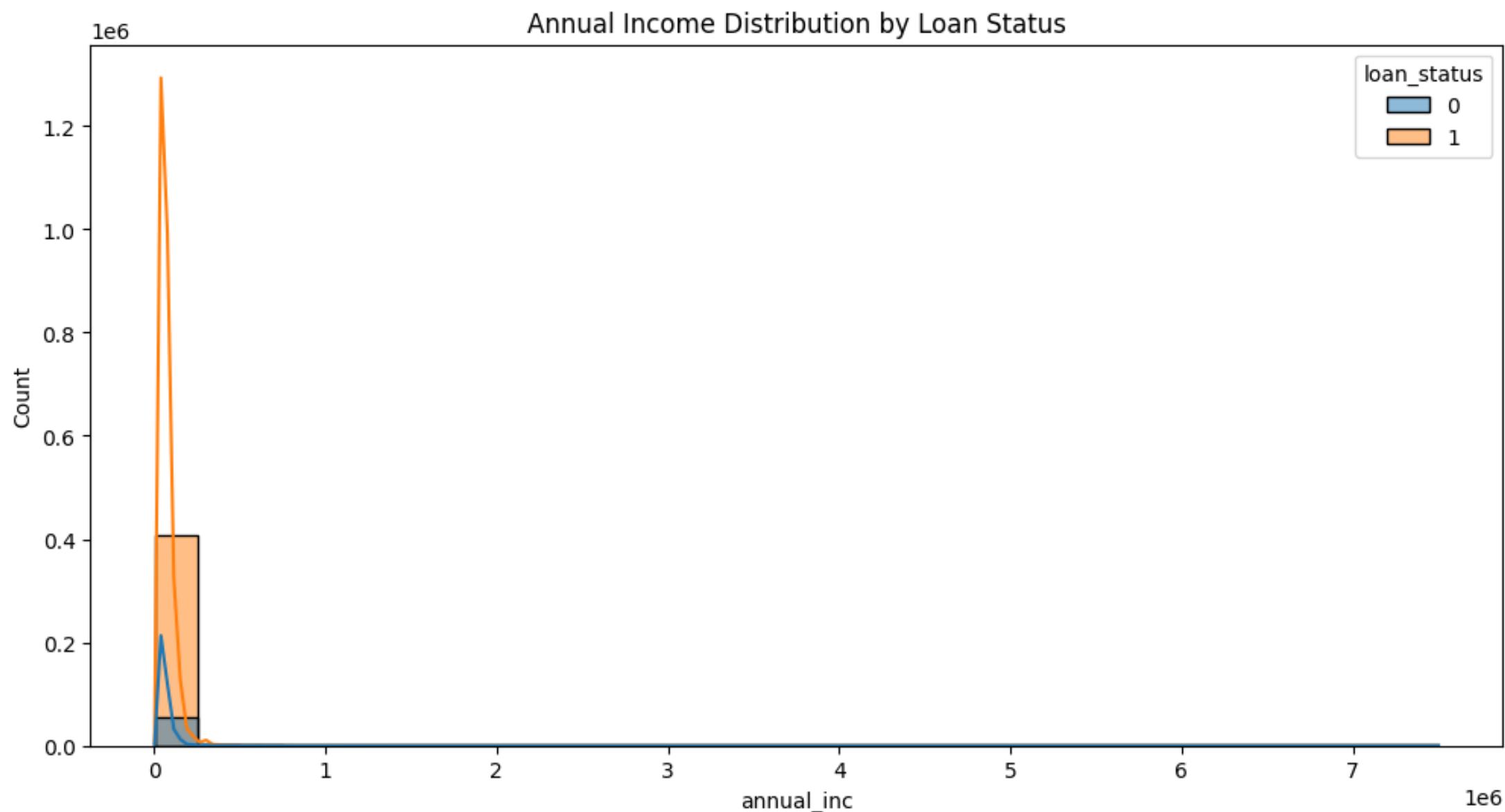
Eksplorasi Data (Exploratory Data Analysis - EDA)

