Click here for more details.

CREDIT RISK PREDICTION

BY: DEA DAHLILA

About Me

“I'm a Fresh Graduate of Geophysics Engineering with a keen interest in data science and analysis. I'm passionate about extracting meaningful insights from data and using them to drive informed decision-making.”

Let's Connect

ddahlila6@gmail.com

linkedin.com/in/dea-dahlila/

github.com/Dea-dahlila

WHAT IS THE PROBLEM?

BACKGROUND	As an Intern Data Scientist at ID/X Partners, I am involved in a collaborative project for a lending company. The focus of the project is to develop technology solutions to manage credit risk.
GOALS	Build an accurate credit risk prediction model to increase the efficiency of the loan approval process and reduce the risk of default.
OBJECTIVES	Create machine learning model credit score prediction
BUSINESS METRICS	• Accuracy prediction • Credit score prediction model

DATASET

```
RangeIndex: 8580 entries, 0 to 8579
Data columns (total 75 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Unnamed: 0                  8580 non-null   int64
 1   id                          8580 non-null   int64
 2   member_id                   8580 non-null   int64
 3   loan_amnt                   8580 non-null   int64
 4   funded_amnt                 8580 non-null   int64
 5   funded_amnt_inv             8580 non-null   float64
 6   term                        8580 non-null   object
 7   int_rate                    8580 non-null   float64
 8   installment                 8580 non-null   float64
 9   grade                       8580 non-null   object
 10  sub_grade                   8580 non-null   object
 11  emp_title                   8009 non-null   object
 12  emp_length                  8268 non-null   object
 13  home_ownership              8580 non-null   object
 14  annual_inc                  8580 non-null   float64
 15  verification_status         8580 non-null   object
 16  issue_d                     8580 non-null   object
 17  loan_status                 8580 non-null   object
 18  pymnt_plan                  8580 non-null   object
 19  url                         8580 non-null   object
 20  desc                        4881 non-null   object
 21  purpose                     8580 non-null   object
 22  title                       8580 non-null   object
 23  zip_code                    8580 non-null   object
 24  addr_state                  8580 non-null   object
 25  dti                         8580 non-null   float64
 26  delinq_2yrs                 8580 non-null   float64
 27  earliest_cr_line            8580 non-null   object
 28  inq_last_6mths              8580 non-null   float64
 29  mths_since_last_delinq      2716 non-null   float64
 30  mths_since_last_record      371 non-null    float64
 31  open_acc                    8580 non-null   float64
 32  pub_rec                     8580 non-null   float64
 33  revol_bal                   8580 non-null   int64
 34  revol_util                  8577 non-null   float64
 35  total_acc                   8580 non-null   float64
 36  initial_list_status         8579 non-null   object
 37  out_prncp                   8579 non-null   float64
 38  out_prncp_inv               8579 non-null   float64
 39  total_pymnt                 8579 non-null   float64
 40  total_pymnt_inv             8579 non-null   float64
 41  total_rec_prncp             8579 non-null   float64
 42  total_rec_int               8579 non-null   float64
 43  total_rec_late_fee          8579 non-null   float64
 44  recoveries                  8579 non-null   float64
 45  collection_recovery_fee     8579 non-null   float64
 46  last_pymnt_d                8565 non-null   object
 47  last_pymnt_amnt             8579 non-null   float64
 48  next_pymnt_d                861 non-null    object
 49  last_credit_pull_d          8579 non-null   object
 50  collections_12_mths_ex_med  8579 non-null   float64
 51  mths_since_last_major_derog 0 non-null      float64
 52  policy_code                 8579 non-null   float64
 53  application_type            8579 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              8579 non-null   float64
 58  tot_coll_amt                0 non-null      float64
 59  tot_cur_bal                 0 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            0 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
```

Shape

8580 data rows, 75 features.

Dtype

float64 (47 features), int64 (6 features), object (22 features).

DATA CLEANSING & PREPROCESING

1 Missing Value

Drop features that have a missing value above 15% of data.

