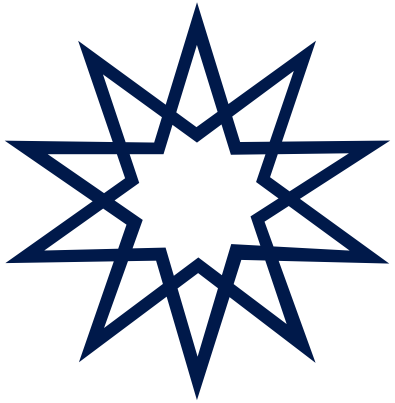
PROJECT PRESENTATION - CREDIT CARD FRAUD DETECTION

BLM5110 Machine Learning

ÖNDER GÖRMEZ

21501035



YTÜ YILDIZ TEKNİK ÜNİVERSİTESİ

Agenda

I. Literature Review

II. Dataset

III. Experiments and Results

IV. Conclusions

V. References

VI. Q & A

Literature Review

Dolandırıcılık Tespit Yöntemleri:

- Geleneksel Dolandırıcılık Tespit Yöntemleri

- Geleneksel Makine Öğrenmesi Yöntemleri
  - Decision Trees
  - K-Nearest Neighbors (KNNs)
  - Support Vector Machine (SVM)

- Derin Öğrenme Yöntemleri
  - Yapay Sinir Ağları (ANN)
  - Konvolüsyonel Sinir Ağları (CNN)
  - Recurrent Neural Networks (RNN)

Literature Review

2020 International Conference on E-Commerce and Internet Technology (ECIT)

Credit Card Fraud Detection Using Lightgbm Model

Dingling Ge,
Northeastern University,
Boston, United States,
ge.di@husky.neu.edu,

Shunyu Chang,
Changchun University of Science and Technology,
Jilin, China,
changshunyu@yullioner.com,

Jianyang Gu,
Nankai University,
Tianjin, China,
gjy1198350167@163.com,

JingHui Cai,
JiNan University,
Guangzhou, China,
legolascai@163.com.

Literature Review

2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE 2021)

CatBoost for Fraud Detection in Financial Transactions

Yeming Chen
ClarityAI
Beijing, China
cymcsg@gmail.com

Xinyuan Han
ClarityAI
Beijing, China
eric@clarityai.tech

Literature Review

2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)

A Data Mining Based Fraud Detection Hybrid Algorithm in E-bank

Zijian Song
University of Rochester
Newyork, United States
zsong6@u.rochester.edu

Literature Review

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 20, 2022, accepted April 8, 2022, date of publication April 12, 2022, date of current version April 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3166891

Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms

FAWAZ KHALED ALARFAJ[1], IQRA MALIK[2], HIKMAT ULLAH KHAN[3], NAIF ALMUSALLAM[1], MUHAMMAD RAMZAN[2], AND MUZAMIL AHMED[3]

[1]Department of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11564, Saudi Arabia
[2]Department of Computer Science and Information Technology, University of Sargodha, Sargodha 40100, Pakistan
[3]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt 47040, Pakistan

Corresponding author: Hikmat Ullah Khan (hikmat.ullah@ciitwah.edu.pk)

This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University through the Research Group under Grant RG-21-51-01.

Literature Review

Received 29 February 2024, accepted 19 March 2024, date of publication 22 March 2024, date of current version 23 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3380823

Identifying Fraudulent Credit Card Transactions Using Ensemble Learning

JABER JEMAI[1], ANIS ZARRAD[2], AND ALI DAUD[3]

[1]CIS Division, Higher Colleges of Technology, Abu Dhabi, United Arab Emirates
[2]University of Birmingham Dubai, Dubai, United Arab Emirates
[3]Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates

Corresponding author: Ali Daud (alimsdb@gmail.com)

Agenda

I. Literature Review
II. Dataset
III. Experiments and Results
IV. Conclusions
V. References
VI. Q & A

Dataset

Dataset Features:

- A research collaboration of Worldline and the Machine Learning Group (http://mlg.ulb.ac.be) of ULB (Université Libre de Bruxelles)
- Transactions, September 2013 by European cardholders
- In two days, 492 frauds out of 284,807 transactions
- Input variables which are the result of a PCA transformation
- +5K Users studied in Kaggle

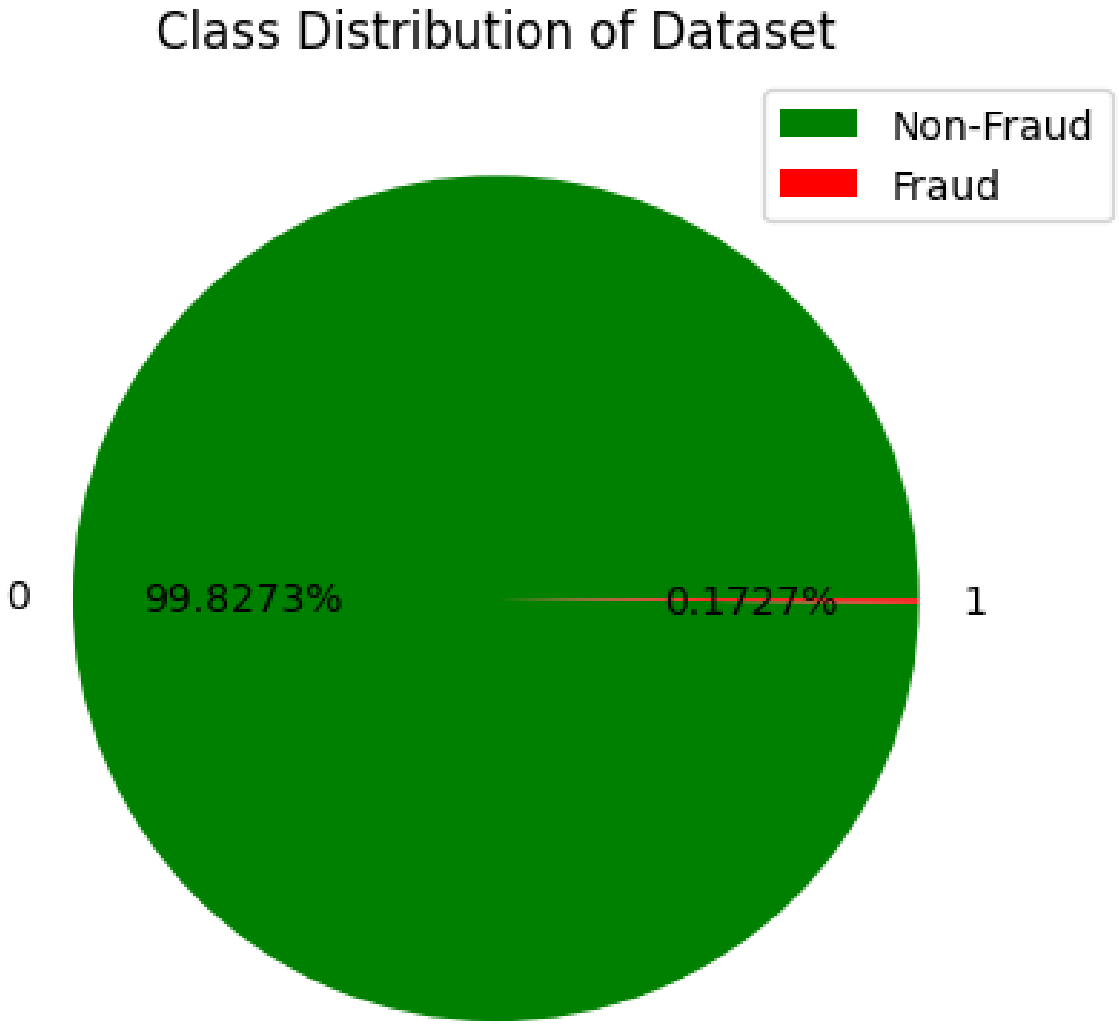
Dataset

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...

5 rows × 31 columns

...	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
...	-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.189115	0.133558	-0.021053	149.62	0
...	-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.125895	-0.008983	0.014724	2.69	0
...	0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.139097	-0.055353	-0.059752	378.66	0
...	-0.108300	0.005274	-0.190321	-1.175575	0.647376	-0.221929	0.062723	0.061458	123.50	0
...	-0.009431	0.798278	-0.137458	0.141267	-0.206010	0.502292	0.219422	0.215153	69.99	0

Dataset


Class Distribution of Dataset
Non-Fraud
Fraud
0
99.8273%
0.1727%
1

Dataset

Splitting the Dataset:

%80 Train, %20 Test set olarak ayrıldı

- Train set size: 227845
  - Non-Fraud transactions in the training set: 227451 samples, 99.8271%
  - Fraud transactions in the training set: 394 samples, 0.1729%

- Test set size: 56962
  - Non-Fraud transactions in the test set: 56864 samples, 99.8280%
  - Fraud transactions in the test set: 98 samples, 0.1720%