Gambar tersebut menampilkan perbandingan antara jumlah pinjaman ('loan_amnt') dan besar suku bunga (int_rate)



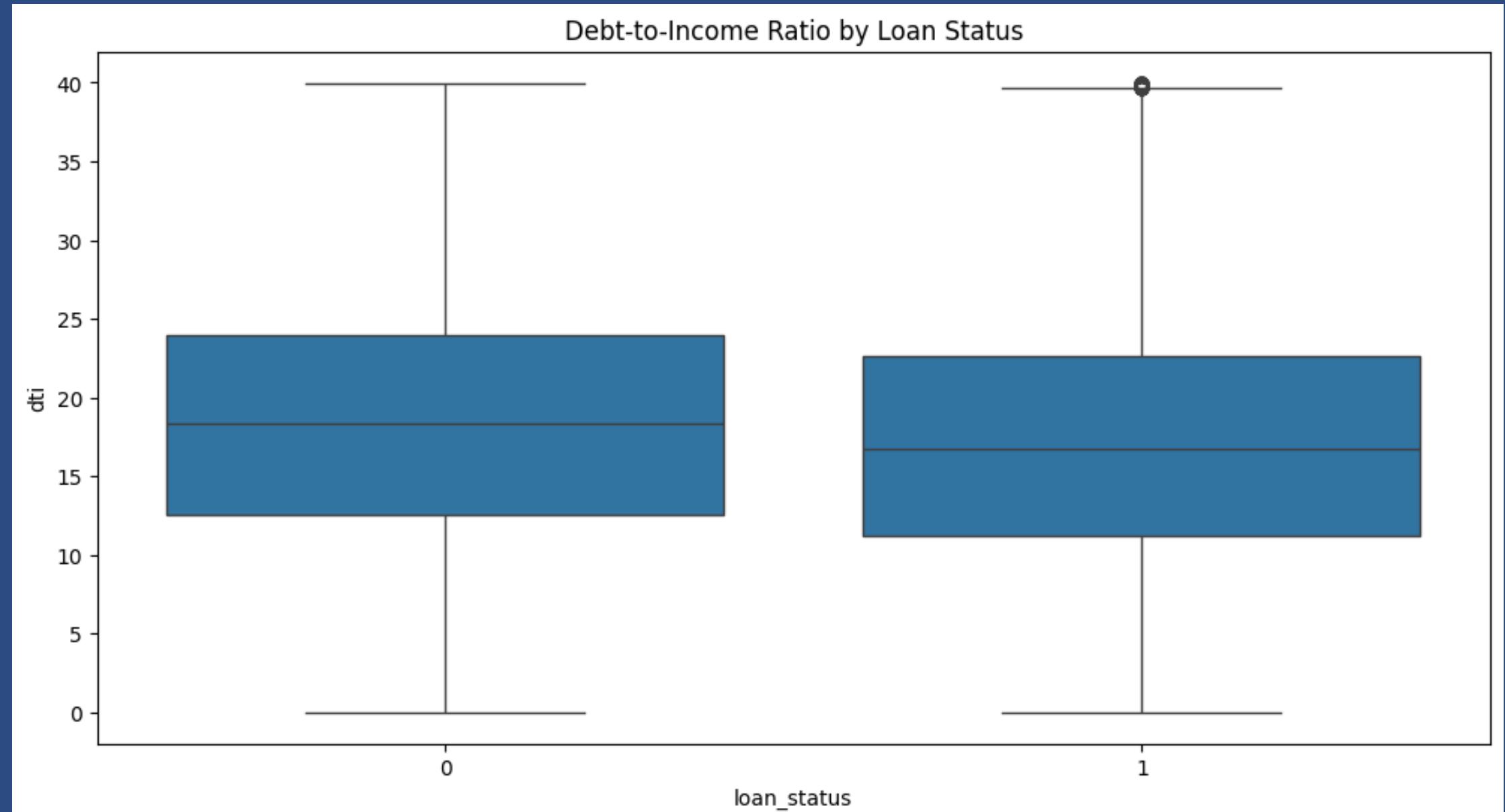
Eksplorasi Data (Exploratory Data Analysis - EDA)

Gambar tersebut menunjukkan jumlah peminjam berdasarkan status pinjaman dan rentang pendapatan tahunan. Rentang pendapatan tahunan yang ditunjukkan adalah dari 0 hingga 1 juta dengan interval 0.2 juta.



