FRAUD DETECTION PROJECT

- LOGISTIC REGRESSION
- RANDOM FOREST
- XGBOOST
- ANN
- MODEL DEPLOYMENT

G2 – Capstone Project3

CONTENT

- Group Members
- Data Analysis
- EDA
- Logistic Regression without SMOTE
- Logistic Regression with SMOTE
- Random Forest Classifier without SMOTE
- Random Forest Classifier with SMOTE
- XGB Classifier without SMOTE
- XGB Classifier with SMOTE
- ANN
- Model Deployment

Group Member

- F3662-Muhsin
- F3634-İbrahim

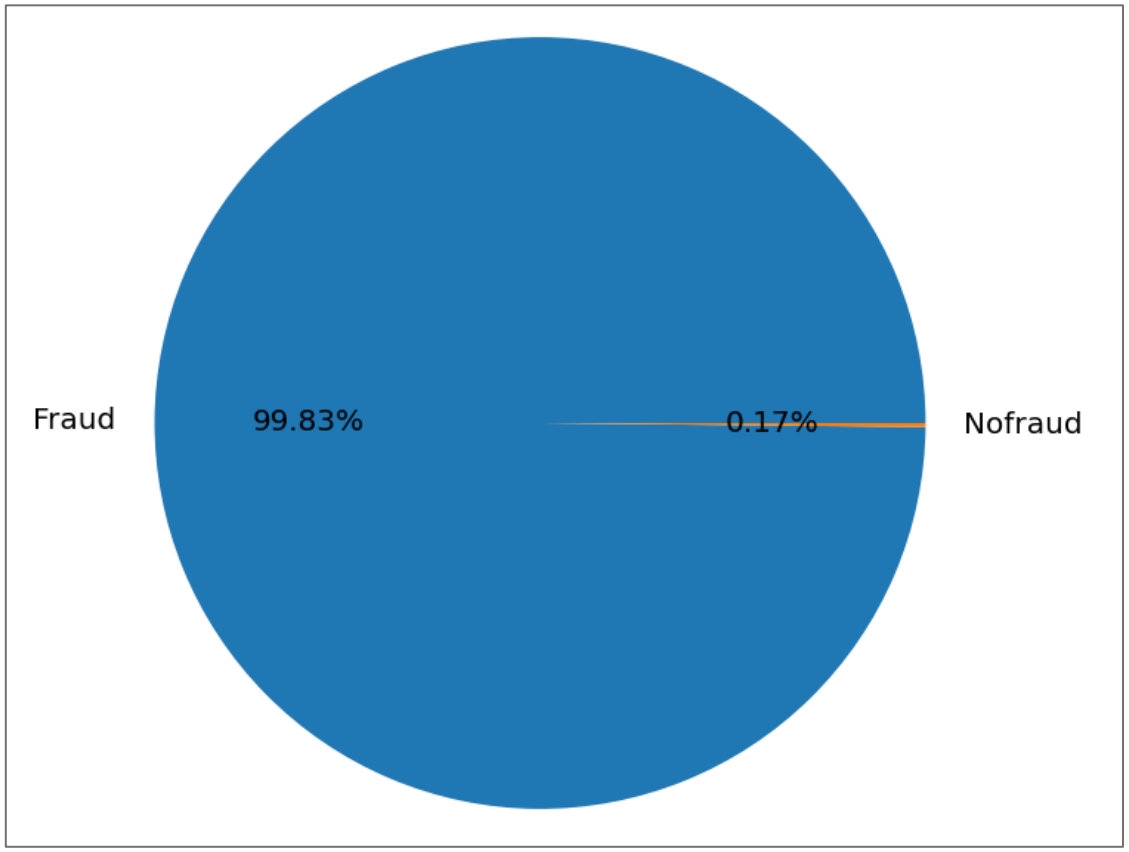
```
df.info()
```

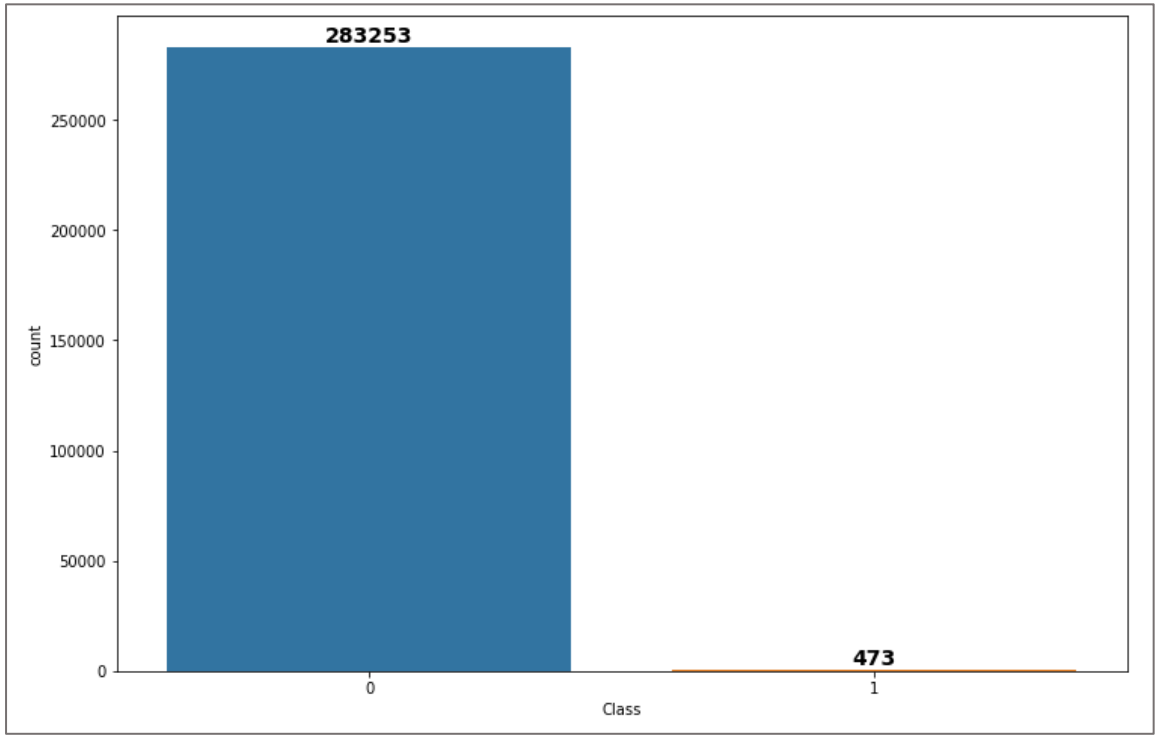
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
 #   Column  Non-Null Count   Dtype  
---  ------  --------------   -----  
 0   Time    284807 non-null  float64
 1   V1      284807 non-null  float64
 2   V2      284807 non-null  float64
 3   V3      284807 non-null  float64
 4   V4      284807 non-null  float64
 5   V5      284807 non-null  float64
 6   V6      284807 non-null  float64
 7   V7      284807 non-null  float64
 8   V8      284807 non-null  float64
 9   V9      284807 non-null  float64
 10  V10     284807 non-null  float64
 11  V11     284807 non-null  float64
 12  V12     284807 non-null  float64
 13  V13     284807 non-null  float64
 14  V14     284807 non-null  float64
 15  V15     284807 non-null  float64
 16  V16     284807 non-null  float64
 17  V17     284807 non-null  float64
 18  V18     284807 non-null  float64
 19  V19     284807 non-null  float64
 20  V20     284807 non-null  float64
 21  V21     284807 non-null  float64
 22  V22     284807 non-null  float64
 23  V23     284807 non-null  float64
 24  V24     284807 non-null  float64
 25  V25     284807 non-null  float64
 26  V26     284807 non-null  float64
 27  V27     284807 non-null  float64
 28  V28     284807 non-null  float64
 29  Amount  284807 non-null  float64
 30  Class   284807 non-null  int64  
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

DATA ANALYSIS

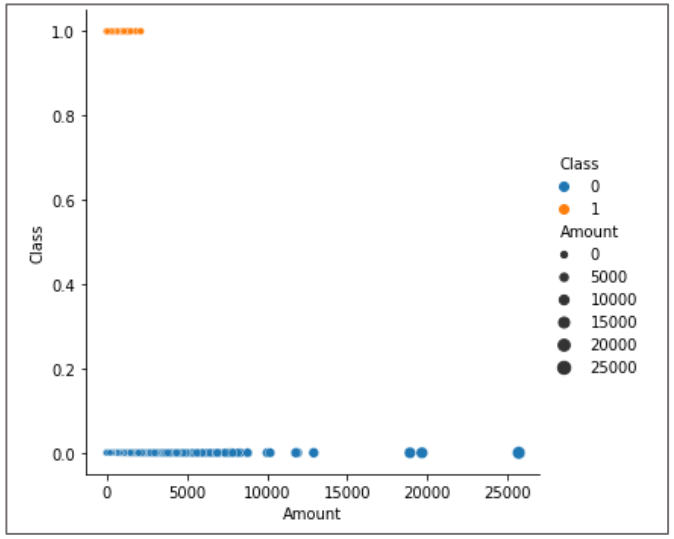
- Records : 284.807
- Rows : 30
- Duplicates: 1081