Agenda

I. Literature Review

II. Dataset

III. Experiments and Results

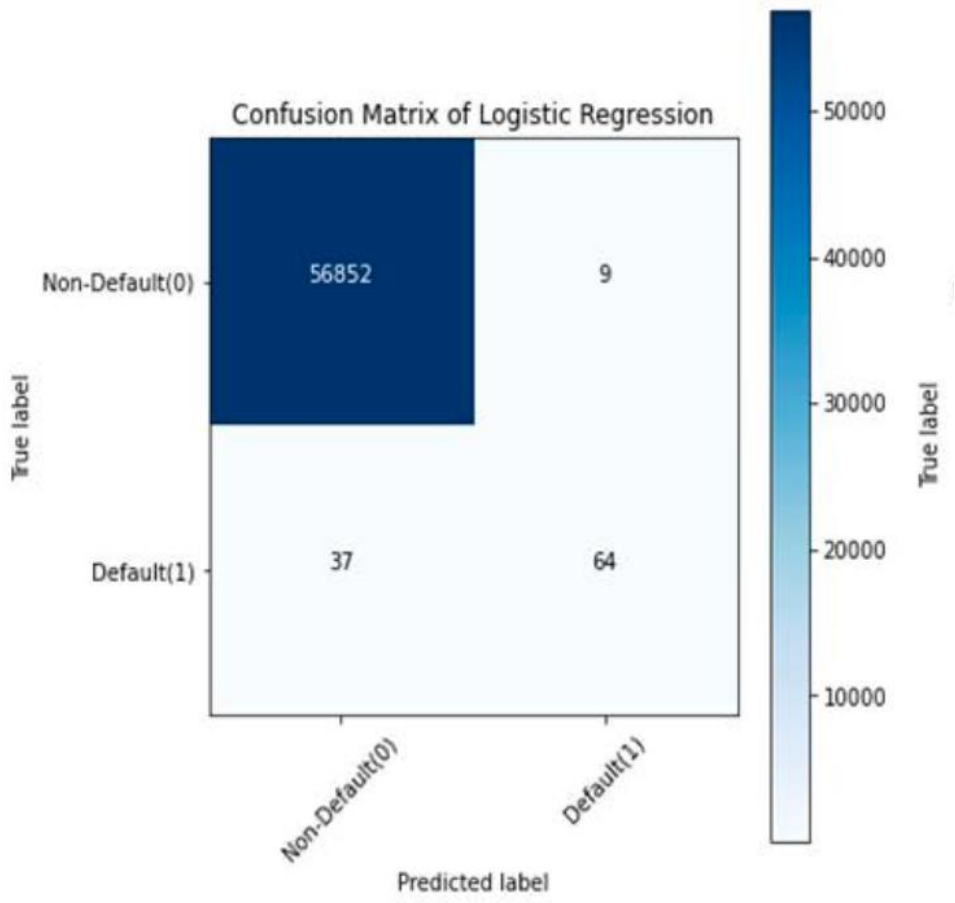
IV. Conclusions

V. References

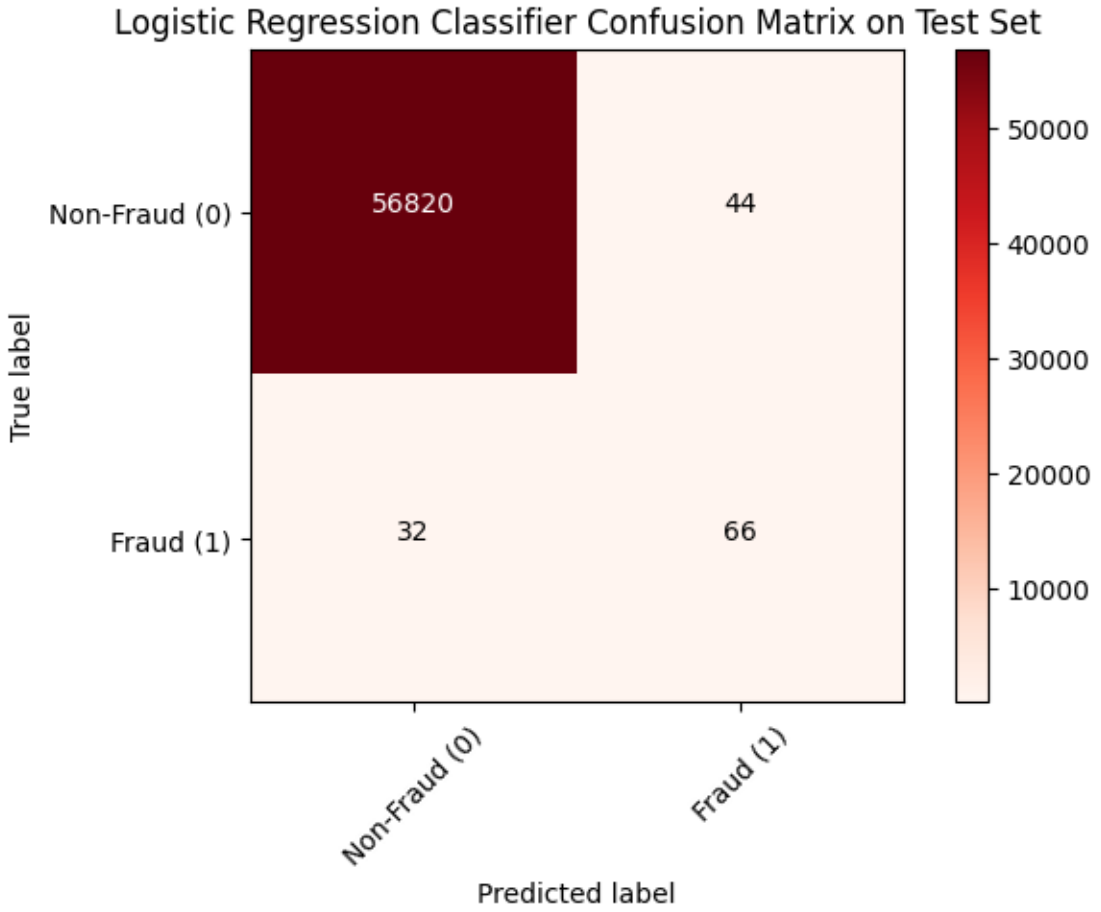
VI. Q & A

Experiments and Results - Logistic Regression

Result on Paper

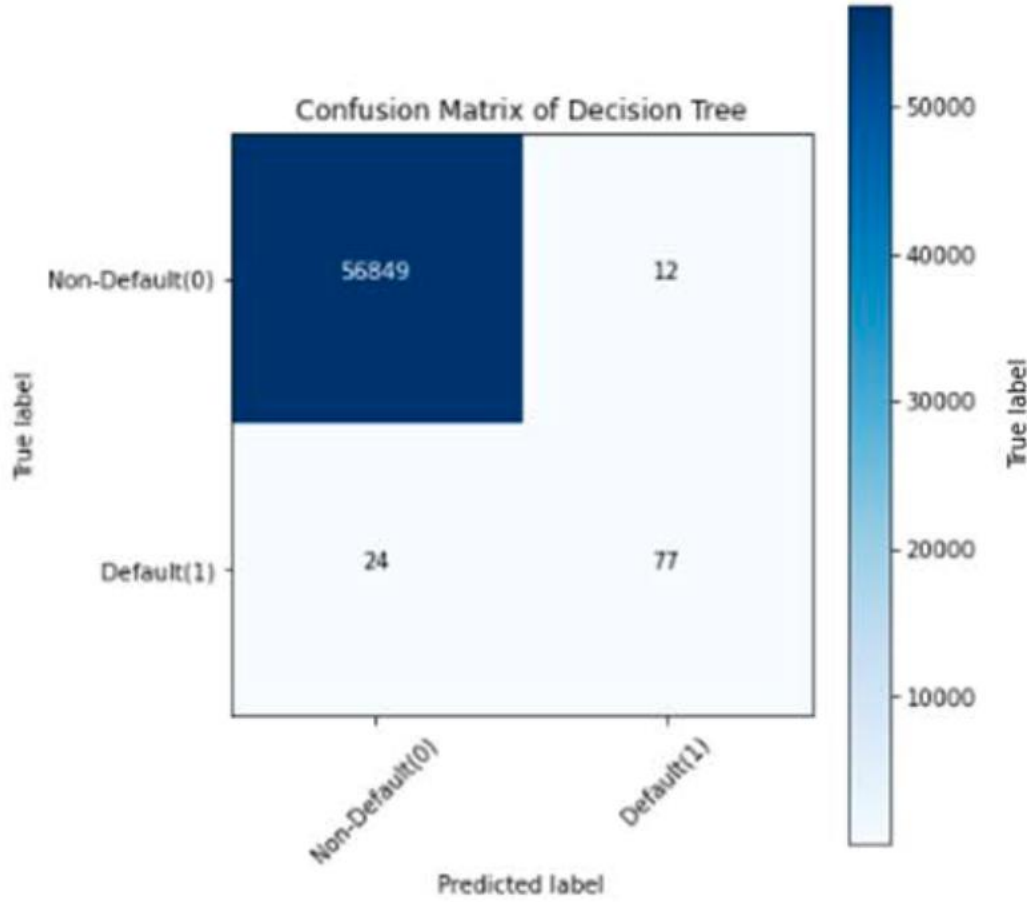


Our Result

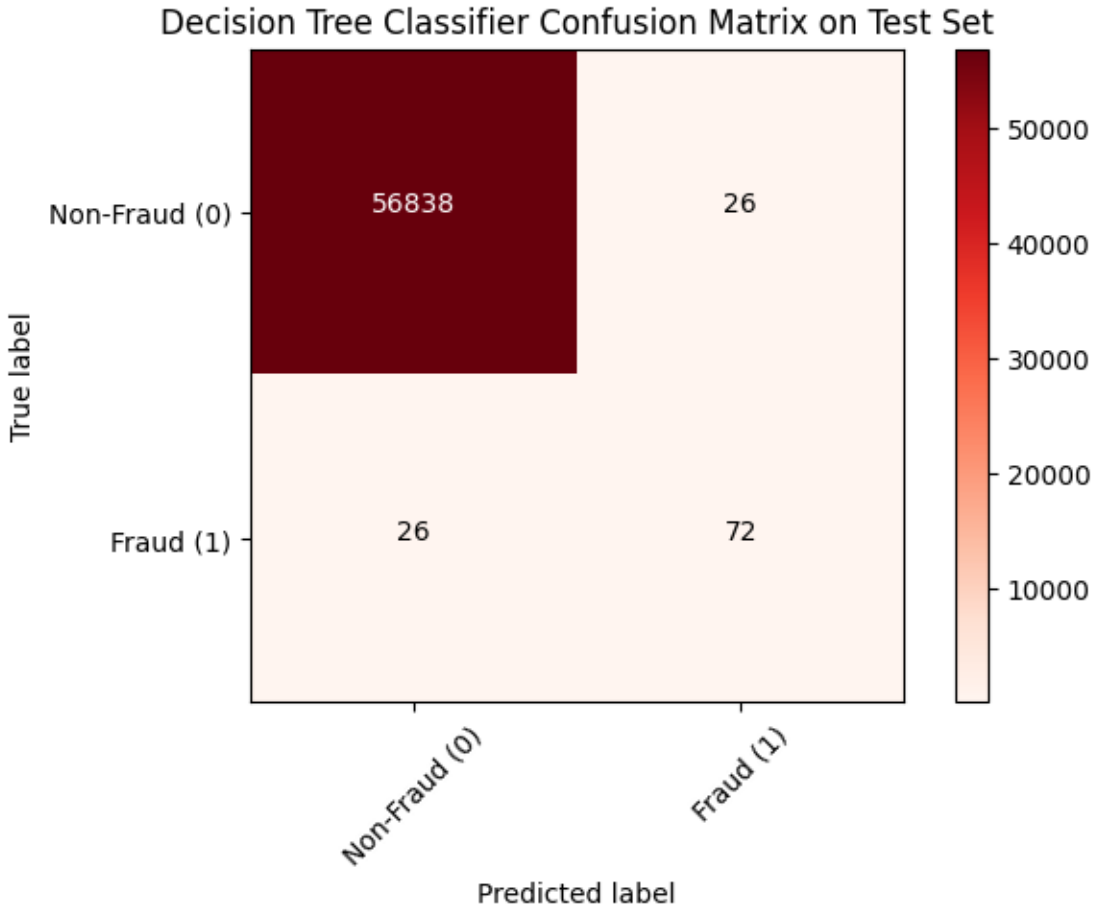


Experiments and Results - Decision Tree

Result on Paper

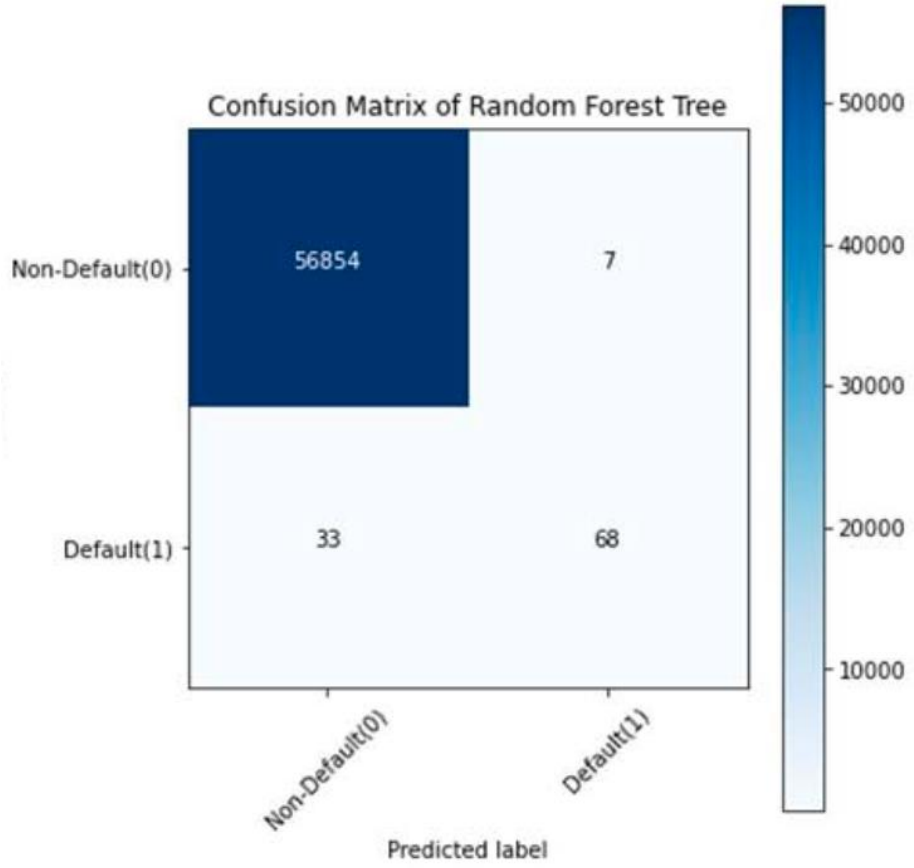