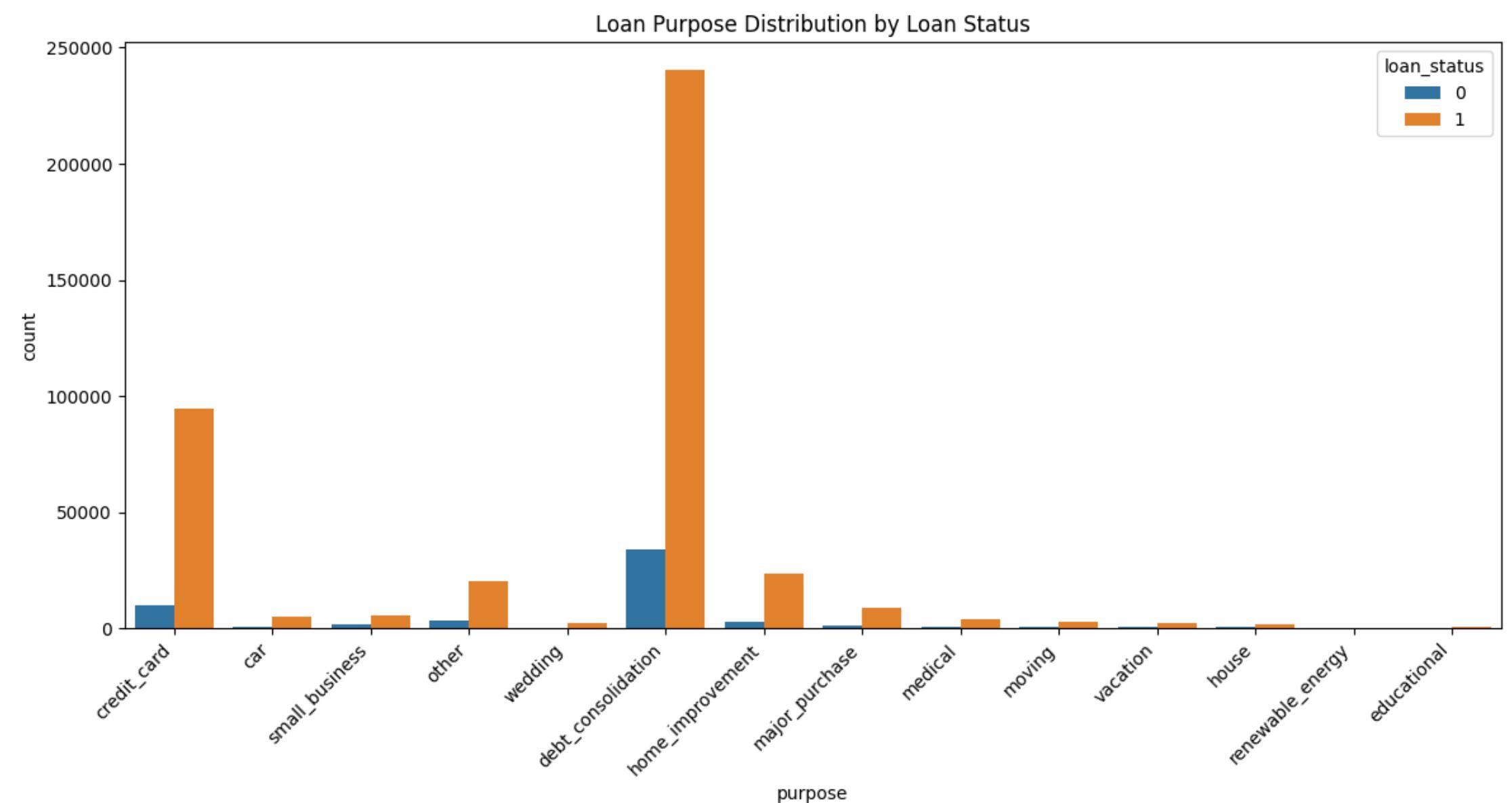
Eksplorasi Data (Exploratory Data Analysis - EDA)