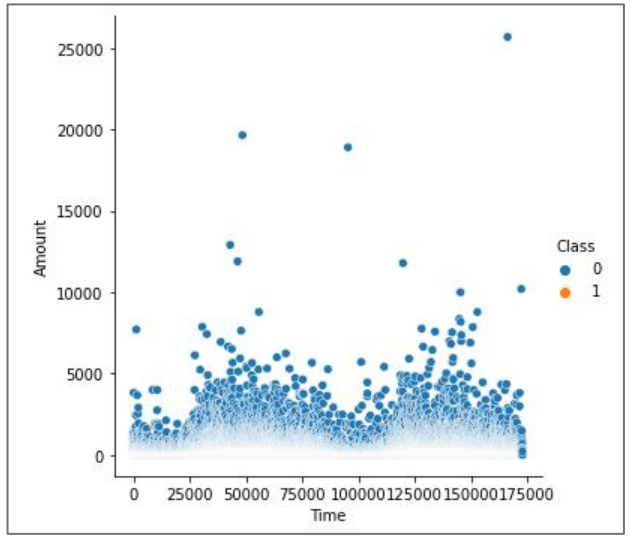
DATA ANALYSIS

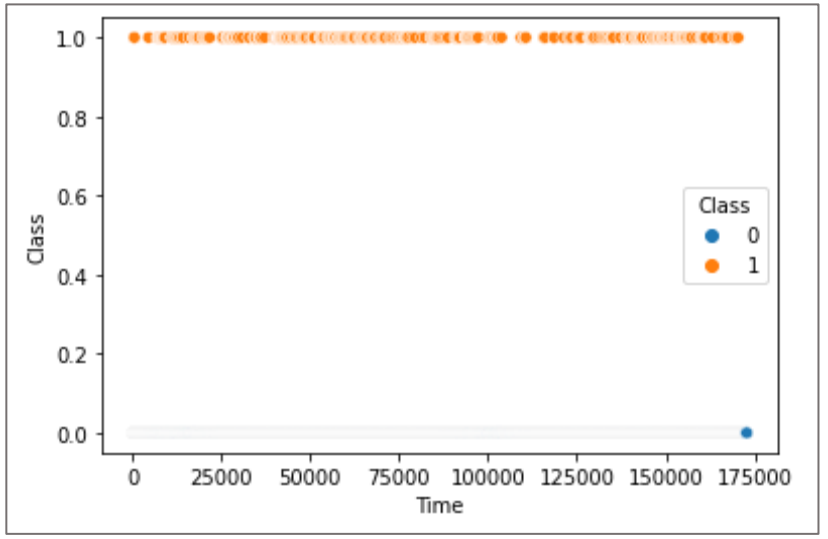




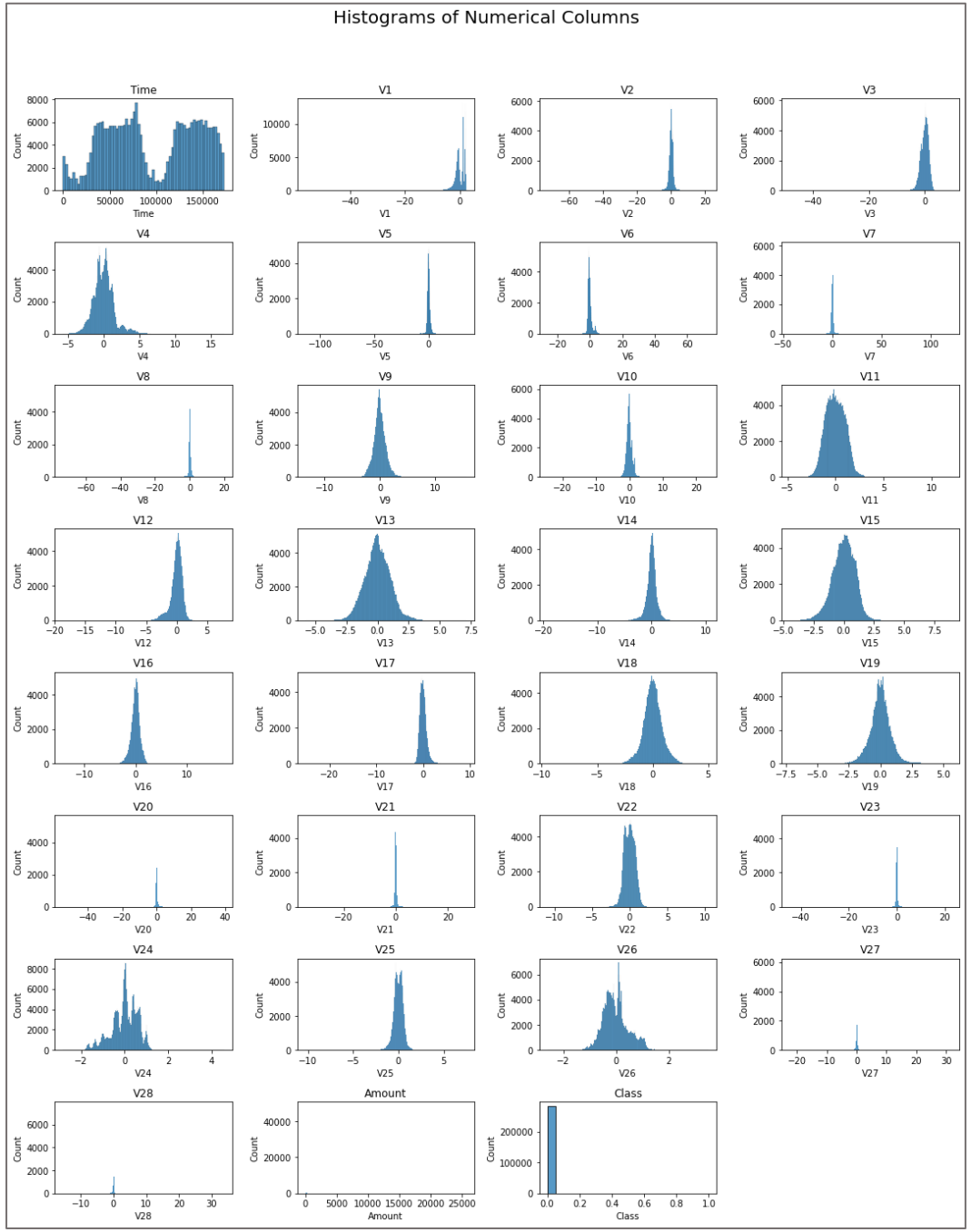
DATA ANALYSIS



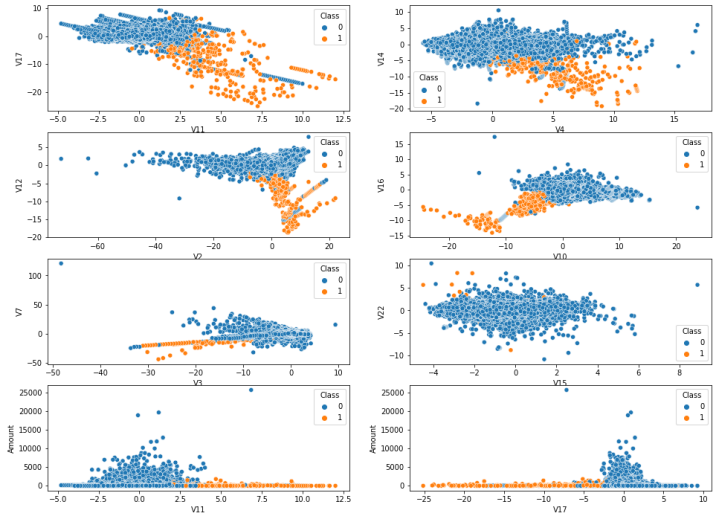


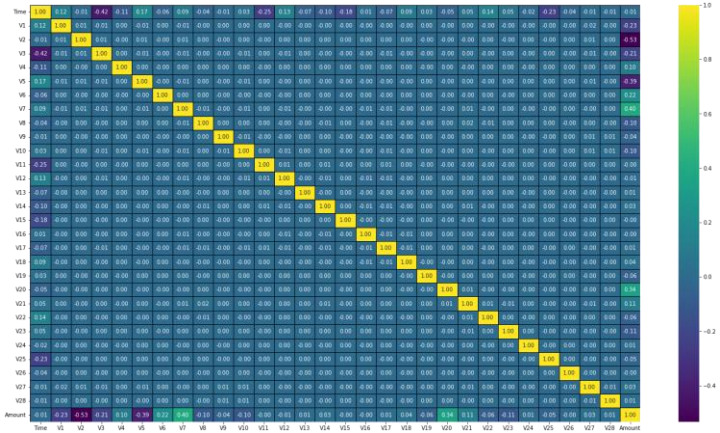


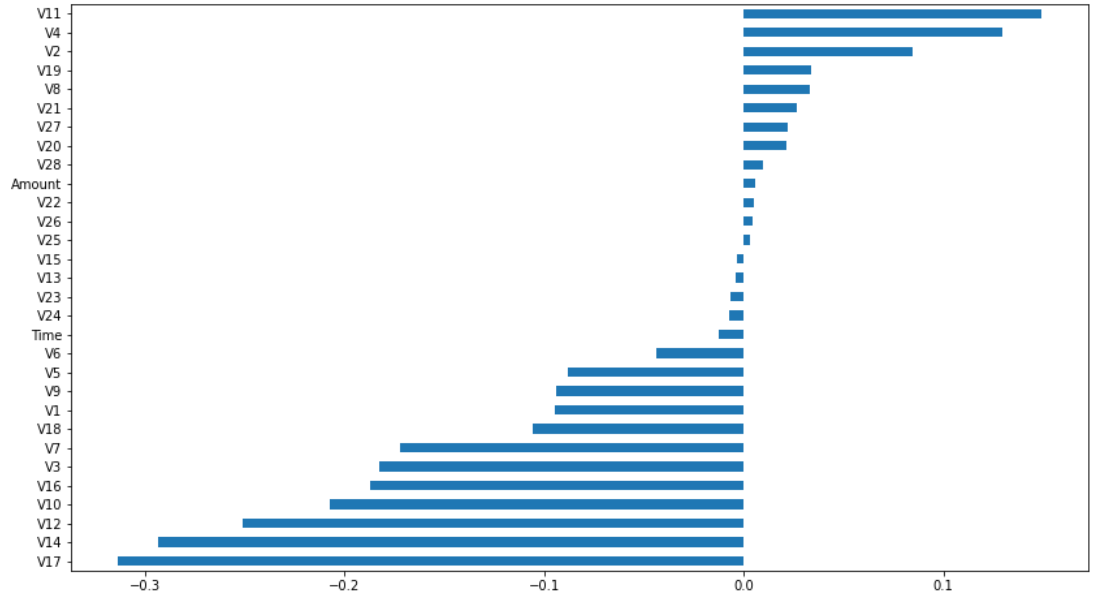
DATA ANALYSIS



DATA ANALYSIS







```
df_Fraud.shape
```

```
(473, 31)
```

```
df2 = pd.concat([df_Fraud, df_No_Fraud])
```

```
df2 = df2.sample(frac=1).reset_index(drop=True)
```