Our Result

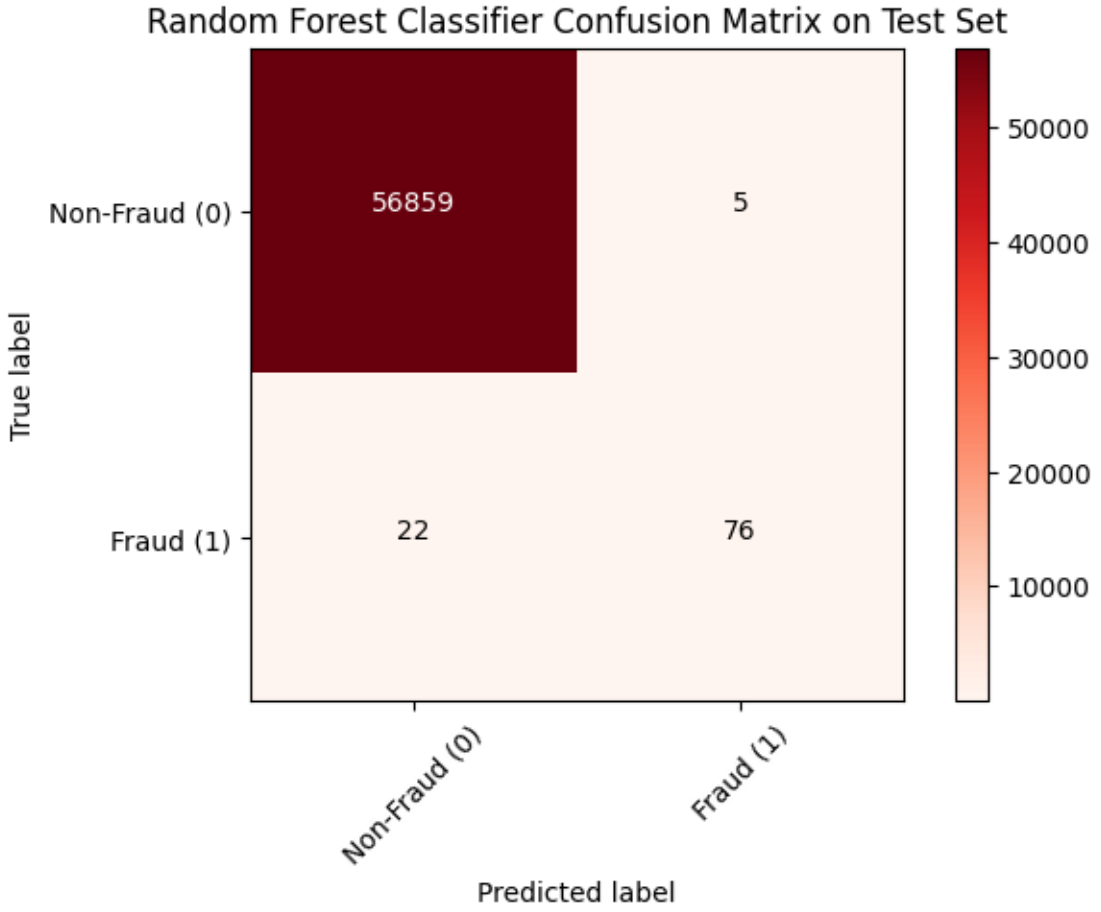


Experiments and Results - Random Forest

Result on Paper



Our Result



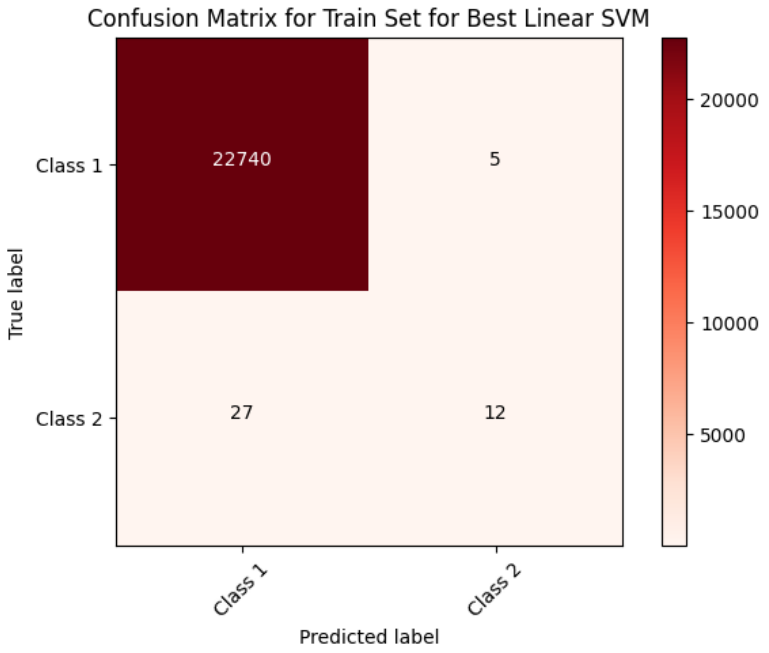
Experiments and Results - SVM - Grid Search w. 10% of Data

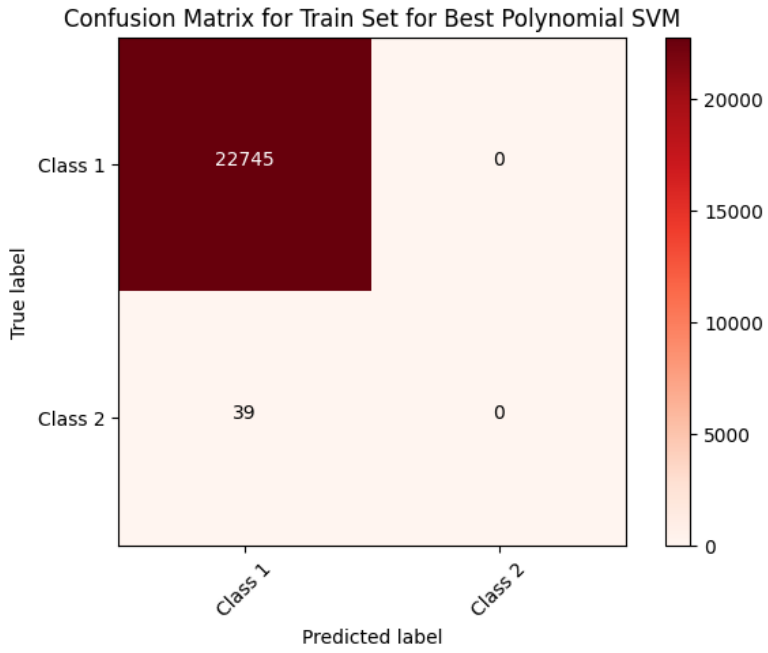
Kernel Türü	C	degree	gamma
Lineer	0.1, 1, **10**, 100	-	-
Polinomsal	**0.1**, 1, 10, 100	**2**, 3, 4, 5	-
Gaussian RBF	**0.1**, 1, 10, 100	-	**0.01**, 0.1, 1, 10