Gambar tersebut menunjukkan jumlah peminjam berdasarkan status pinjaman dan perbandingan hutang-penghasilan. Perbandingan hutang-penghasilan yang ditunjukkan adalah 0 hingga 40 dengan interval 5.



Eksplorasi Data (Exploratory Data Analysis - EDA)

Gambar tersebut menunjukkan jumlah peminjam berdasarkan status pinjaman dan tujuan peminjaman. Tujuan peminjaman yang ditunjukkan adalah kartu kredit (credit card), mobil (car), bisnis (business), lain-lain (other), pernikahan (wedding), dan beberapa tujuan lainnya.



Preprocessing data

- Membersihkan data: menghapus kolom yang tidak relevan, menghapus data duplikat, dan mengidentifikasi outlier

```
print("shape sebelum data dibersihkan = ", data.shape)
```

```
shape sebelum data dibersihkan = (466285, 75)
```

```
print("shape sesudah data dibersihkan", data.shape)
```

```
shape sesudah data dibersihkan (466285, 56)
```

Pembuatan Model

Dalam study case ini membangun model dengan klasifikasi random forest

```
# Pilih model pembelajaran mesin yang sesuai  
model = RandomForestClassifier(random_state=42)
```

```
# Latih model dengan data pelatihan  
model.fit(X_train_scaled, y_train)
```