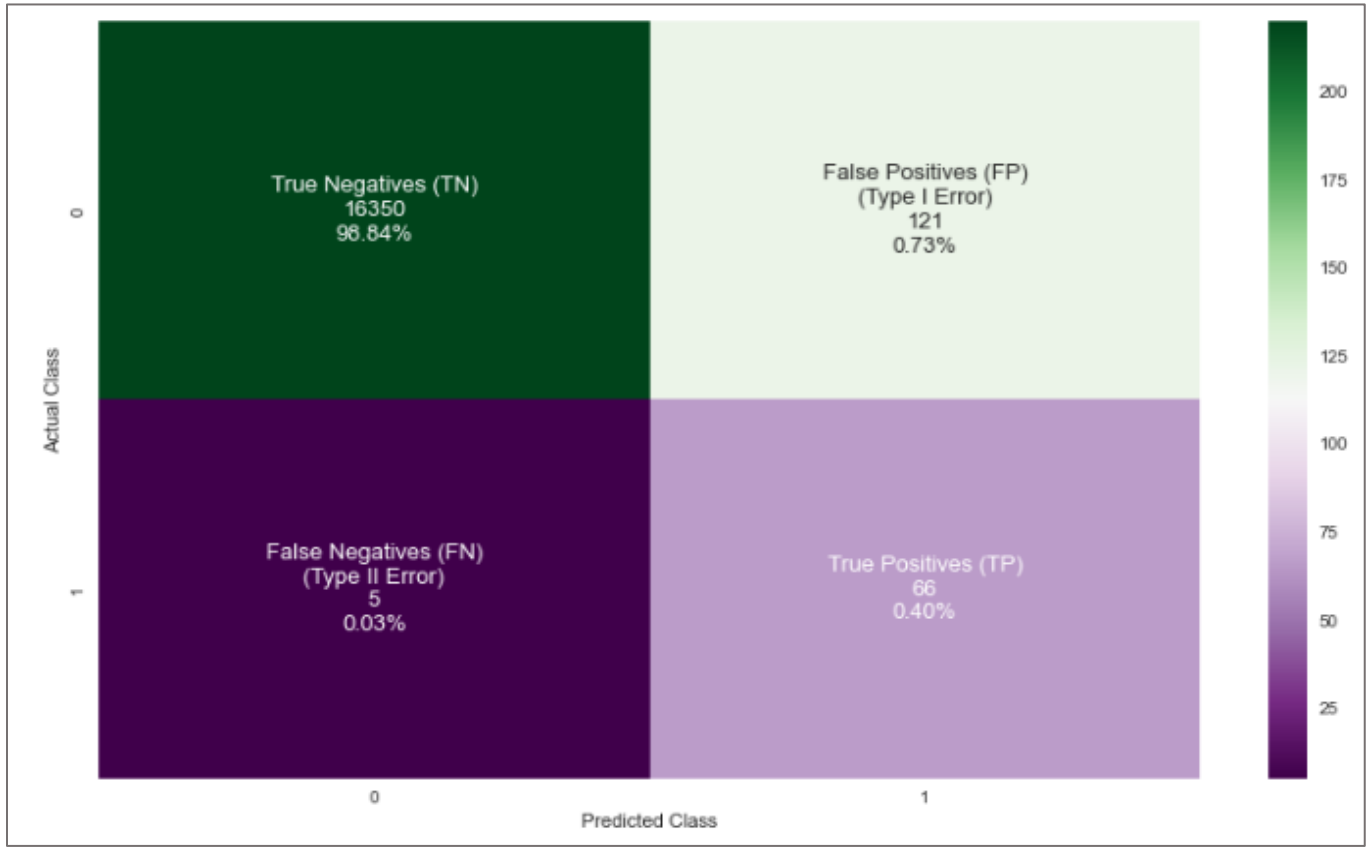
```
df2.info()
```

```
columns = df_No_Fraud.drop("Class", axis=1)
for i in columns:
    q1 = df_No_Fraud[i].quantile(0.25)
    q3 = df_No_Fraud[i].quantile(0.75)
    iqr = q3-q1 #Interquartile range
    fence_low  = q1-1.5*iqr
    fence_high = q3+1.5*iqr
    df_No_Fraud = df_No_Fraud.loc[(df_No_Fraud[i] > fence_low) & (df_No_Fraud[i] < fence_high)]
```

```
df2["Class"].value_counts()
```

```
0    109806
1       473
Name: Class, dtype: int64
```

