أكاديمية سدايا
SDAIA Academy

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

T5 Data Science Bootcamp

Credit card fraud detection

Final Project Writeup

Name: Amnah Nasser Aldayri

Abstract
As the world is rapidly moving towards digitization and money transactions are becoming cashless, the use of credit cards has rapidly increased. The fraud activities associated with it have also been increasing which leads to a huge loss to the financial institutions. Therefore, we need to analyze and detect the fraudulent transaction from the non-fraudulent ones. Various fraud scenarios happen continuously, which has a massive impact on financial losses. Many technologies such as phishing or virus-like Trojans are mostly used to collect sensitive information about credit cards and their owner details. Therefore, efficient technology should be there for identifying the different types of fraudulent conduct in credit cards. In this project, various machine learning and deep learning approaches are used for detecting frauds in credit cards and different algorithms such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes, Random Forest, XG BOOST, and Neural Network are applied. The comparative analysis visualized that the Neural Network algorithm generates better results than other approaches with 98.3% accuracy.

Design

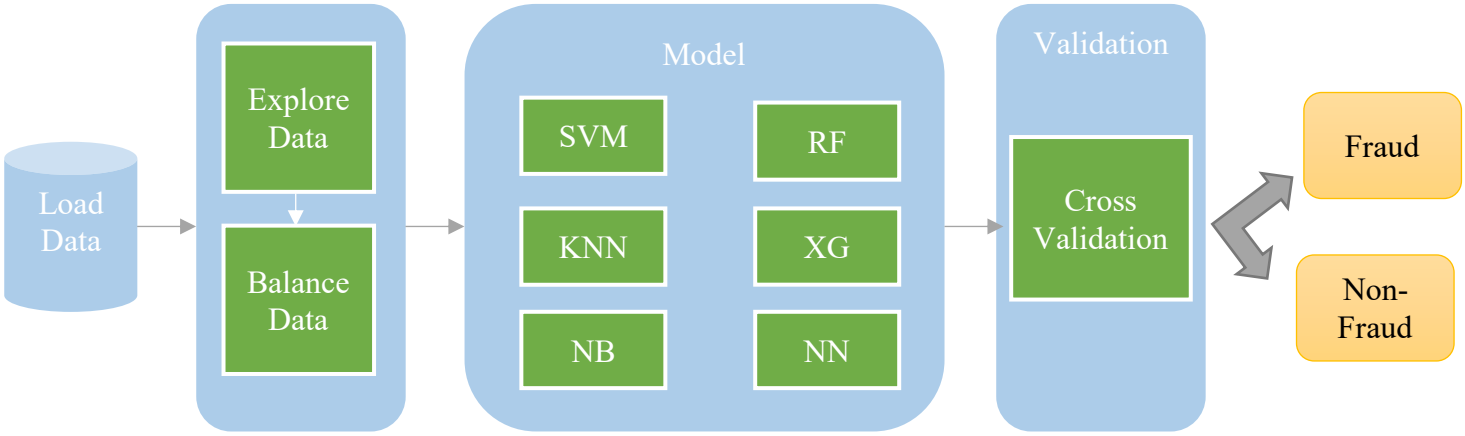

Figure 1. System Architecture

RF: Random Forest **XG**: XG BOOST **NN**: Neural Network **NB**: Naive Bayes

Data

The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. Moreover, it contains 31 features, 29 features are dismal, and 2 features are integer. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.It contains only numerical input variables which are the result of a PCA transformation. Because of privacy issues, the dataset developer didn't provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, to be used to train the model.

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	0.090794	-0.551600	-0.617801	-0.991390	-0.311169	1.468177	-0.470401	0
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	-0.166974	1.612727	1.065235	0.489095	-0.143772	0.635558	0.463917	-0
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	0.207643	0.624501	0.066084	0.717293	-0.165946	2.345865	-2.890083	1
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	-0.054952	-0.226487	0.178228	0.507757	-0.287924	-0.631418	-1.059647	-0
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	0.753074	-0.822843	0.538196	1.345852	-1.119670	0.175121	-0.451449	-0

Figure 2. Dataset

Algorithms
In this project different machine learning and deep learning algorithms ware implemented and then compared to determine the best model performance in term of (Confusion metric, Precision, Recall, Accuracy).