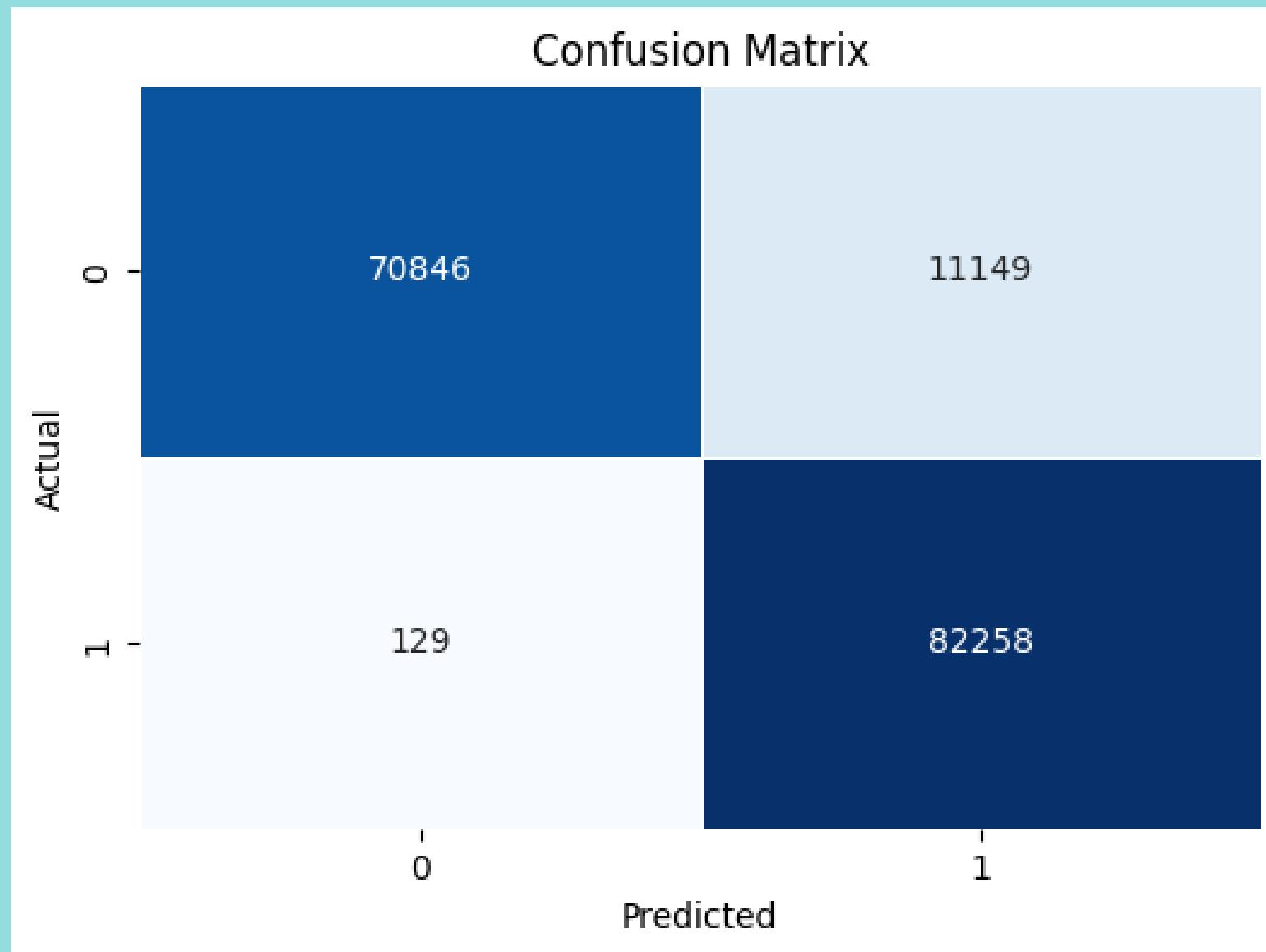


RandomForestClassifier

RandomForestClassifier(random_state=42)