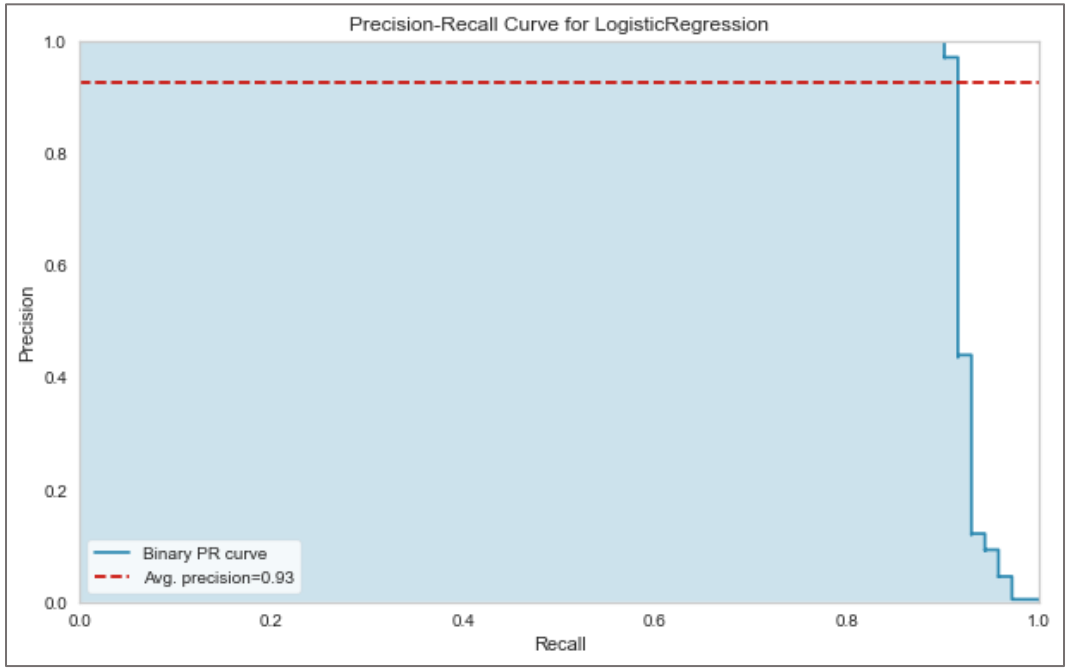
EDA




```
Test_Set
[[16350   121]
 [    5    66]]
              precision    recall  f1-score   support

           0       1.00      0.99      1.00     16471
           1       0.35      0.93      0.51        71

    accuracy                           0.99     16542
   macro avg       0.68      0.96      0.75     16542
weighted avg       1.00      0.99      0.99     16542


Train_Set
[[92604   731]
 [    8   394]]
              precision    recall  f1-score   support

           0       1.00      0.99      1.00     93335
           1       0.35      0.98      0.52       402

    accuracy                           0.99     93737
   macro avg       0.68      0.99      0.76     93737
weighted avg       1.00      0.99      0.99     93737
```

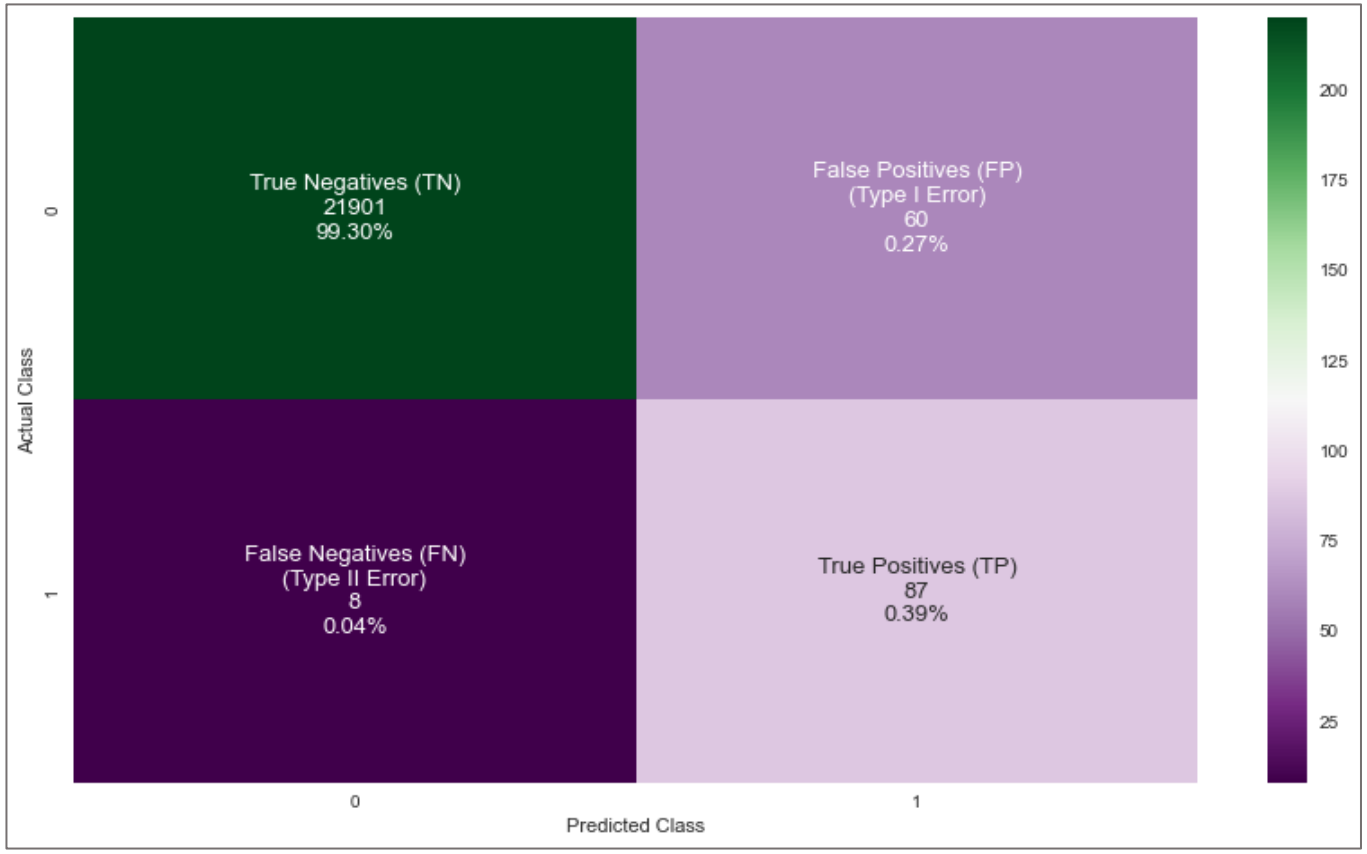
Logistic Regression without SMOTE

	train_set	test_set
Accuracy	0.992134	0.992247
Precision	0.350662	0.350394
Recall	0.981481	0.936842
f1	0.516713	0.510029
roc_auc	0.986830	0.964664
recall_auc	0.643754	0.643754



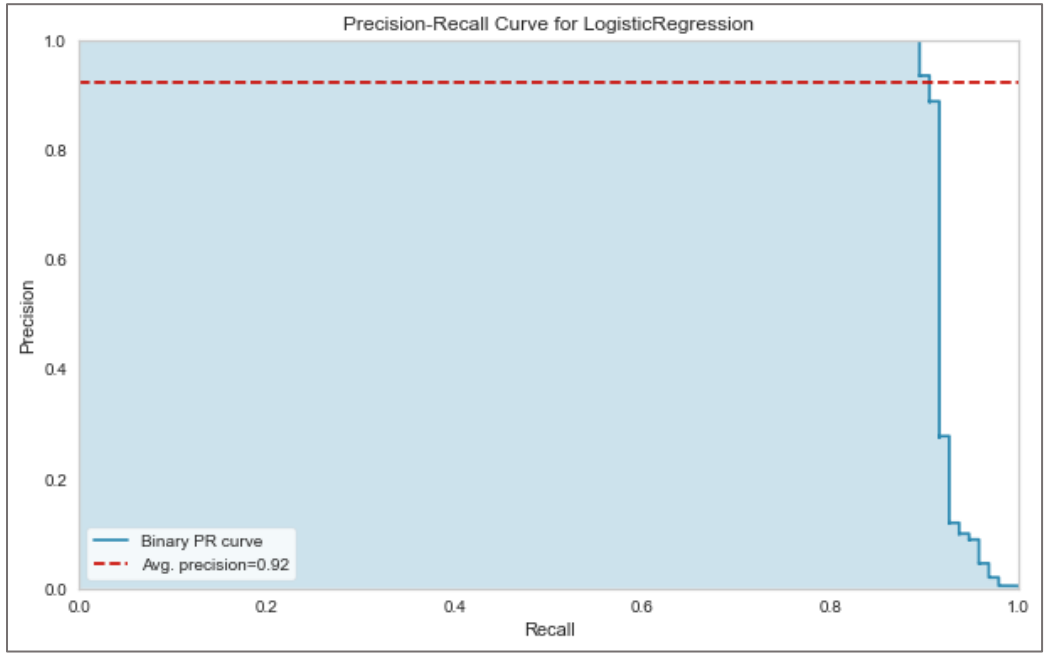
Logistic Regression without SMOTE

```
Test_Set
[[21901    60]
 [    8    87]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     21961
           1       0.59      0.92      0.72        95

    accuracy                           1.00     22056
   macro avg       0.80      0.96      0.86     22056
weighted avg       1.00      1.00      1.00     22056


Train_Set
[[87581   264]
 [   11   367]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     87845
           1       0.58      0.97      0.73       378

    accuracy                           1.00     88223
   macro avg       0.79      0.98      0.86     88223
weighted avg       1.00      1.00      1.00     88223
```