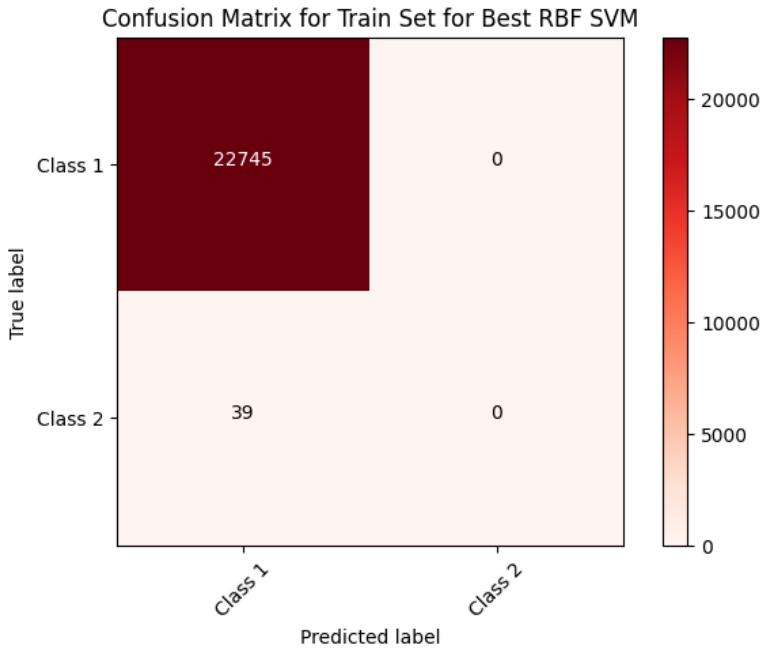
```
[INFO] [2025-01-04T18:44:57.992Z] Non-Fraud transactions in the training set: 22745 samples, 99.8288%
[INFO] [2025-01-04T18:44:57.992Z] Fraud transactions in the training set: 39 samples, 0.1712%
[INFO] [2025-01-04T18:44:57.992Z]
[INFO] [2025-01-04T18:44:57.993Z] Non-Fraud transactions in the test set: 5686 samples, 99.8244%
[INFO] [2025-01-04T18:44:57.993Z] Fraud transactions in the test set: 10 samples, 0.1756%
[INFO] [2025-01-04T18:44:57.994Z]
```

	Model	Data	Accuracy	Precision	Recall	F1 Score
0	Best Linear SVM	Train	99.86%	0.706	0.308	0.429
1	Best Linear SVM	Test	99.81%	0.429	0.300	0.353
2	Best Polynomial SVM	Train	99.83%	0.000	0.000	0.000
3	Best Polynomial SVM	Test	99.82%	0.000	0.000	0.000
4	Best RBF SVM	Train	99.83%	0.000	0.000	0.000
5	Best RBF SVM	Test	99.82%	0.000	0.000	0.000

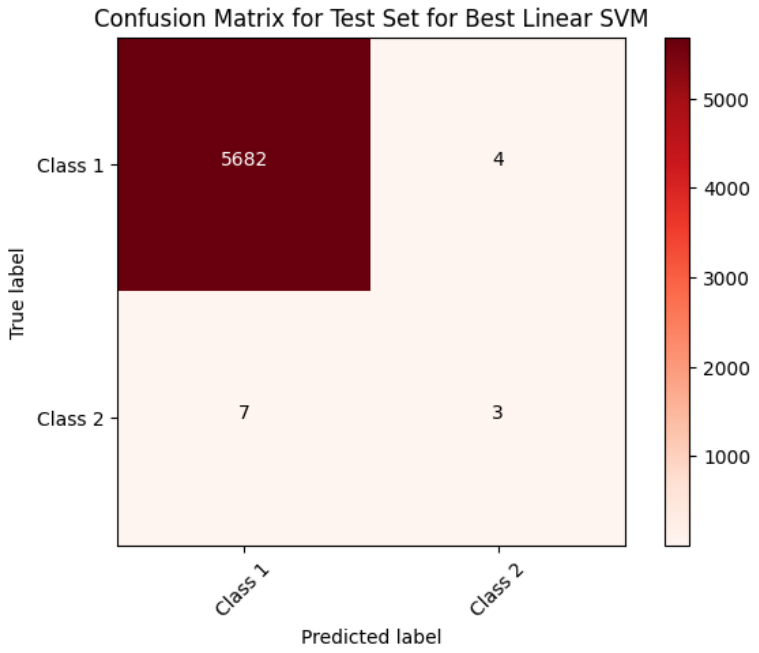
Experiments and Results - SVM - Grid Search w. 10% of Data

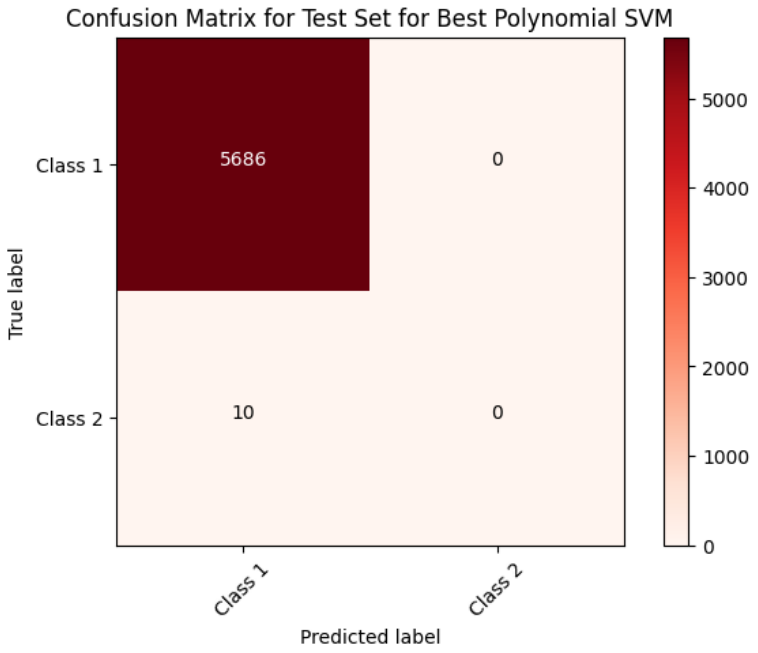


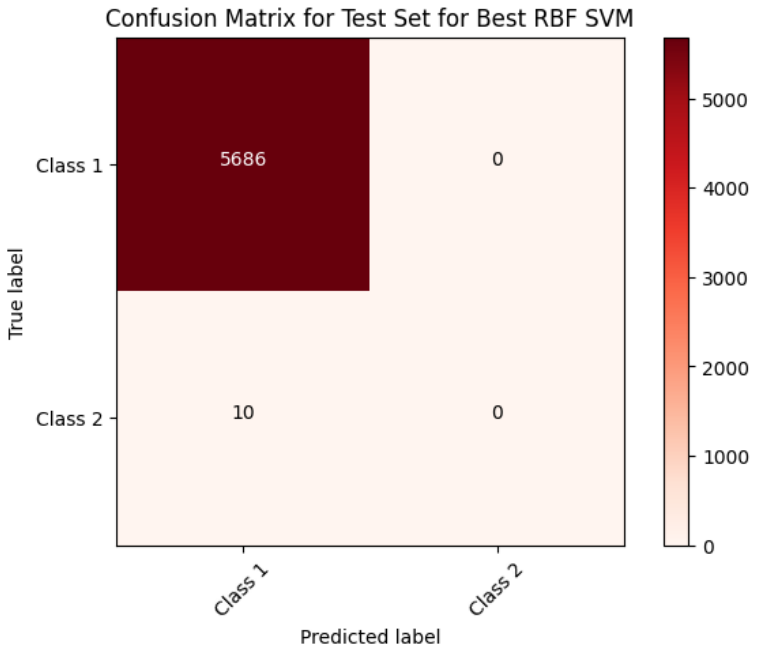




Experiments and Results - SVM - Grid Search w. 10% of Data







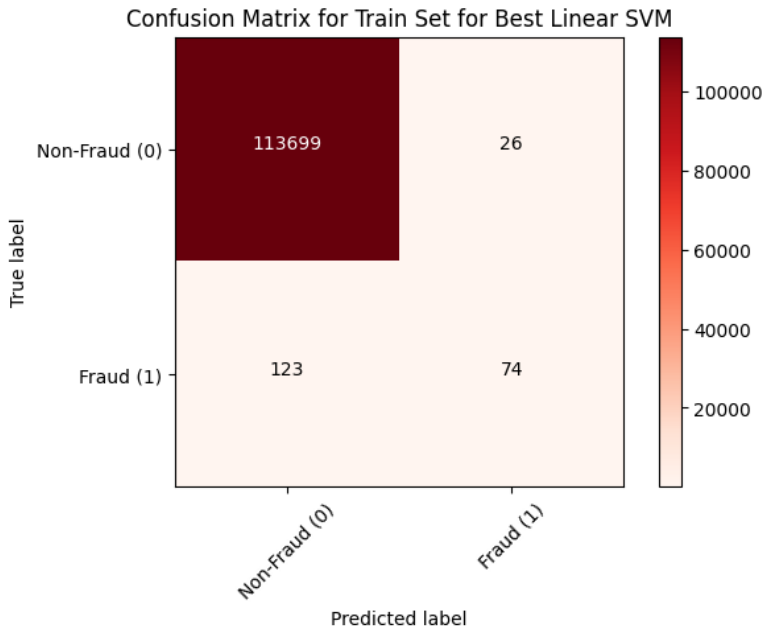
Experiments and Results - SVM - Grid Search w. 50% of Data

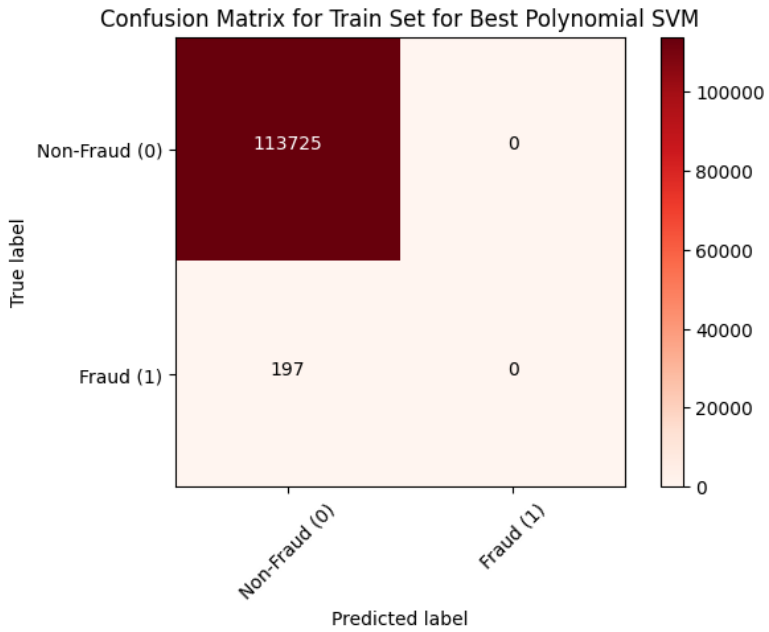
Kernel Türü	C	degree	gamma
Lineer	0.1, **1**, 10, 100	-	-
Polinomsal	**0.1**, 1, 10, 100	**2**, 3, 4, 5	-
Gaussian RBF	0.1, **1**, 10, 100	-	**0.01**, 0.1, 1, 10

