Credit Card Fraud Detection using Data Science

Predictive Modeling Using DNN

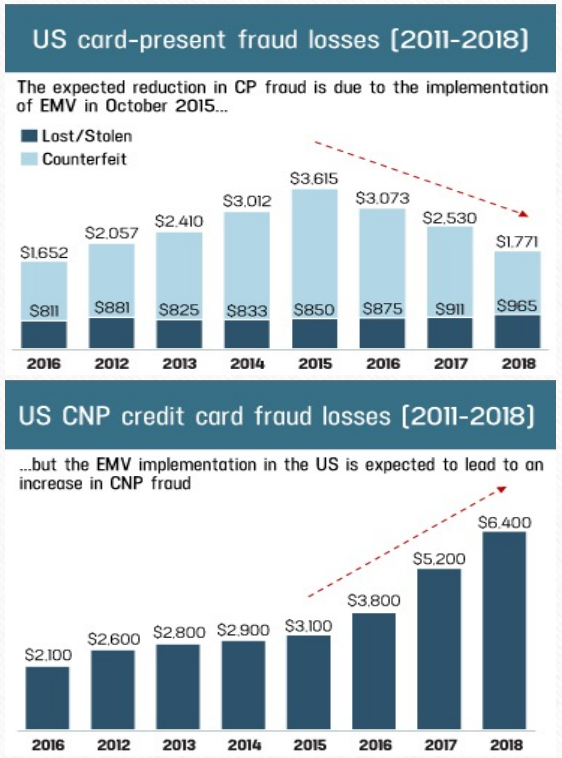
- *Chandramouli Yalamanchili*

Abstract

- All transactions are getting digitized now a days and the usage of credit card transactions has been skyrocketing last few years.
- The industry is coming with a lot of innovations like EMV chip cards, token wallets like apple pay, android pay etc. to prevent fraudsters from capturing the card information and re-using it.
- But the fraudsters have been successful to extract the card/cardholder information one way or the other and perform fraudulent transactions using this information.
- Goal for this project is to build a DNN (Deep Neural Network) to see how efficiently the model can predict the fraudulent transactions.

Domain Background - What are credit card transactions?

- Credit card transactions are electronic transactions that happen when a credit card is used for making a purchase.
- On a very high level, credit card transactions can be of two types - card present, and card not present transactions.
- **Card present transactions** are the ones where the physical card and/or cardholder is present at the merchant location.
- Fraud has reduced in recent years for card present transactions due to enhancements like EMV Chip cards and apple pay, android pay token/digital wallets.
- **Card Not Present (CNP) transactions** are the ones where cardholder and/or card is not present at merchant location, these are usually e-commerce transactions.
- Card not present transactions is where we are seeing most of the fraud transactions in the recent years.