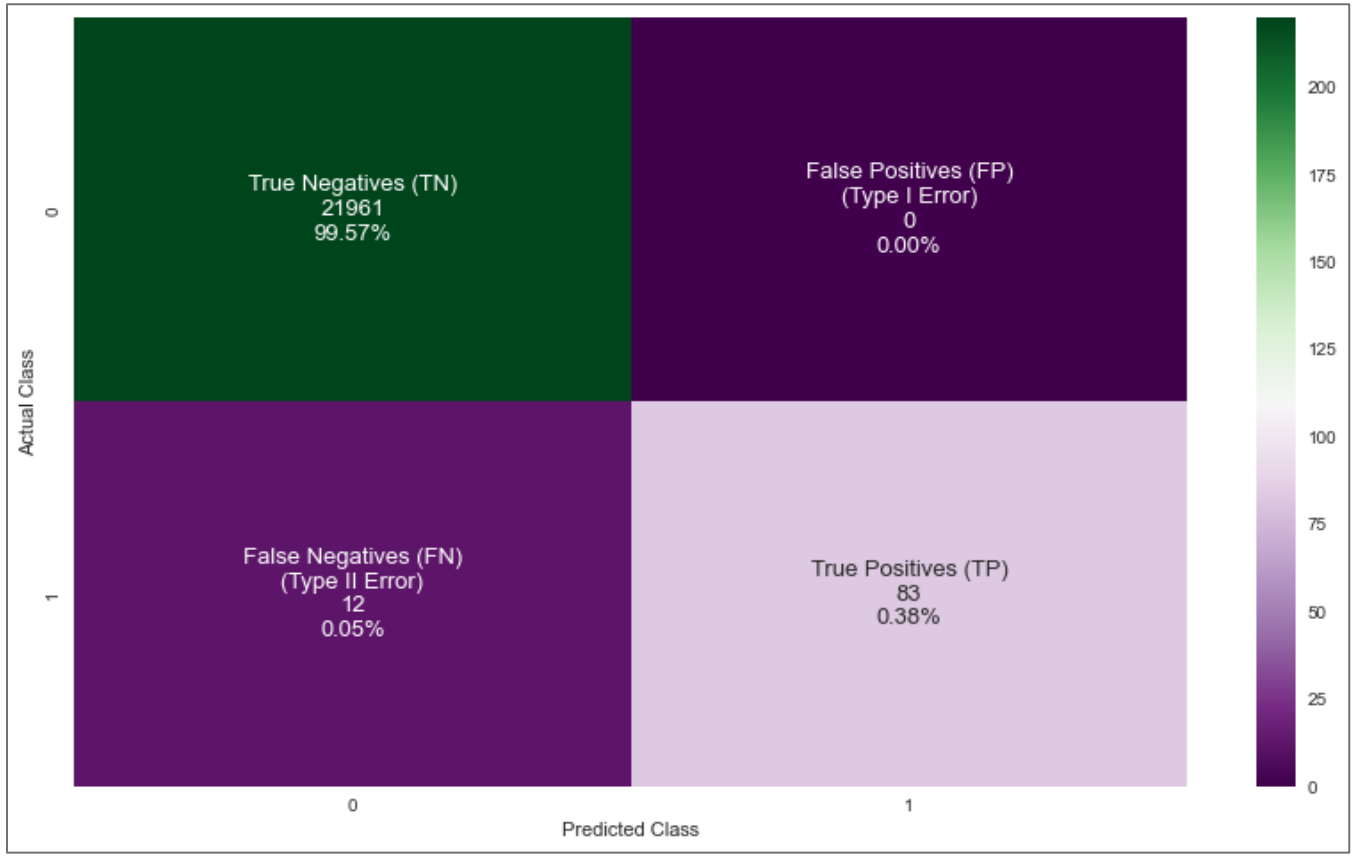


Logistic Regression with SMOTE

	train_set	test_set
Accuracy	0.992134	0.996962
Precision	0.350662	0.595890
Recall	0.981481	0.915789
f1	0.516713	0.721992
roc_auc	0.986830	0.956551
recall_auc	0.756021	0.756021



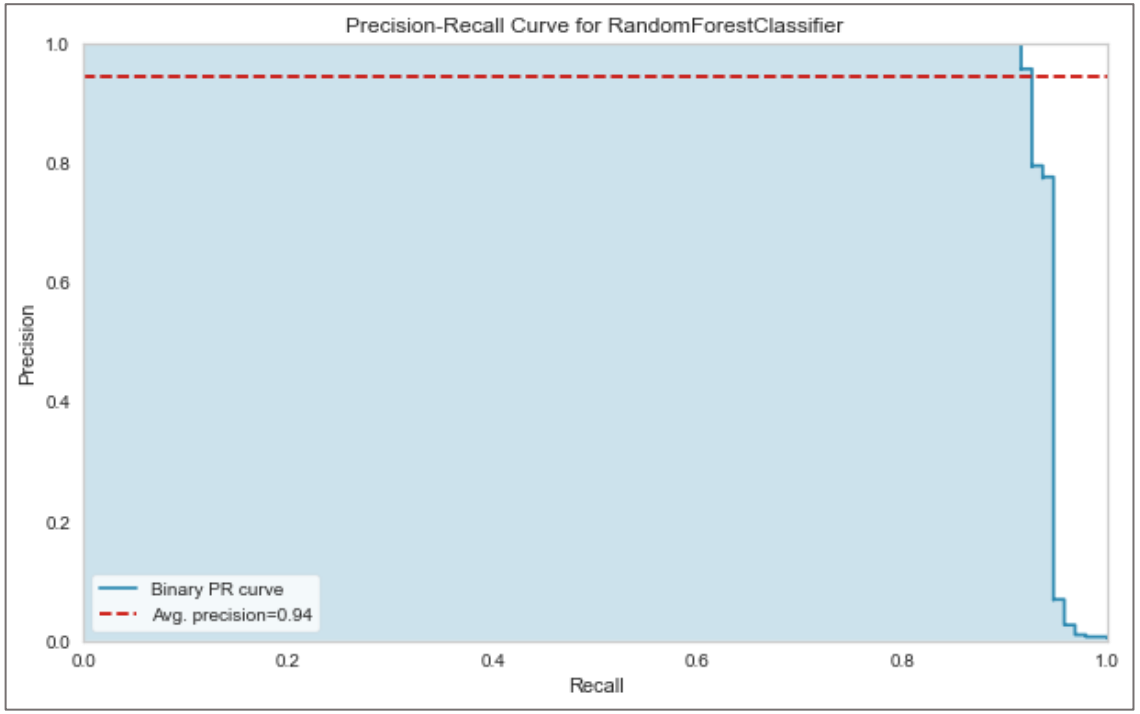
Logistic Regression with SMOTE




```
Test_Set
[[21961     0]
 [   12    83]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     21961
           1       1.00      0.87      0.93        95

    accuracy                           1.00     22056
   macro avg       1.00      0.94      0.97     22056
weighted avg       1.00      1.00      1.00     22056


Train_Set
[[87845     0]
 [   21   357]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     87845
           1       1.00      0.94      0.97       378

    accuracy                           1.00     88223
   macro avg       1.00      0.97      0.99     88223
weighted avg       1.00      1.00      1.00     88223
```