2 Unique value

Drop features that have a very high unique value (high cardinality) and features that have only one unique value

3 Corellation Check

Drop features which has a high correlation (0.7)

4 Encoding

Handled with one-hot encoding

5 Split Data

Data Train 70%
Data Test 30%

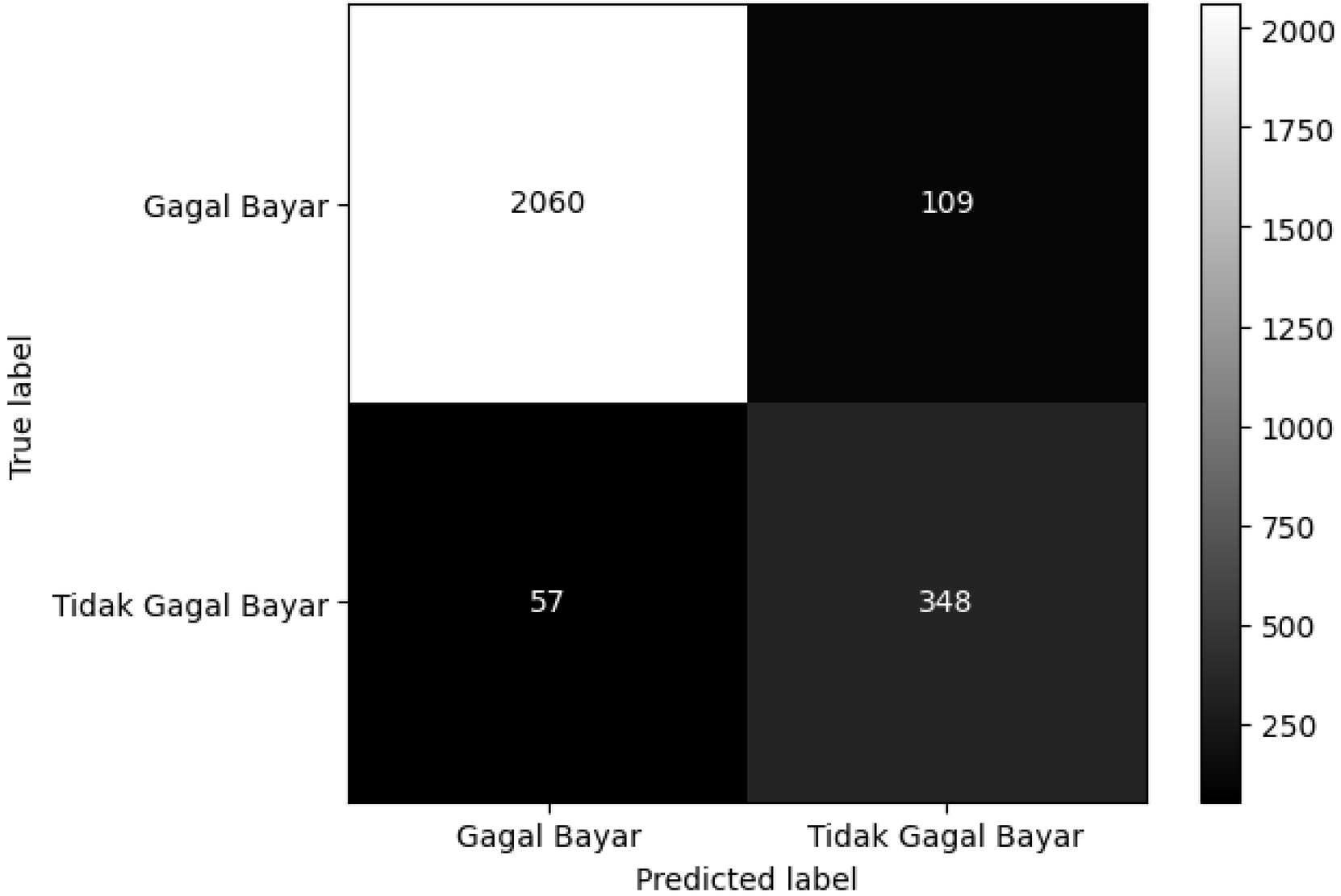
6 Scaling

Use StandarScaler

MODELING

Metrics \ Model	Logistic Regression	XGBoost	Desicion Tree
Accuracy	0.94	0.97	0.96
Precision	0.93	0.98	0.88
Recall	0.86	0.85	0.85
ROC-AUC	0.97	0.98	0.92
F1	0.81	0.91	0.87

CONFUSION MATRIX



Without Machine Learning

18.69%

DEFAULT

With ML

5.02%

DEFAULT

THANK YOU