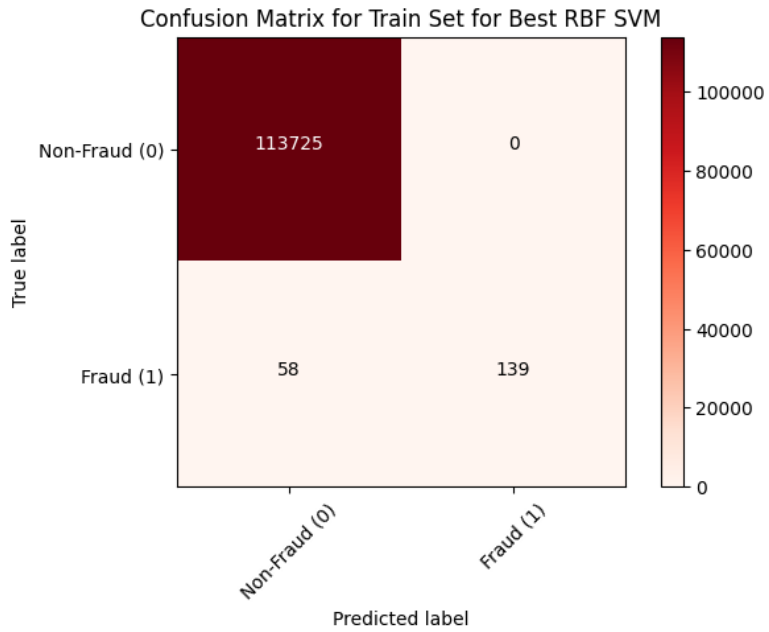
```
[INFO] [2025-01-04T20:51:43.590Z] Non-Fraud transactions in the training set: 113725 samples, 99.8271%
[INFO] [2025-01-04T20:51:43.591Z] Fraud transactions in the training set: 197 samples, 0.1729%
[INFO] [2025-01-04T20:51:43.591Z]
[INFO] [2025-01-04T20:51:43.592Z] Non-Fraud transactions in the test set: 28432 samples, 99.8280%
[INFO] [2025-01-04T20:51:43.592Z] Fraud transactions in the test set: 49 samples, 0.1720%
[INFO] [2025-01-04T20:51:43.593Z]
```

	Model	Data	Accuracy	Precision	Recall	F1 Score
0	Best Linear SVM	Train	99.87%	0.740	0.376	0.498
1	Best Linear SVM	Test	99.87%	0.714	0.408	0.519
2	Best Polynomial SVM	Train	99.83%	0.000	0.000	0.000
3	Best Polynomial SVM	Test	99.83%	0.000	0.000	0.000
4	Best RBF SVM	Train	99.95%	1.000	0.706	0.827
5	Best RBF SVM	Test	99.83%	1.000	0.020	0.040

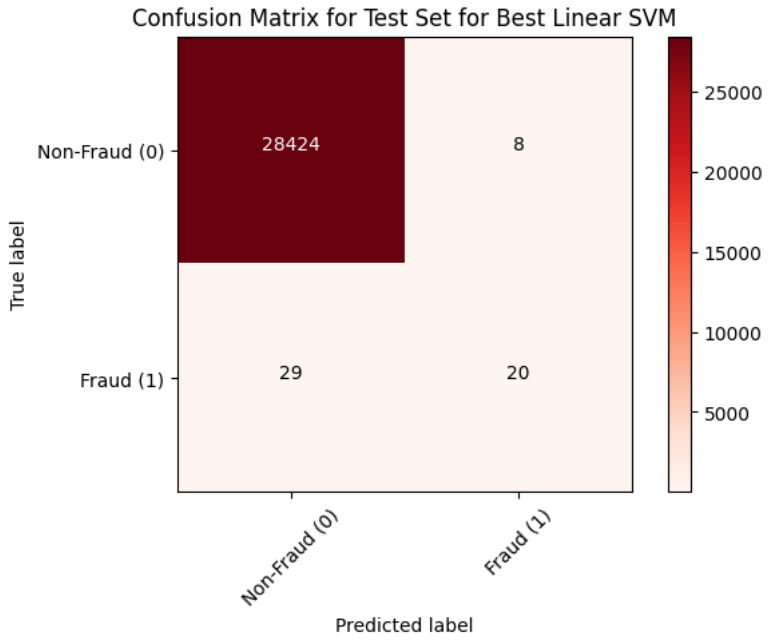
Experiments and Results - SVM - Grid Search w. 50% of Data