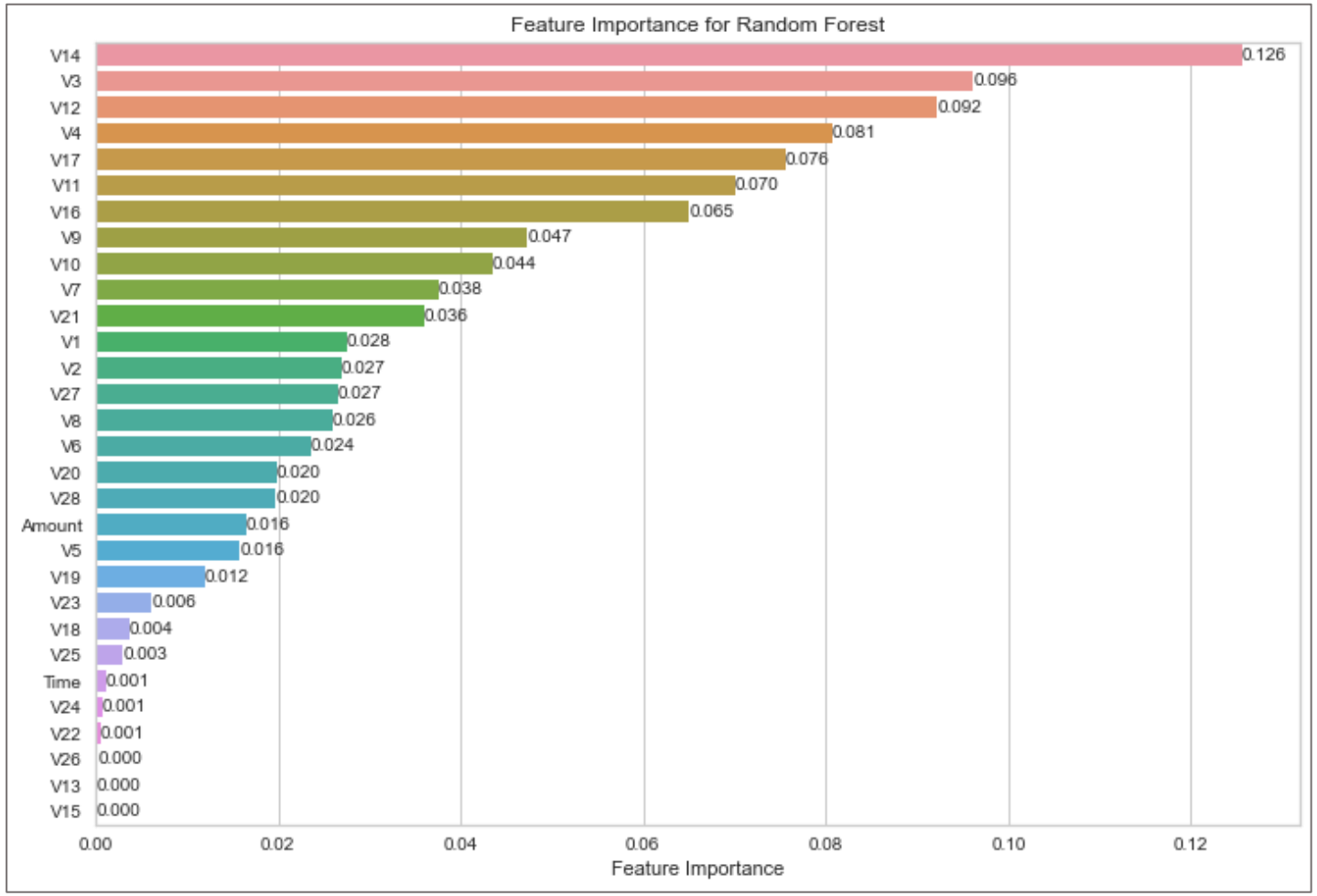
Random Forest Classifier without SMOTE

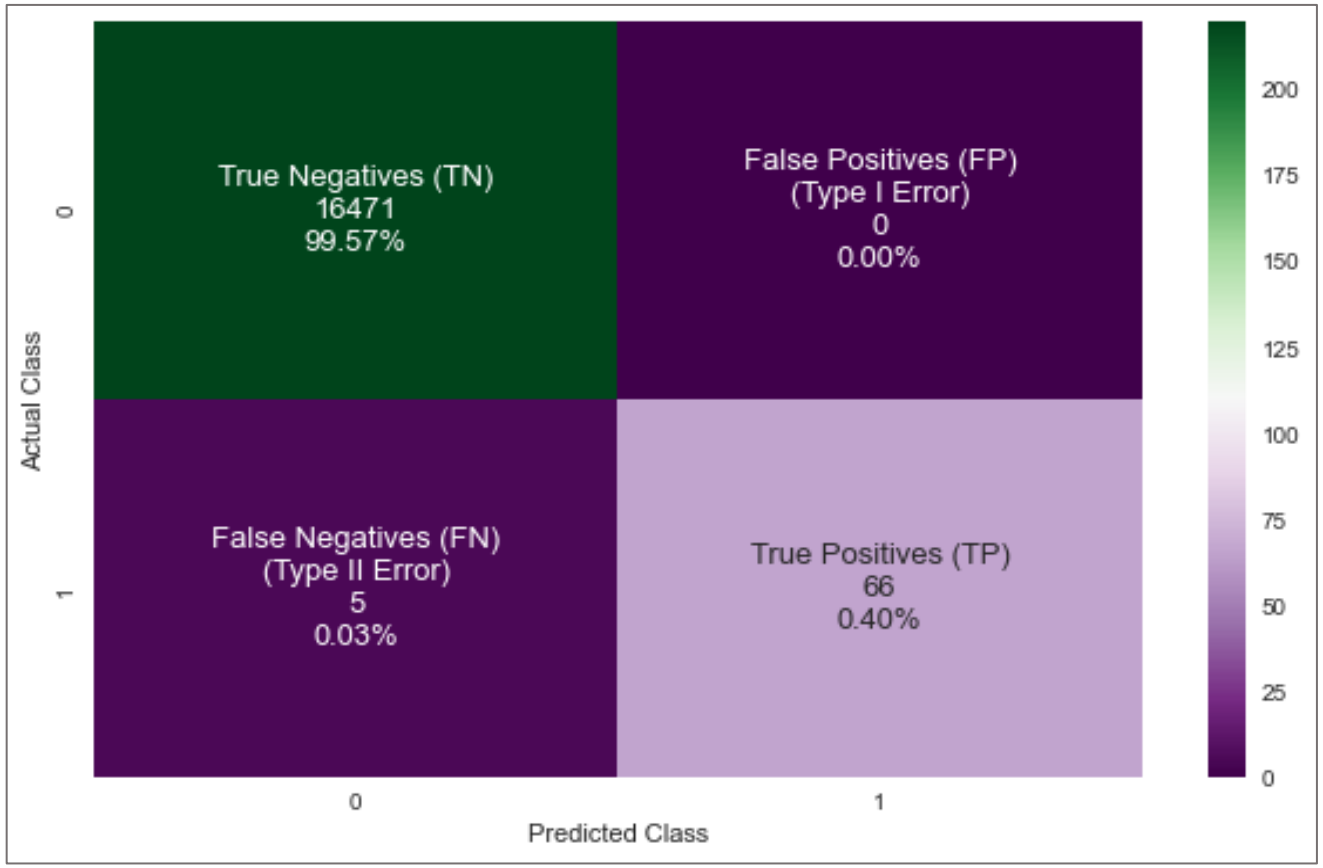
	train_set	test_set
Accuracy	0.999762	0.999456
Precision	1.000000	1.000000
Recall	0.944444	0.873684
f1	0.971429	0.932584
roc_auc	0.972222	0.936842
recall_auc	0.937114	0.937114