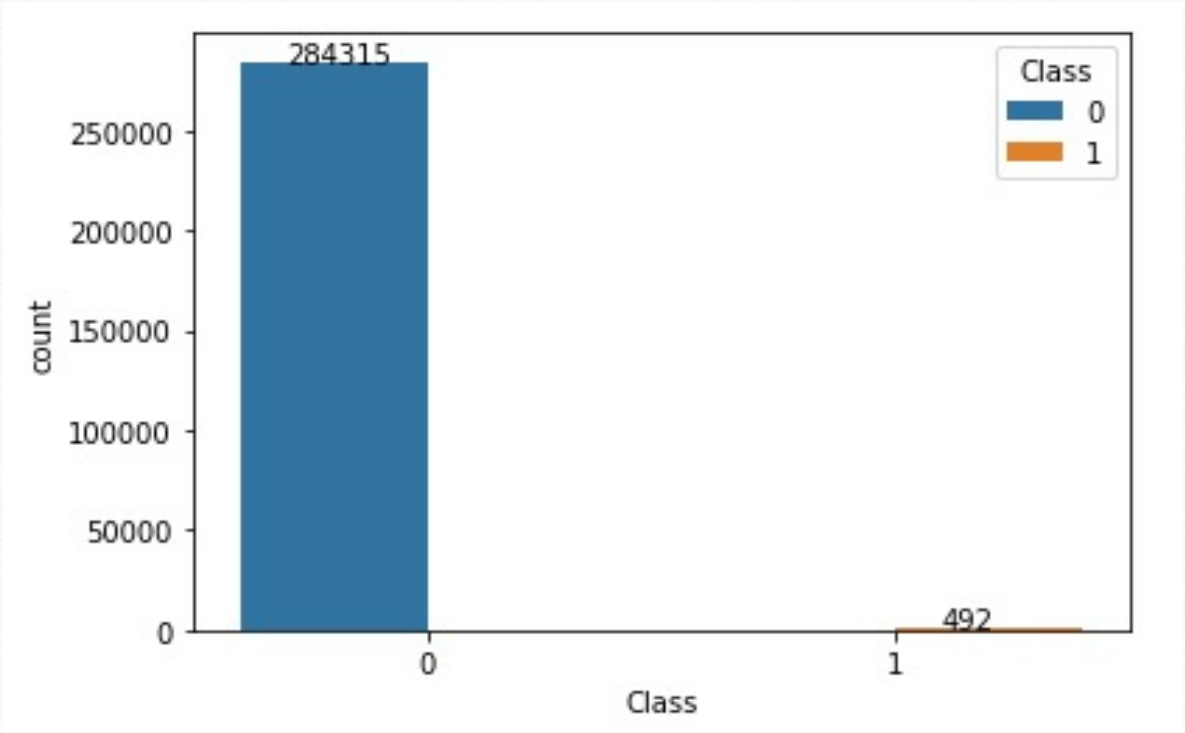
Domain Background - Solutions available today

Solution Type	Drawbacks
Rules Based	❖ Complex to build and manages the rules. ❖ Clients/banks could set up bad rules resulting in more false positives.
Score Based	❖ Static models are used for predicting fraud, it would take at least 3 months to produce new scoring models to meet current market needs.
All other methods	❖ Highly dependent on human support. ❖ Some applications detect fraud after the transaction got approved, these would stop future fraud transactions, but initial transactions are still impacted.

Exploratory Data Analysis - Input dataset

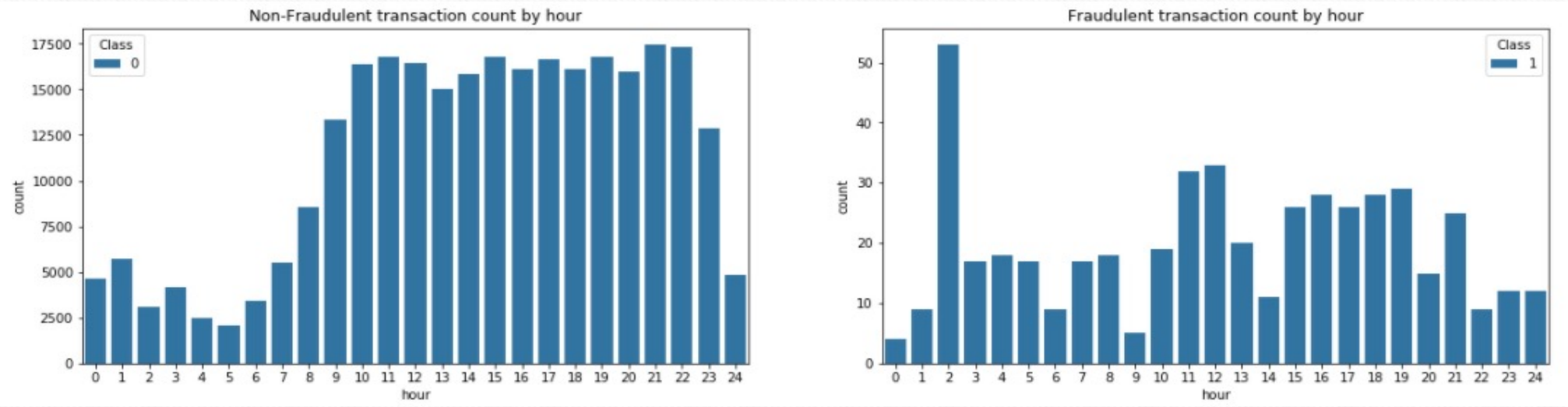
- Dataset - https://www.kaggle.com/mlg-ulb/creditcardfraud.
  - 284,807 transactions, with 492 fraudulent transactions. (0.172%).
- Features:
  - Total 31 features.
  - All fields are principal components obtained with PCA transformation.
  - Class, Time, and Amount features are not transformed.
  - Class is our target variable with value of 1 for fraudulent transactions and 0 for non-fraudulent transactions.