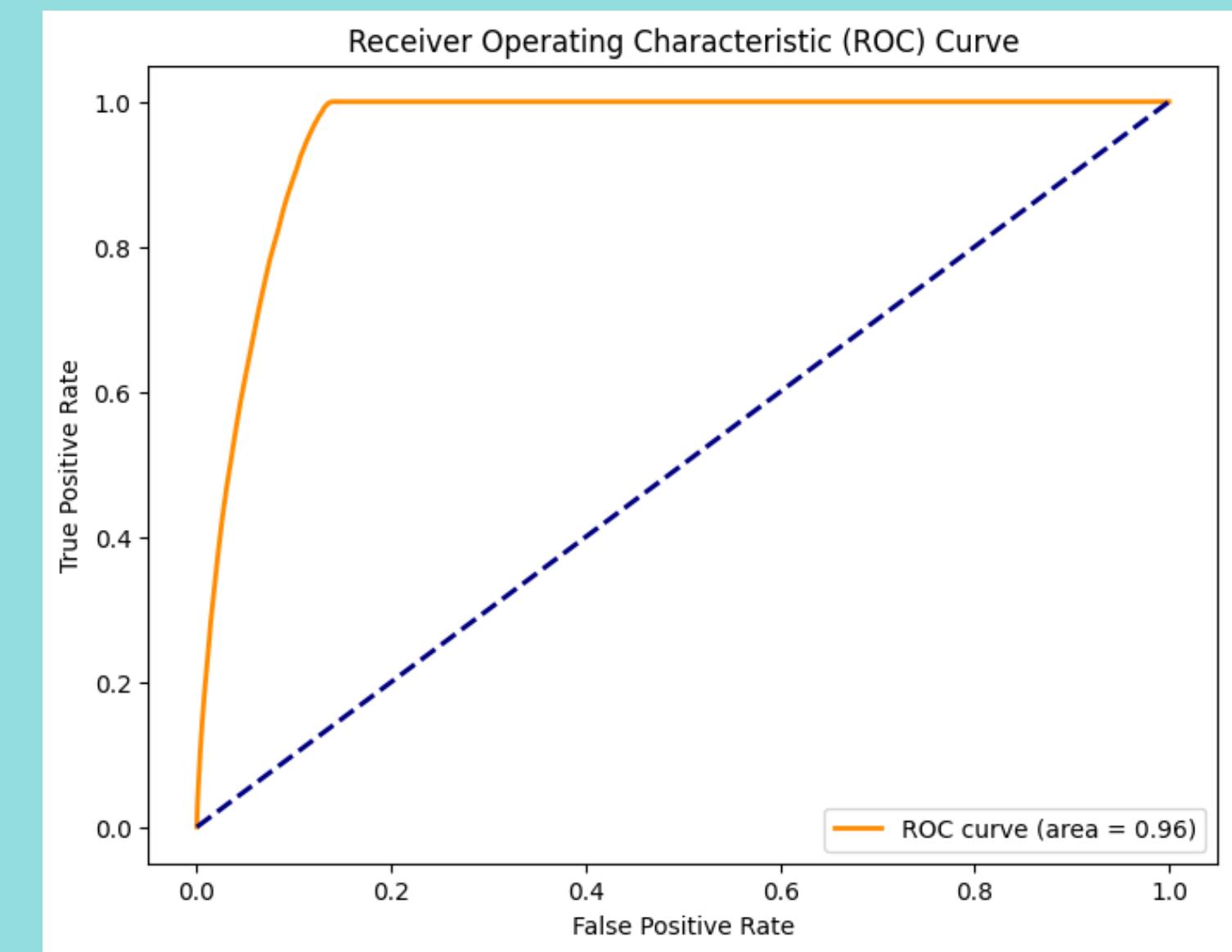
Visualisasi Hasil



Model ini memberikan prediksi dengan
82258 True Positif dan 70846 True Negatif.

Visualisasi Hasil

kurva ROC memiliki area 0,96, yang menunjukkan bahwa model klasifikasi ini memiliki kinerja yang baik.



Evaluasi Hasil

Hasil evaluasi model klasifikasi dengan random forest adalah sebagai berikut:

1. Accuracy (Akurasi): Akurasi mengukur sejauh mana model dapat memprediksi dengan benar. Nilai akurasi sebesar 0.9314 atau 93.14% menunjukkan bahwa model dapat memprediksi dengan tingkat keseluruhan yang tinggi.
2. Precision (Presisi): Presisi mengukur seberapa banyak dari yang diprediksi positif oleh model yang benar-positif. Dalam konteks ini, nilai presisi sebesar 0.8806 atau 88.06% menunjukkan bahwa ketika model memprediksi risiko kredit (positif), sekitar 88.06% prediksinya benar.
3. Recall (Recall atau Sensitivitas): Recall mengukur seberapa banyak dari keseluruhan kelas yang positif dapat diidentifikasi oleh model. Nilai recall sebesar 0.9984 atau 99.84% menunjukkan bahwa model dapat mengidentifikasi sebagian besar risiko kredit yang sebenarnya.
4. F1 Score: F1 Score adalah rata-rata harmonik antara presisi dan recall. Nilai F1 Score sebesar 0.9358 atau 93.58% mencerminkan keseimbangan yang baik antara presisi dan recall.
5. ROC AUC (Area di Bawah Kurva ROC): ROC AUC mengukur kinerja model dalam memisahkan kelas positif dan negatif. Nilai ROC AUC sebesar 0.9312 atau 93.12% menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan antara kredit yang aman dan kredit yang berisiko.

Secara keseluruhan, hasil evaluasi model menunjukkan kinerja yang tinggi, dan model ini dapat dianggap efektif dalam memprediksi risiko kredit dengan akurat.

Terima Kasih...