Random Forest Classifier with SMOTE

Random Forest Classifier with SMOTE *Feature-Importance*

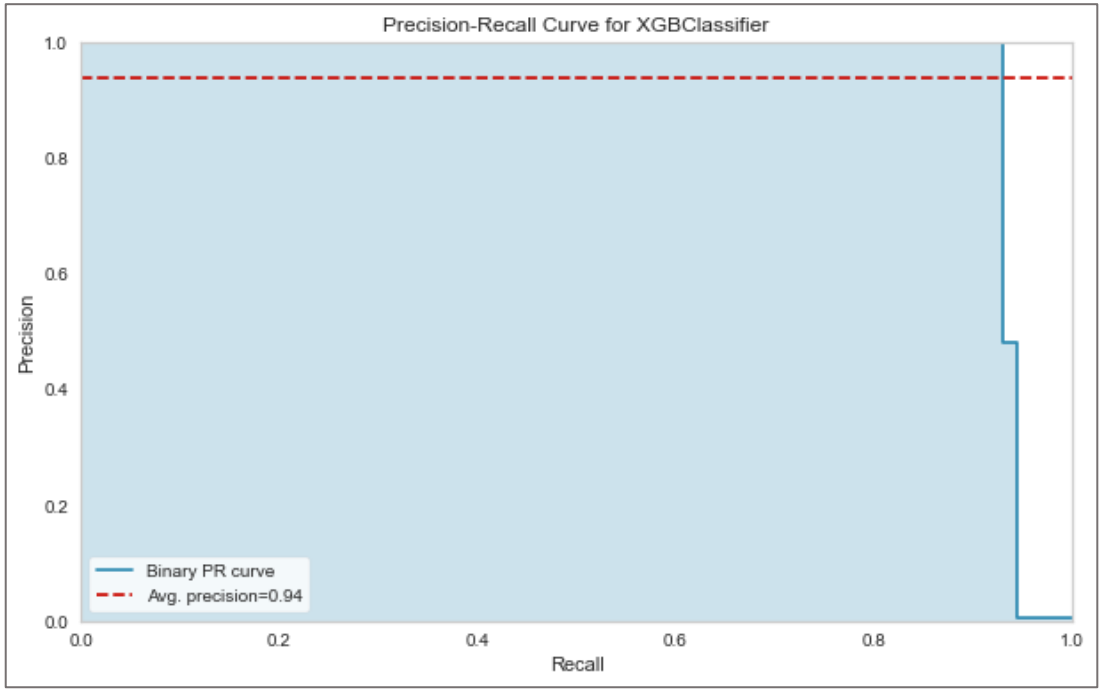




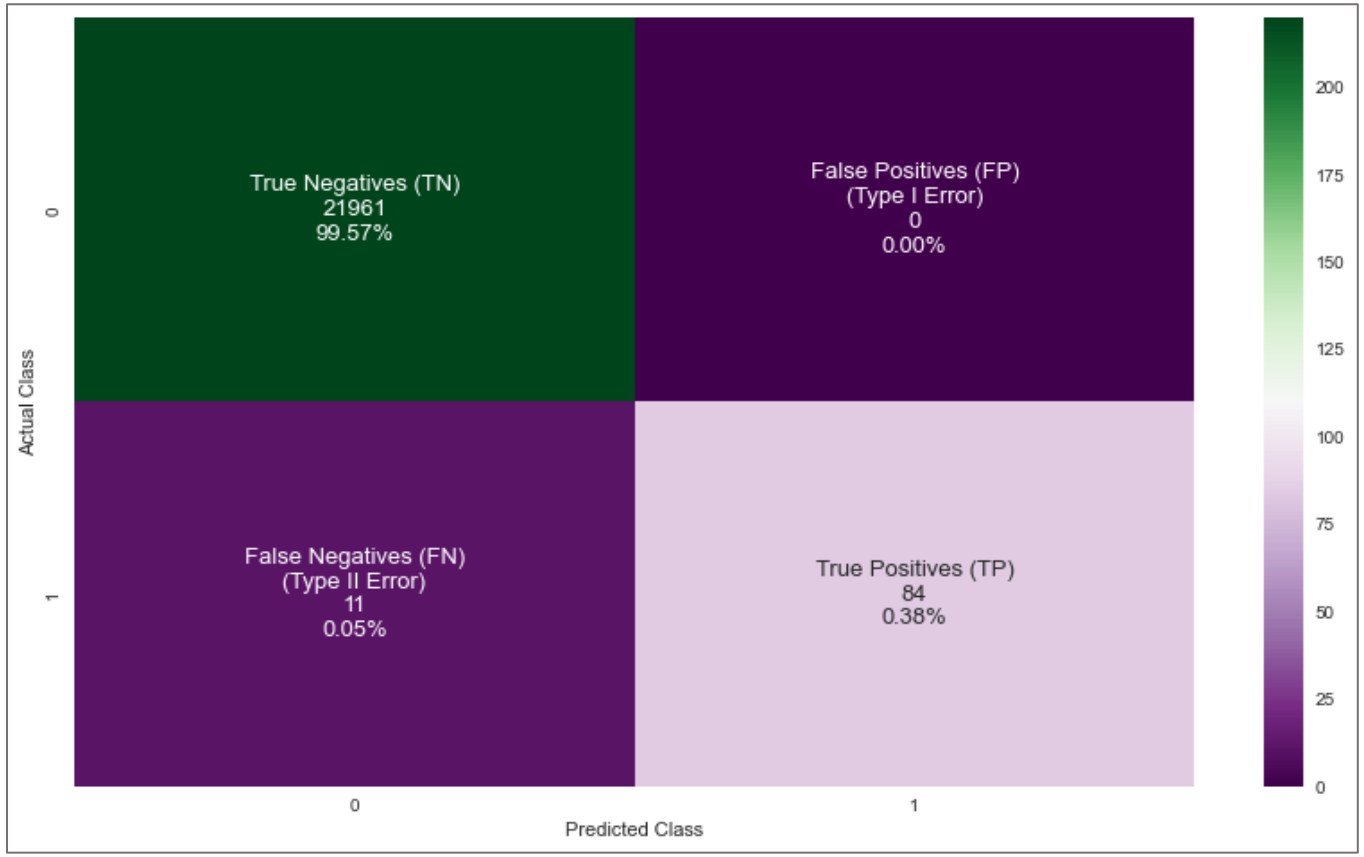

```
Test_Set
[[16471     0]
 [    5    66]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     16471
           1       1.00      0.93      0.96        71

    accuracy                           1.00     16542
   macro avg       1.00      0.96      0.98     16542
weighted avg       1.00      1.00      1.00     16542


Train_Set
[[93335     0]
 [   19   383]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     93335
           1       1.00      0.95      0.98       402

    accuracy                           1.00     93737
   macro avg       1.00      0.98      0.99     93737
weighted avg       1.00      1.00      1.00     93737
```

XGB Classifier without SMOTE

	train_set	test_set
Accuracy	0.999797	0.999698
Precision	1.000000	1.000000
Recall	0.952736	0.929577
f1	0.975796	0.963504
roc_auc	0.976368	0.964789
recall_auc	0.964940	0.964940



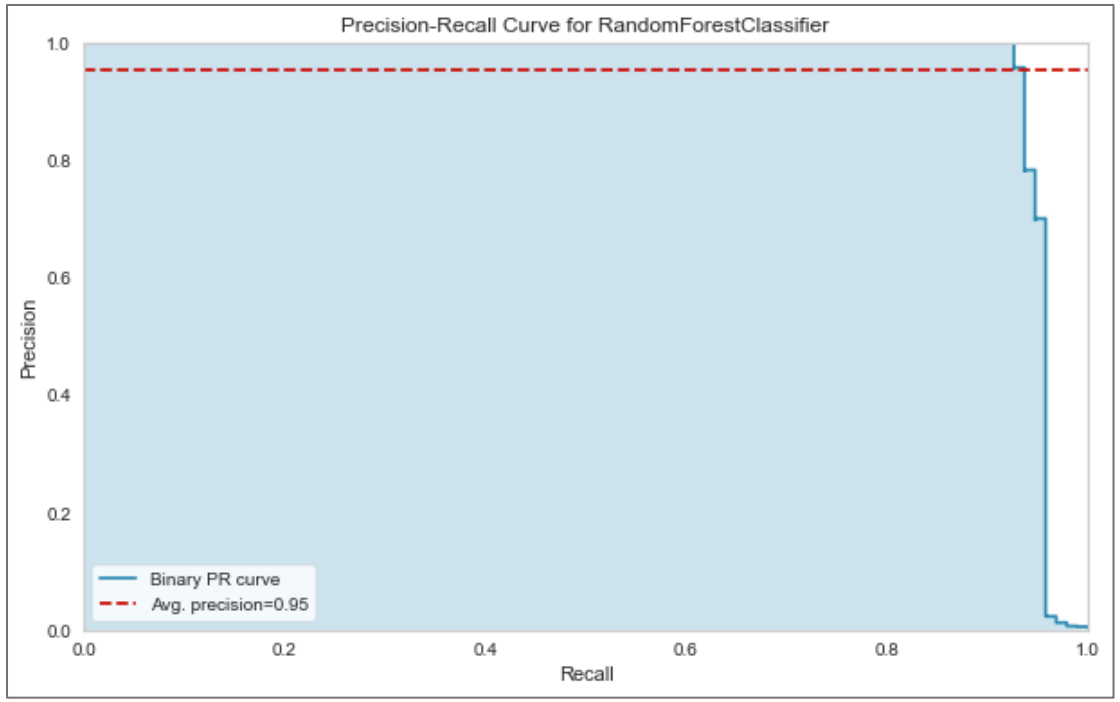
XGB Classifier without SMOTE




```
Test_Set
[[21961     0]
 [   11    84]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     21961
           1       1.00      0.88      0.94        95

    accuracy                           1.00     22056
   macro avg       1.00      0.94      0.97     22056
weighted avg       1.00      1.00      1.00     22056


Train_Set
[[87845     0]
 [   23   355]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     87845
           1       1.00      0.94      0.97       378

    accuracy                           1.00     88223
   macro avg       1.00      0.97      0.98     88223
weighted avg       1.00      1.00      1.00     88223
```