Exploratory Data Analysis – Transaction distribution by Class



The concentration of non-fraud transactions in the dataset is very low (0.172% of all transactions).

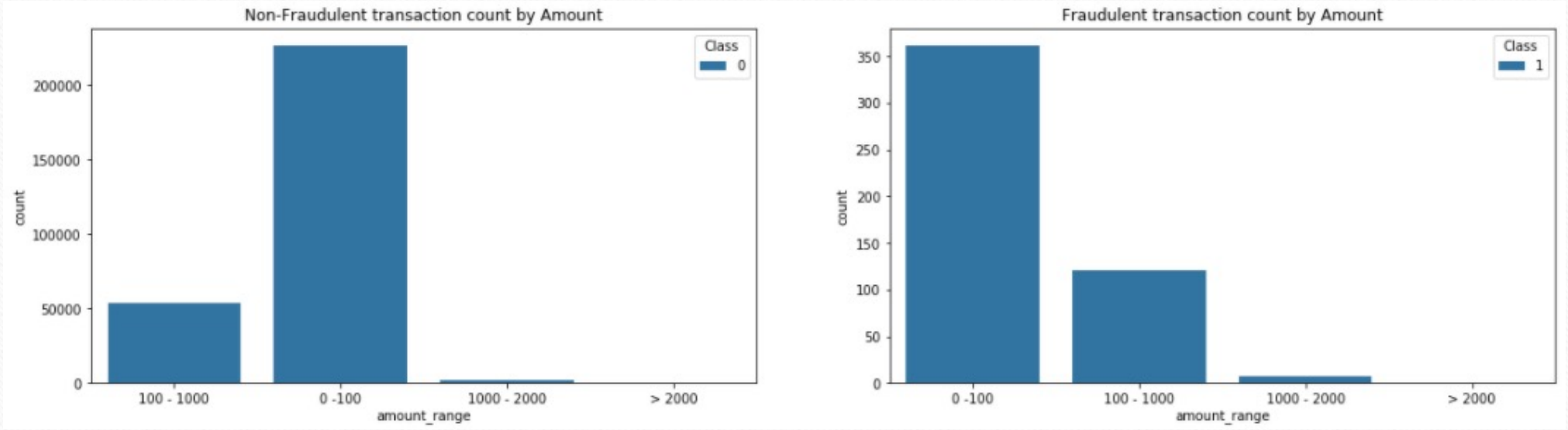
Exploratory Data Analysis – Transaction distribution by Time



As expected, we can see there is a high number of transactions happening during the day than night.

We can also see that the fraudulent transactions are higher at night.