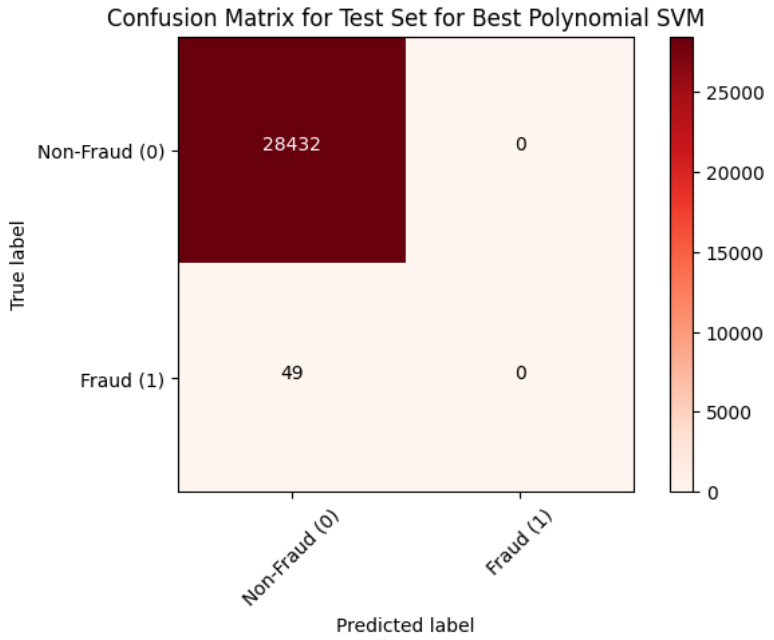


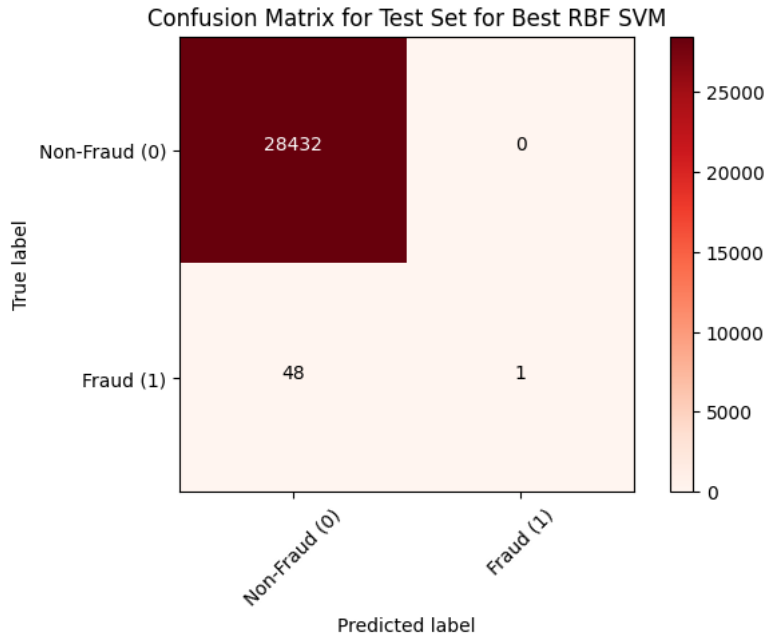




Experiments and Results - SVM - Grid Search w. 50% of Data

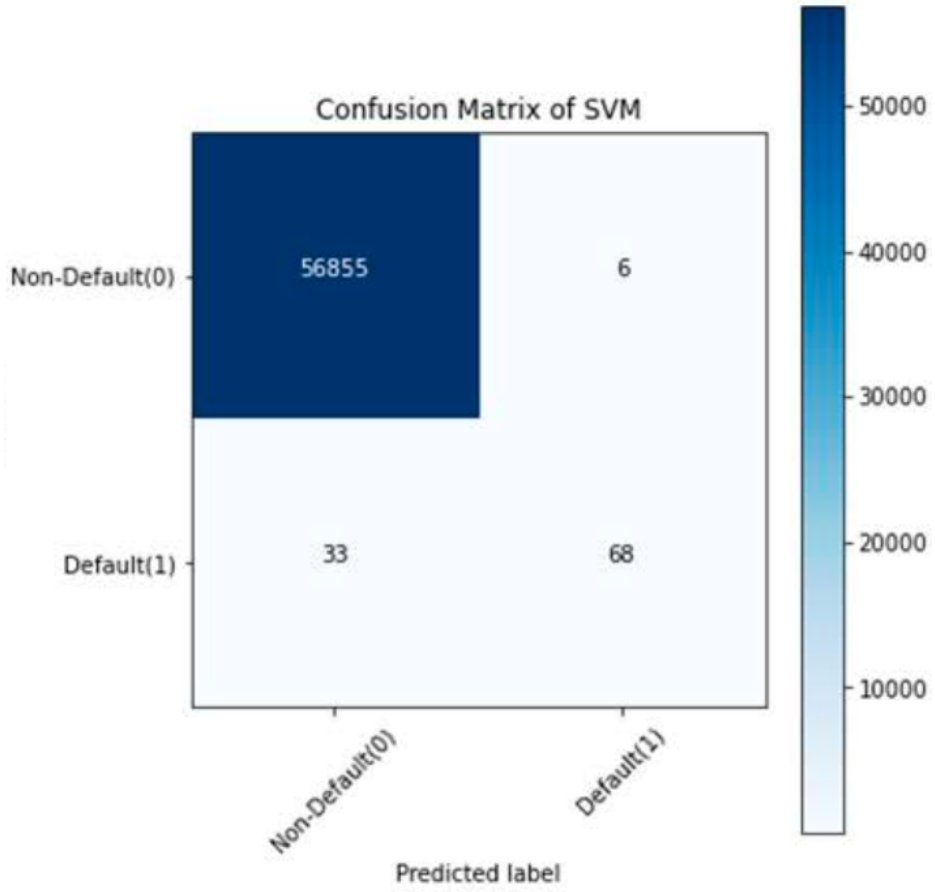




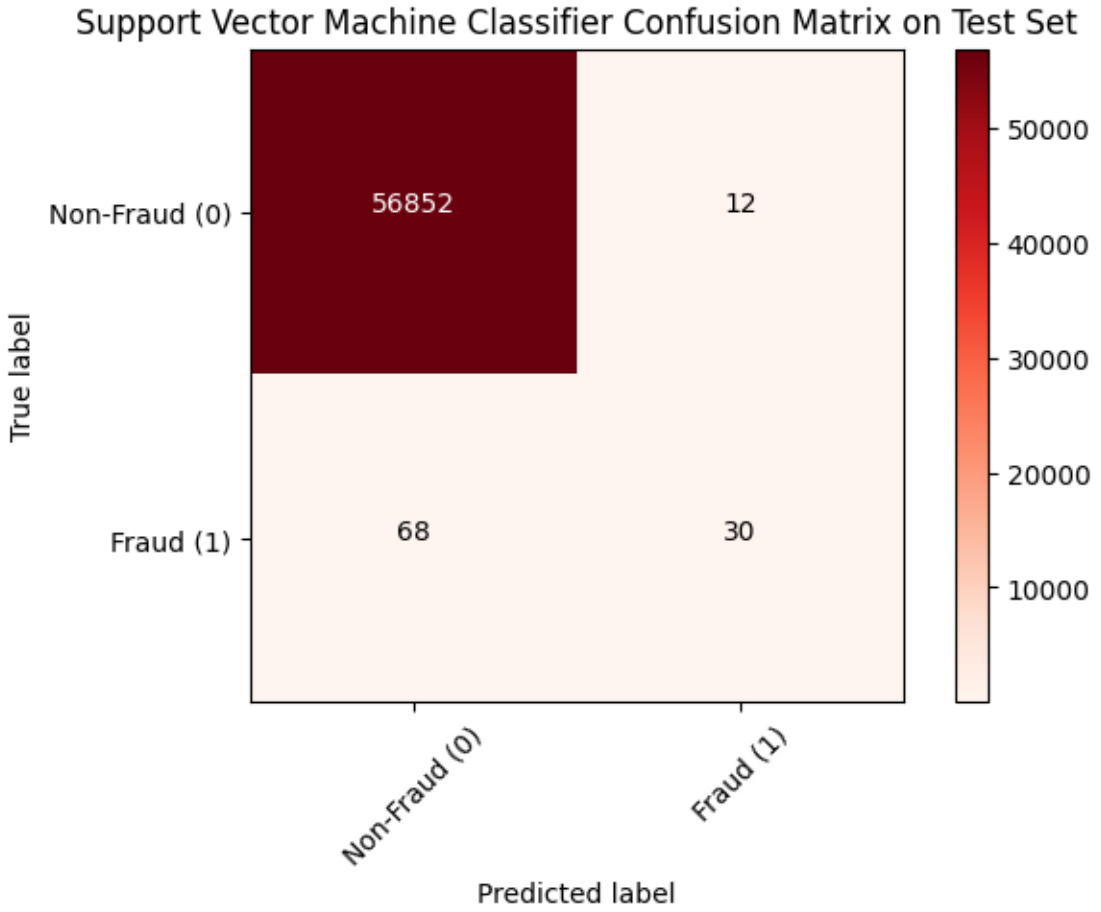


Experiments and Results - SVM

Result on Paper

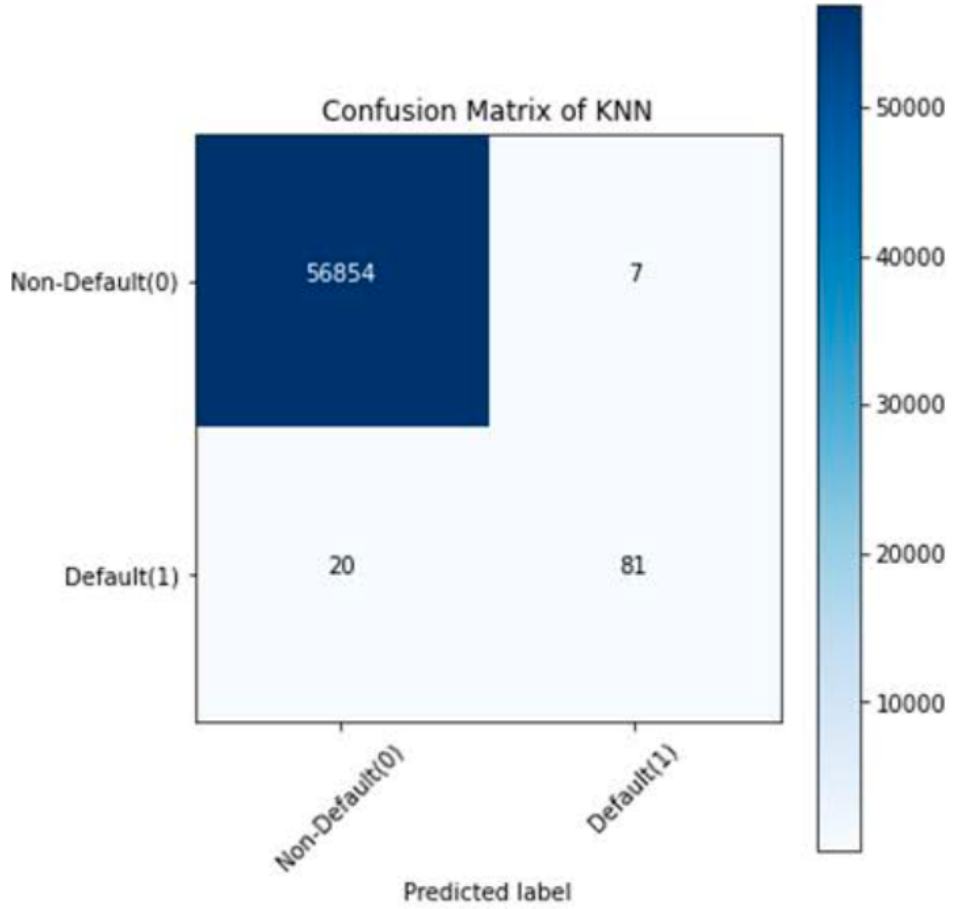


Our Result

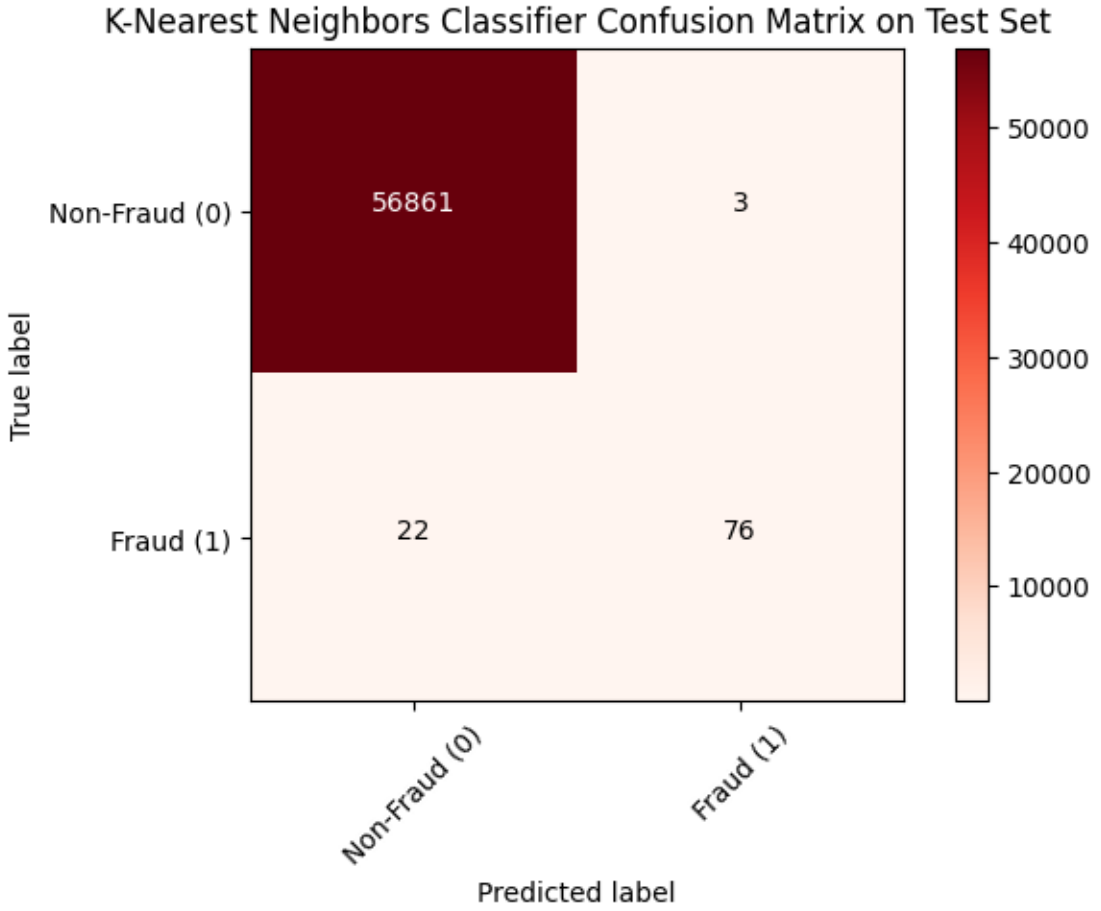


Experiments and Results - KNN

Result on Paper

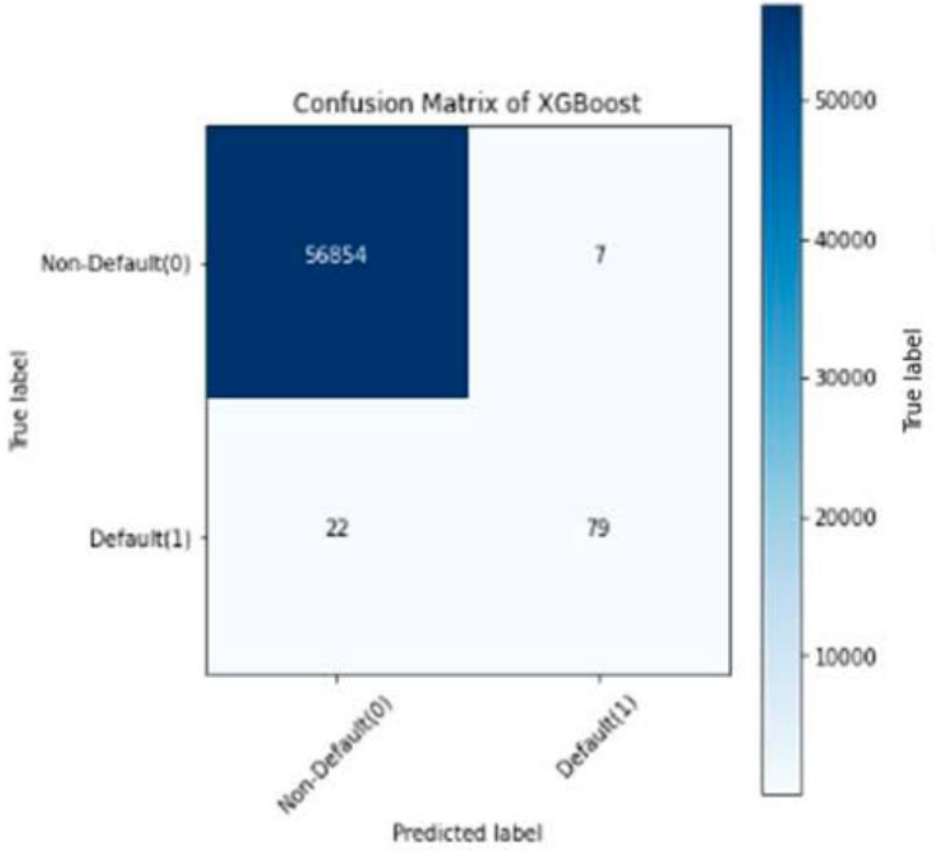


Our Result

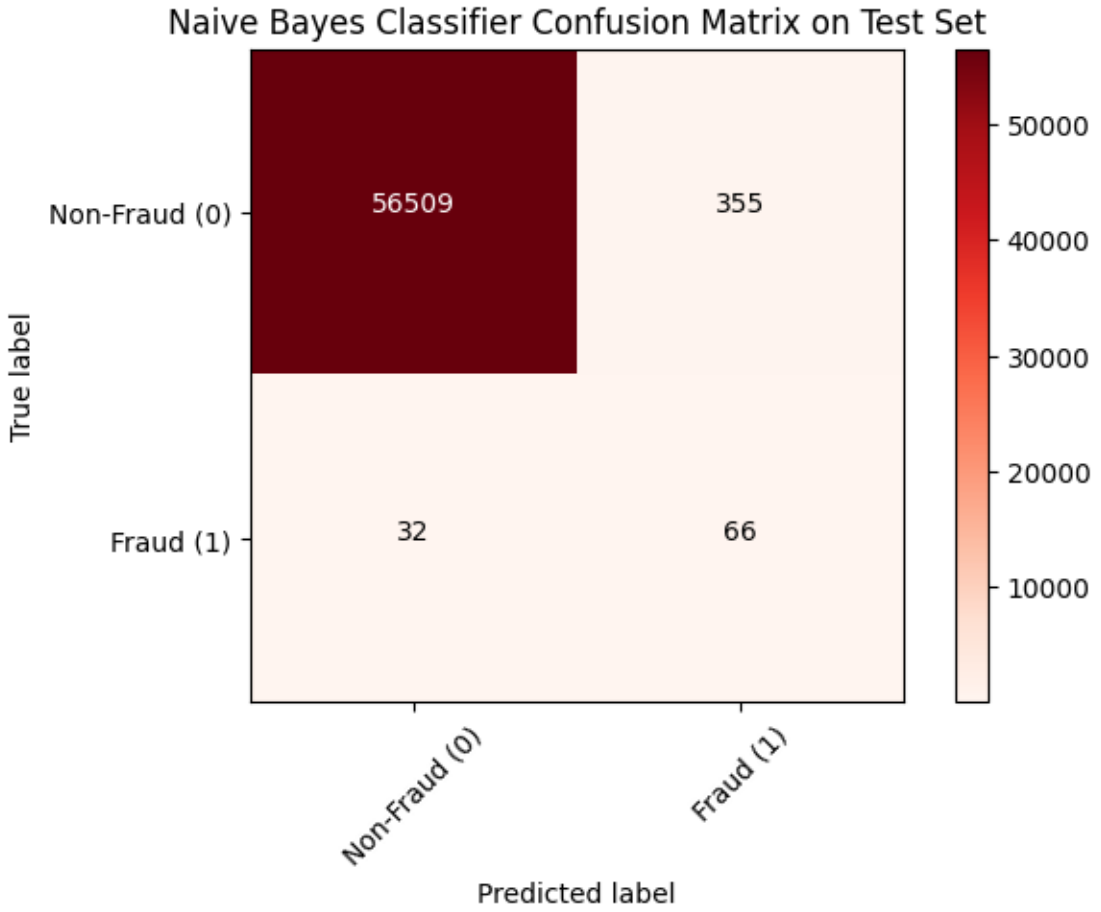


Experiments and Results - XGBoost vs Naive Bayes

Result on Paper



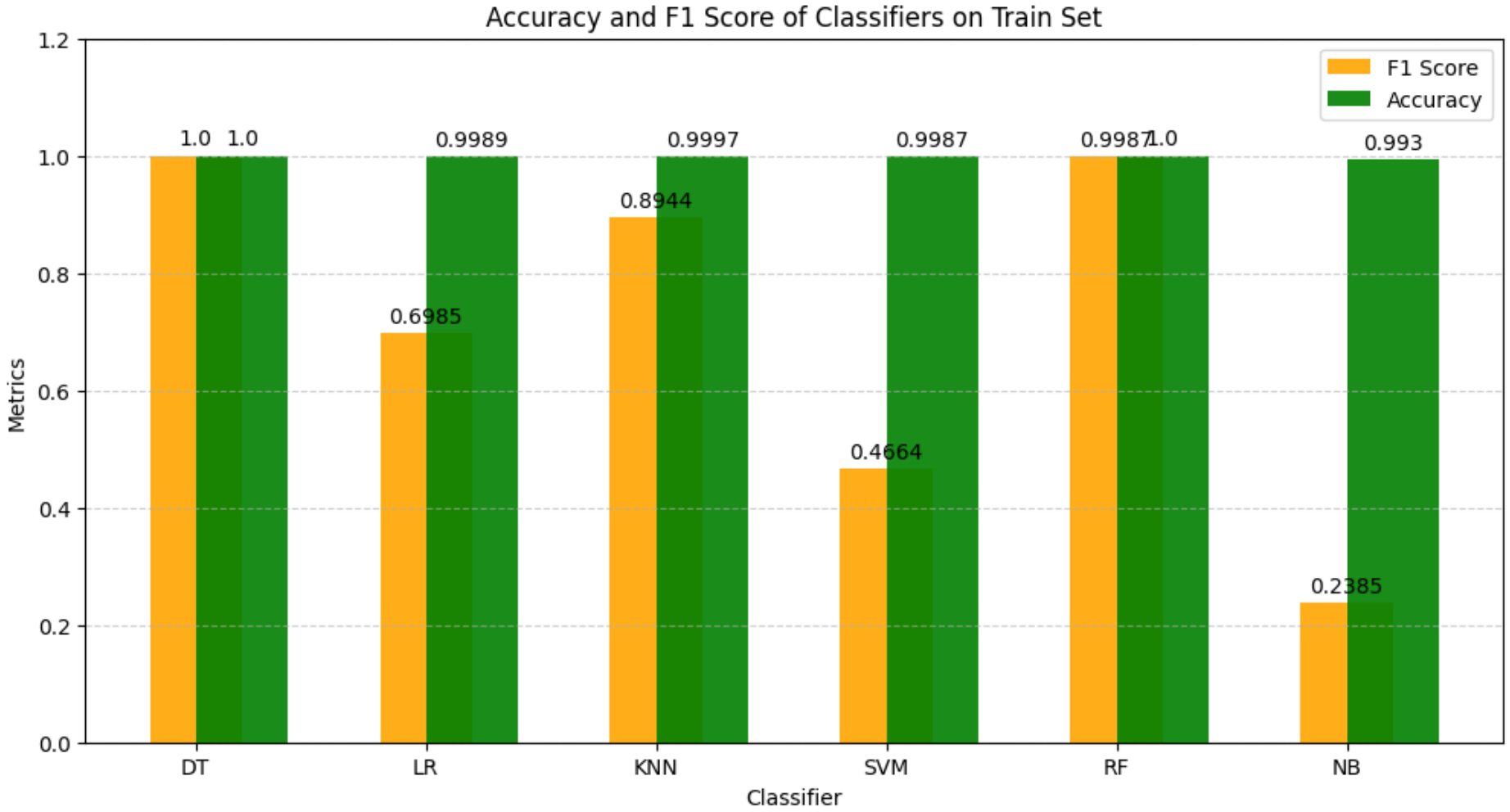
Our Result



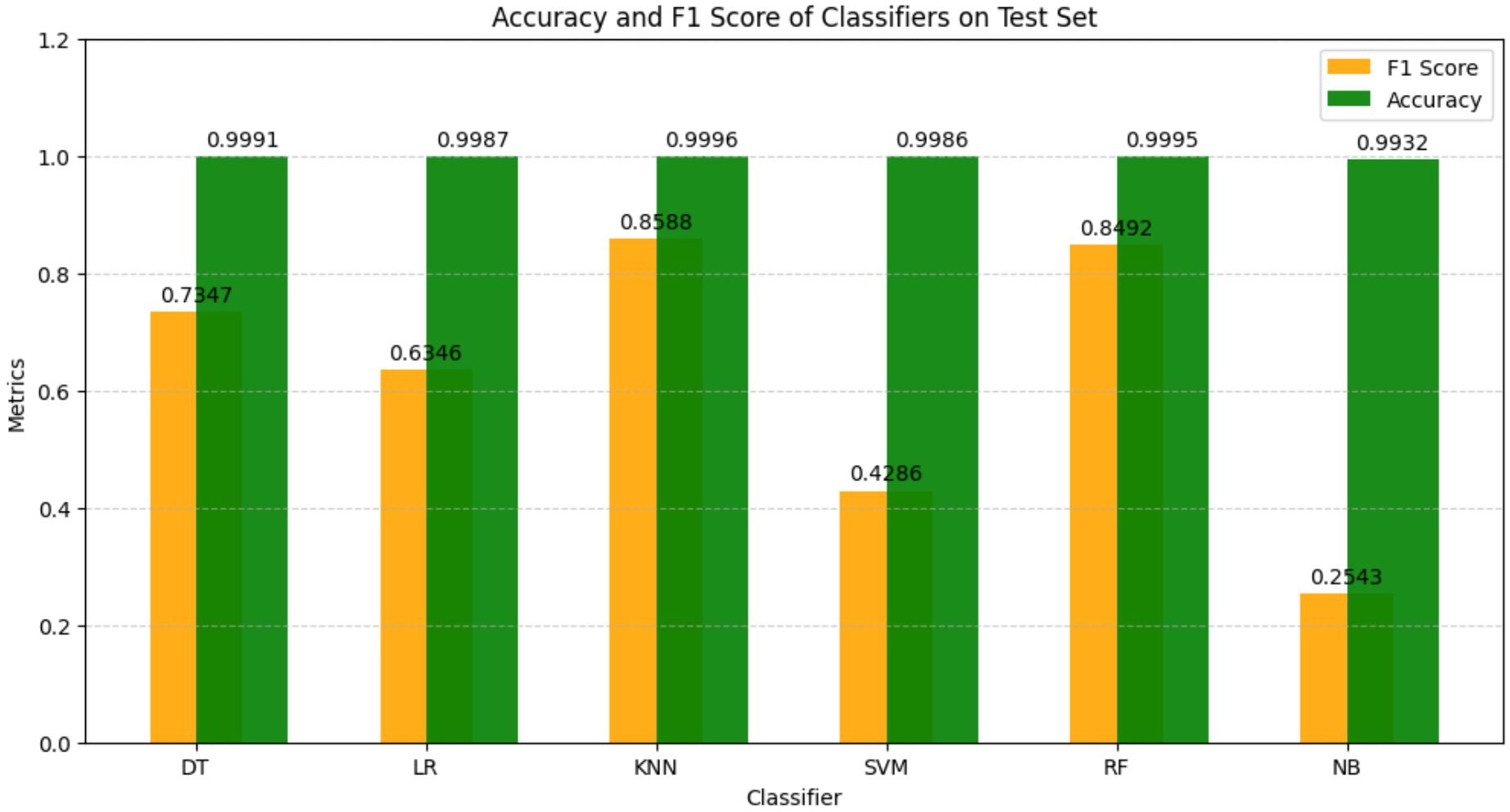
Experiments and Results

	Model	Data	Accuracy	Precision	Recall	F1 Score	Support
0	DT	Train	1.0000	1.0000	1.0000	1.0000	394
0	DT	Test	0.9991	0.7347	0.7347	0.7347	98
0	LR	Train	0.9989	0.6915	0.7056	0.6985	394
0	LR	Test	0.9987	0.6000	0.6735	0.6346	98