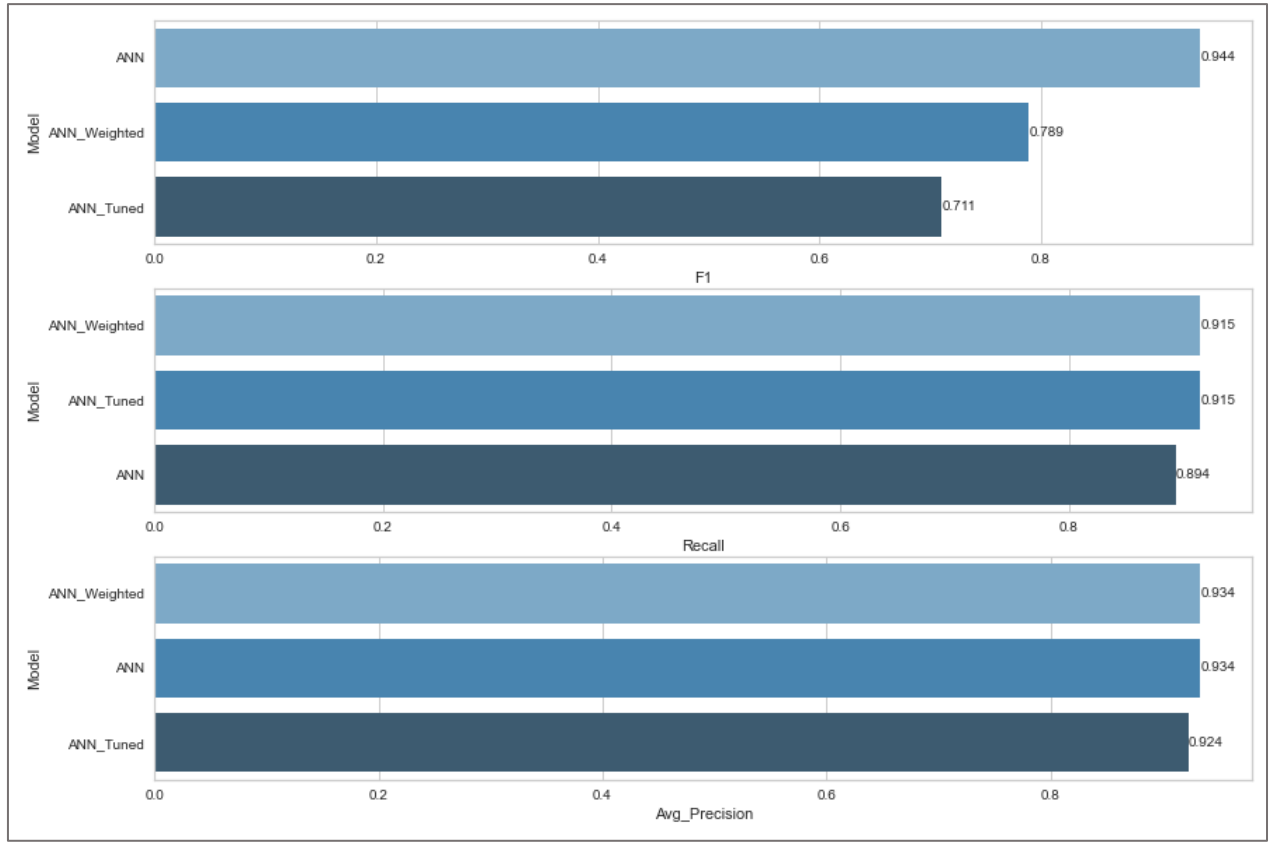
XGB Classifier with SMOTE

	train_set	test_set
Accuracy	0.999989	0.999683
Precision	0.997361	1.000000
Recall	1.000000	0.926316
f1	0.998679	0.961749
roc_auc	0.999994	0.963158
recall_auc	0.963317	0.963317



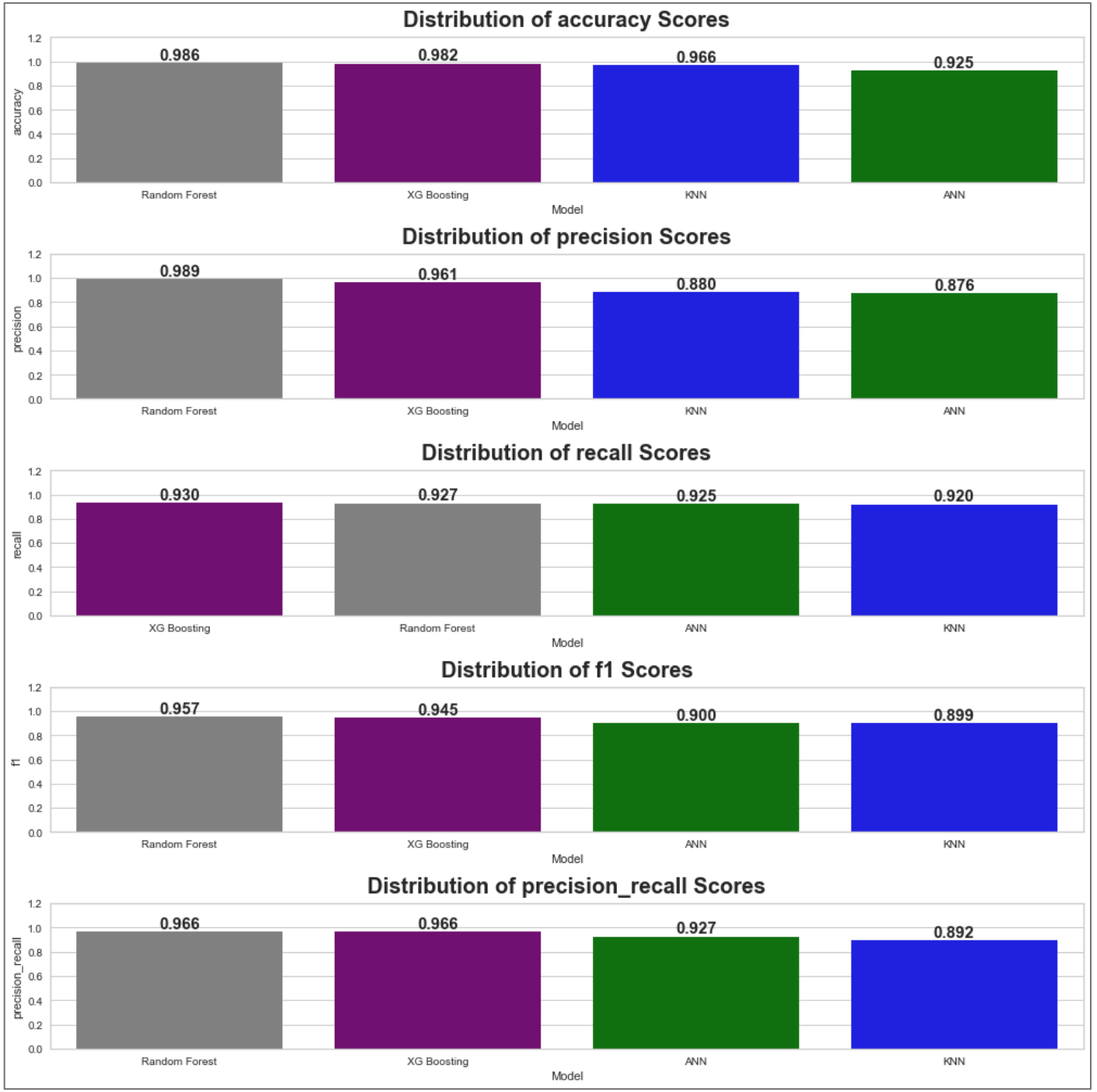
XGB Classifier with SMOTE

	train_set	test_set
Accuracy	0.999345	0.998912
Precision	0.918794	0.857143
Recall	0.929577	0.893617
f1	0.924154	0.875000
roc_auc	0.964612	0.946490
recall_auc	0.875607	0.875607



ANN

COMPARISON OF MODELS



	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V1
242763	151,637.0000	-1.2714	0.8585	-0.0084	-0.4146	0.0777	-0.0689	1.2721	0.3991	-0.6595	-1
43633	41,586.0000	-0.3481	0.7840	1.0291	0.9633	0.8802	1.0956	0.5998	0.2331	-0.7193	-0
153819	100,200.0000	2.0272	-0.0630	-0.7288	0.2303	-0.1056	-0.6661	-0.2370	-0.1626	1.6373	-0
12442	21,825.0000	1.0126	0.1854	1.8570	2.6257	-0.8137	0.7412	-1.0505	0.4114	1.0782	0
75593	56,144.0000	-0.7583	0.3529	2.6328	0.3969	-0.5290	0.0648	0.2130	-0.3154	-1.0414	1

MODEL DEPLOYMENT