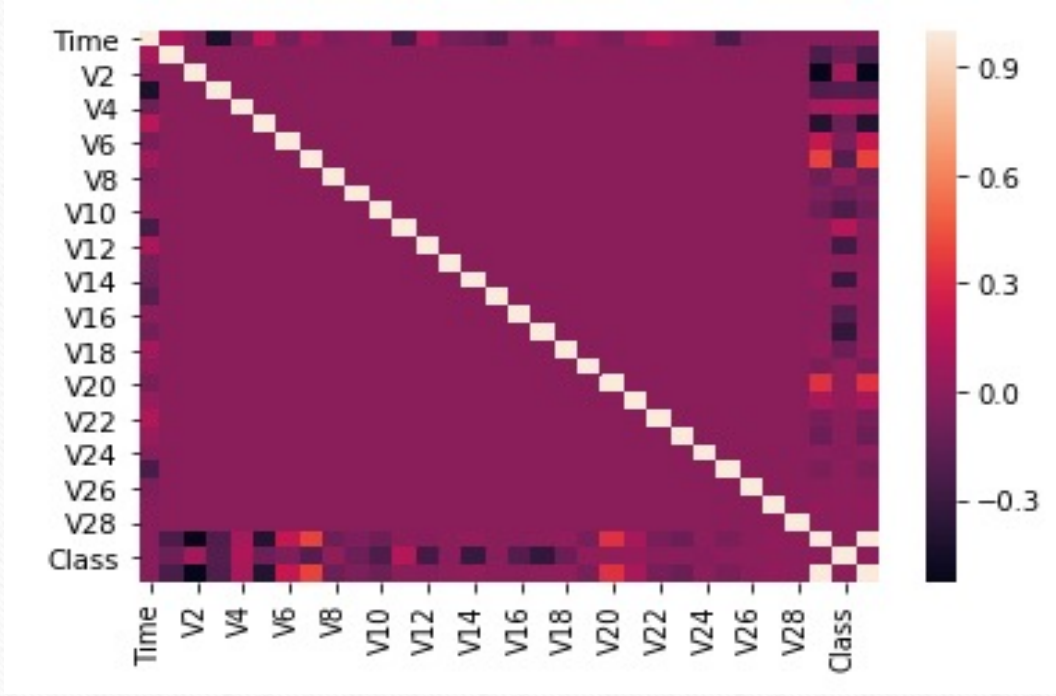
Exploratory Data Analysis – Transaction distribution by Amount



As expected, we can see there is a high number of transactions between 0-100 pounds.

We can also see that the fraudulent transactions are higher at lower denominations as well.

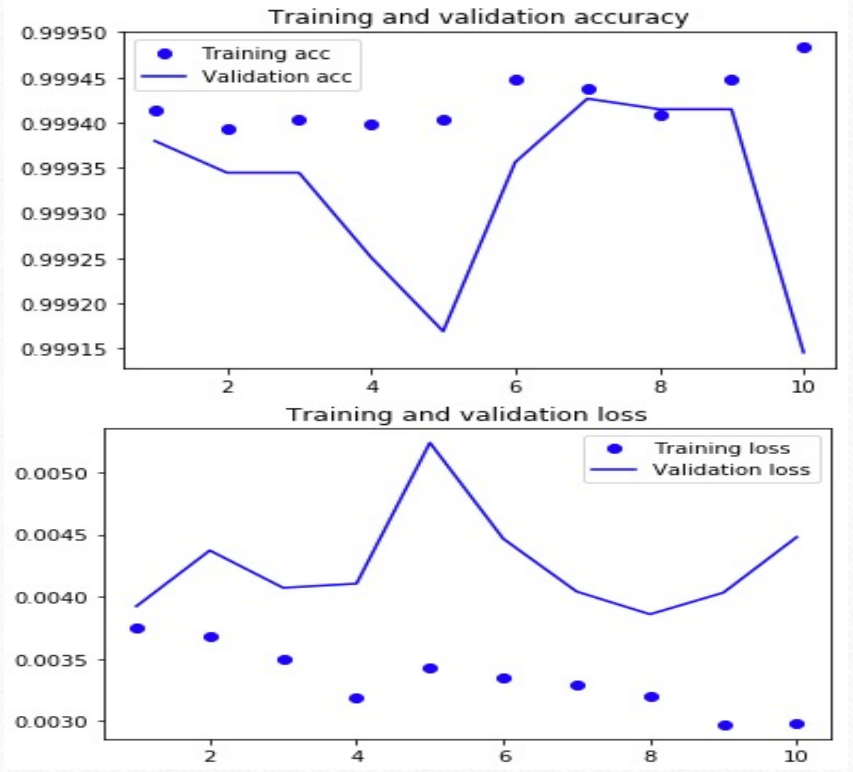
Exploratory Data Analysis - Correlation Between Features



As we can see only a very few features have considerable correlation with the class feature.