Models
- SVM
- KNN
- Naive Bayes
- Random Forest
- XG BOOST
- Neural Network

Model Evaluation and Selection
- Confusion Metrices
- Precision
- Recall
- Accuracy

Algorithms	Accuracy
SVM	0.96%
KNN	0.92%
Naive Bayes	0.92%
Random Forest	0.96%
XG BOOST	94.95%
Neural Network	98.3%

Tools
- NumPy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting

Communication

Fraud is a major problem for the whole credit card industry that grows bigger with the increasing popularity of electronic money transfers. To effectively prevent the criminal actions that lead to the leakage of bank account information leak, skimming, counterfeit credit cards, the theft of billions of dollars annually, and the loss of reputation and customer loyalty, credit card issuers should consider the implementation of advanced Credit Card Fraud Prevention and Fraud Detection methods. Machine Learning-based methods and deep learning-base methods can continuously improve the accuracy of fraud prevention based on information about each cardholder's behavior.

Through this project and after comparing a group of algorithms, we can say that neural networks gave a better result in detecting fraud and non-fraud operations by a greater percentage than the rest of the algorithms. The following figure Figure 3, Figure 4 shows the accuracy and loss rate of the neural networks.

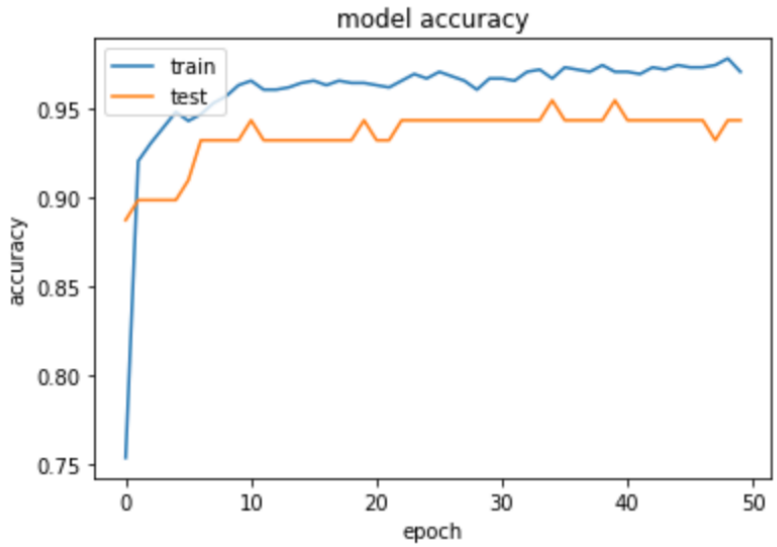

Figure 4. Model Accuracy

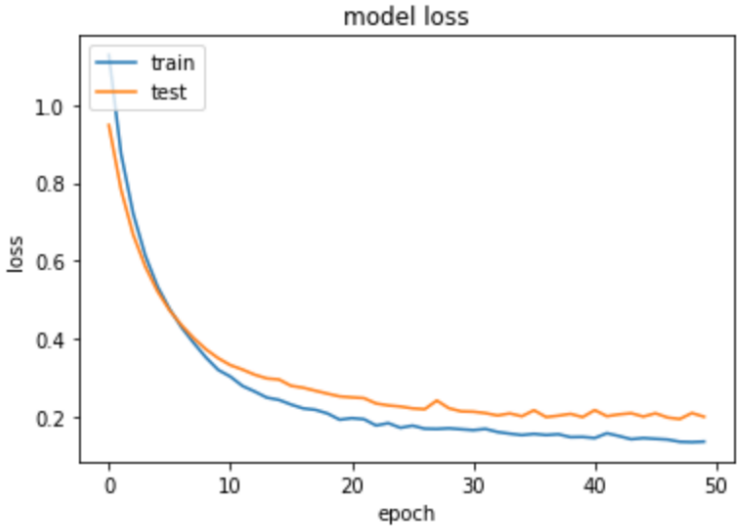

Figure 5. Model Loss