0	KNN	Train	0.9997	0.9731	0.8274	0.8944	394
0	KNN	Test	0.9996	0.9620	0.7755	0.8588	98
0	SVM	Train	0.9987	0.7674	0.3350	0.4664	394
0	SVM	Test	0.9986	0.7143	0.3061	0.4286	98
0	RF	Train	1.0000	1.0000	0.9975	0.9987	394
0	RF	Test	0.9995	0.9383	0.7755	0.8492	98
0	NB	Train	0.9930	0.1470	0.6320	0.2385	394
0	NB	Test	0.9932	0.1568	0.6735	0.2543	98

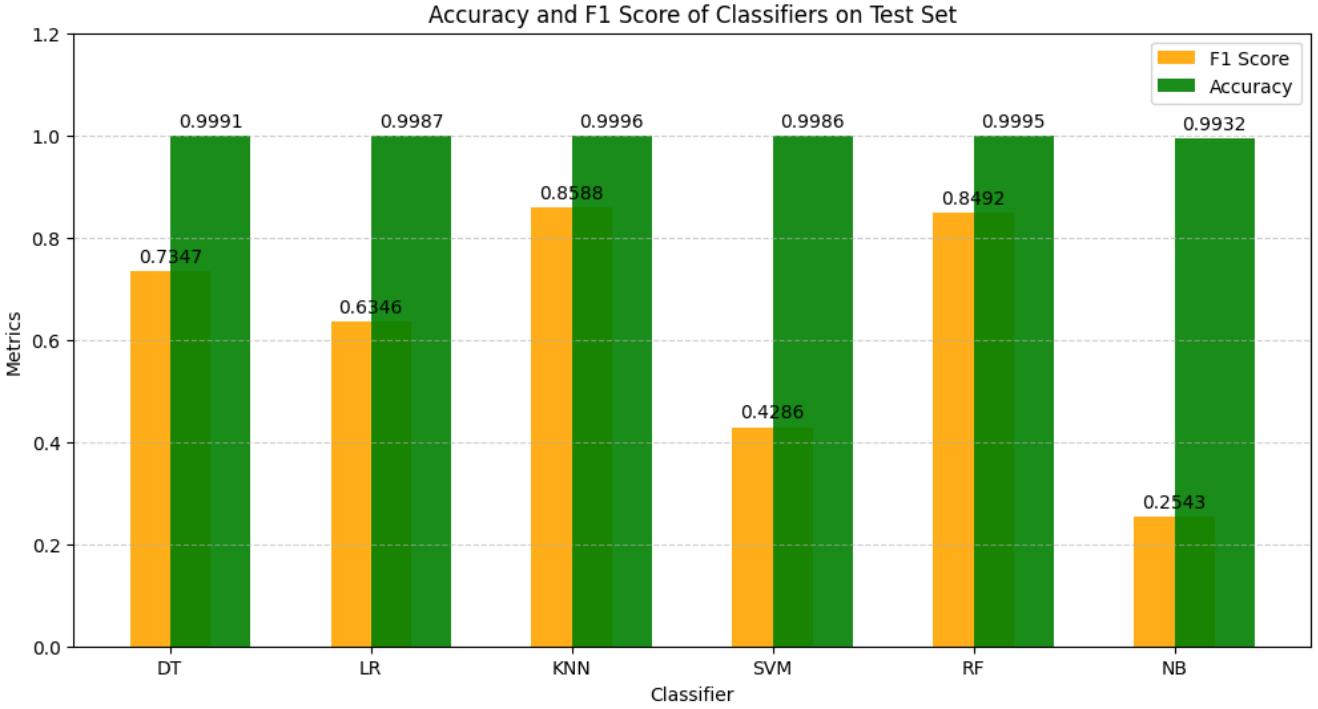
Experiments and Results

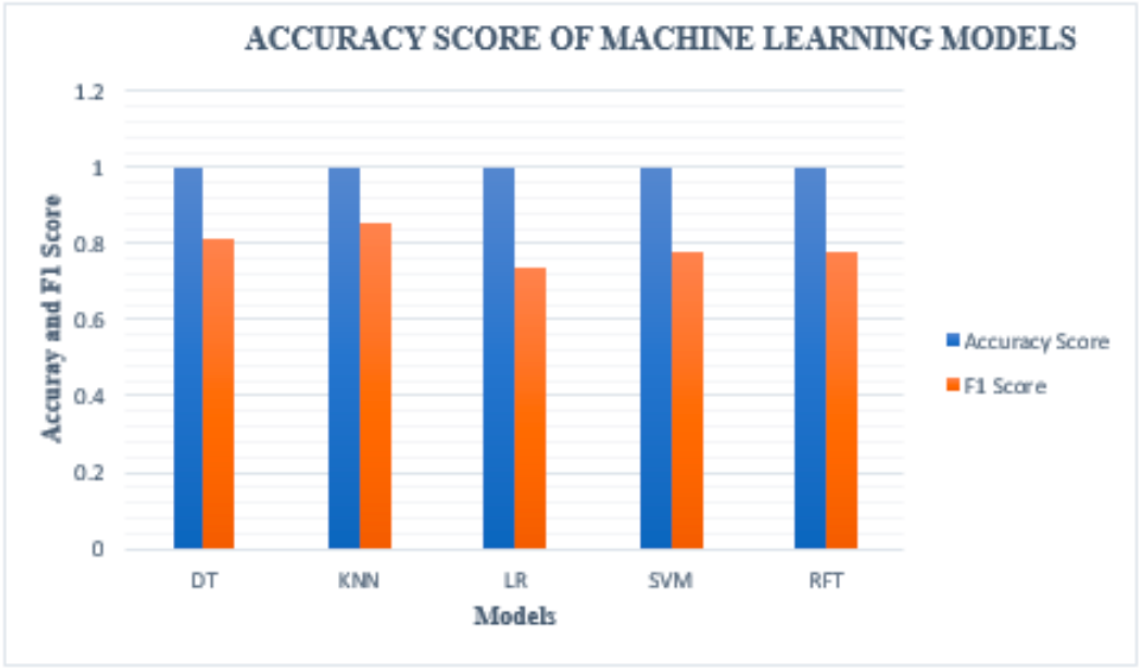


Experiments and Results



Experiments and Results





Agenda

I. Literature Review
II. Dataset
III. Experiments and Results
IV. Conclusions
V. References
VI. Q & A

Conclusions

- Başarı Sırası: KNN > RF > DT > LR > SVM > NB
- Her makine öğrenmesi yöntemi her problem için uygun değildir.
- Veri seti ön işleme modelin çalışabilmesi / performansı için önemlidir

Future Works

- LightGBM (Light Gradient Boosting Machine) ile çalışma yapılabilir.
- CatBoost ile çalışma yapılabilir.

Conclusions

Future Works

- Derin öğrenme yöntemleri üzerinden başarım ölçümü yapılabilir.
- Imbalanced veri setini balanced bir veri seti haline getirerek sınıflandırma performansları ölçülebilir.
  - Random Oversampling
  - Random Undersampling
  - SMOTE (Synthetic Minority Over-sampling Technique)

Agenda

I. Literature Review

II. Dataset

III. Experiments and Results

IV. Conclusions

V. References

VI. Q & A

References

[1] Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms

[2] Machine Learning Group - ULB - Credit Card Fraud Detection Dataset

[3] Credit Card Fraud Detection Using Lightgbm Model

[4] CatBoost for Fraud Detection in Financial Transactions

[5] A Data Mining Based Fraud Detection Hybrid Algorithm in E-bank

[6] Identifying Fraudulent Credit Card Transactions Using Ensemble Learning

Agenda

I. Literature Review

II. Dataset

III. Experiments and Results

IV. Conclusions

V. References

VI. Q & A

Q & A

THANK YOU FOR LISTENING…