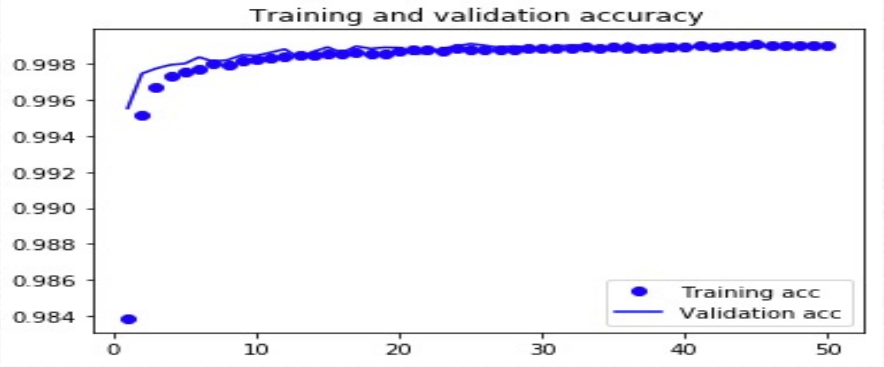
Modeling - DNN without sampling

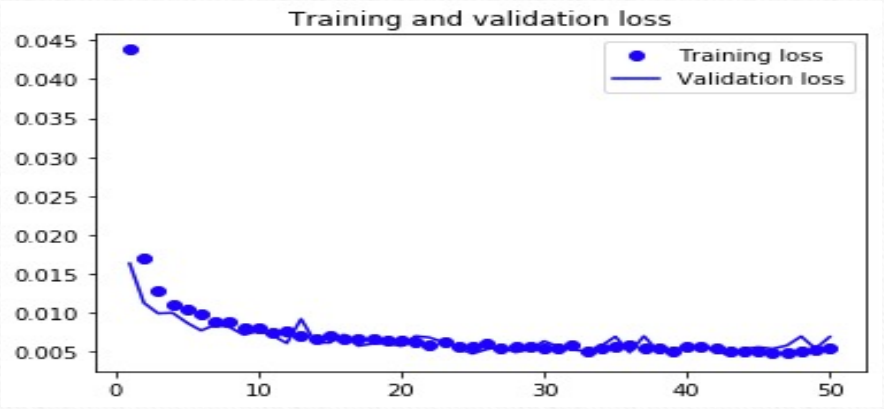
- Used StandardScaler (sklearn package) to standardize the time and amount features.
- Keras Sequential DNN model with 6 layers, with 4 hidden layers.
- In the first experiment, I have used the input data as it is available, where the class 1 (Fraudulent) samples are very less compared to class 0 (Genuine) samples.
- I have fitted with 10 epochs and I could clearly see that the model is overfitted and the validation loss has peaked at 10th EPOC.



Modeling - DNN with data sampling

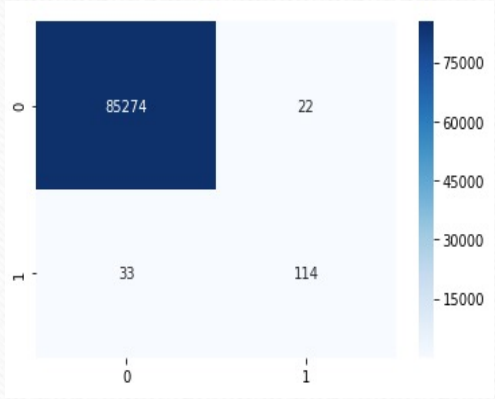
- For the second experiment I have used the same model definition as first experiment as well the same base data.
- But this time, I have used SMOTE to oversample the minority class in the dataset to create a balanced dataset for the model training.
- I have fitted with 50 epochs and I could clearly see that the model got stabilized at around 20 epochs keeping the validation loss to minimum.

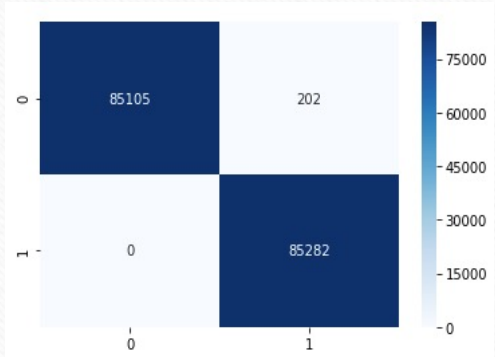




Modeling – DNN Model Comparison

Model	Loss	Accuracy Score	Precision Score	Recall Score	F1 Score
DNN Model with imbalanced sample data	0.45%	99.94%	83.82%	77.55%	80.57%
DNN Model with over-sampling technique	0.69%	99.88%	99.76%	100%	99.88%





Future Scope

- As next steps, I would like to review different options available from references to see what can be done to integrate this DNN model into the transaction processing. I want to evaluate if there are any low-latency integration options that can detect the fraudulent transactions in real-time before the transaction decision is sent back to the merchant POS (Point Of Sale) device.
- In addition to integration into the transaction execution flow, I would also like to research in the future to see how I can get the training automated into the modeling process to ensure model is constantly learning and maintaining the high fraud predicting accuracy and precision.

Conclusion

- Credit card fraud is a growing concern in today's world with both credit card usage and credit card fraud on the rise. As shown in the stats, the number of data breaches, identity theft cases, and credit card fraudulent transactions are rising at an alarming level.
- Data science plays a significant role in improving the fraud prediction tools we are currently using, by analyzing credit card transactions and being able to predict fraud based on transactional data, cardholder data, and historical data.
- Machine learning algorithms running on neural networks can be effective as proven in this project and they should replace the current static rules-based or statis profiles/score-based fraud detection products to improve the fraud prediction precision and to stay up to speed with market trends in detecting new strategies of fraudsters and stopping them.
- One common issue we will have with credit card transactions is that very few transactions would be fraudulent when compared to the overall transaction volume. So, we must use the relevant oversampling techniques to be able to train the models efficiently to do a better job at catching fraudulent transactions.

References

1. Dataset - https://www.kaggle.com/mlg-ulb/creditcardfraud
2. Code - https://github.com/chandu85/data-science/blob/main/Project%201%20-%20Creditcard%20Fraud%20detection/Code/CreditCard%20Fraud%20Detection.ipynb
3. https://www.washingtonpost.com/news/the-switch/wp/2018/03/01/equifax-keeps-finding-millions-more-people-who-were-affected-by-its-massive-data-breach/?noredirect=on&utm_term=.0b854d8c3526
4. Chan, P. K., Fan, W., Prodromidis, A. L., & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. IEEE Intelligent systems, (6), 67- 74.
5. Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. In Proceedings 11th International Conference on Tools with Artificial Intelligence(pp. 103-106). IEEE.
6. Ghosh, S., & Reilly, D. L. (1994, January). Credit card fraud detection with a neural-network. In System Sciences, 1994. Proceedings of the Twenty- Seventh Hawaii International Conference on (Vol. 3, pp. 621-630). IEEE.
7. Chan, P. K., & Stolfo, S. J. (1998, August). Toward Scalable Learning with Non-Uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection. In KDD (Vol. 98, pp. 164-168).
8. Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002, January). Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st international naiso congress on neuro fuzzy technologies (pp. 261-270).
9. Masoumeh Zareapoor, & Pourya Shamsolmoali (2015). Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier. Procedia Computer Science (Vol. 48, pp. 679-685).
10. Stolfo, S., Fan, D. W., Lee, W., Prodromidis, A., & Chan, P. (1997, July). Credit card fraud detection using meta-learning: Issues and initial results. In AAAI-97 Workshop on Fraud Detection and Risk Management.
11. Luke Sun (July 2020). Credit Card Fraud Detection.
12. Dorronsoro, J. R., Ginel, F., Sgnchez, C., & Cruz, C. S. (1997). Neural fraud detection in credit card operations. IEEE transactions on neural networks, 8(4), 827-834.
13. Sánchez, D., Vila, M. A., Cerda, L., & Serrano, J. M. (2009). Association rules applied to credit card fraud detection. Expert systems with applications, 36(2), 3630-3640.