Credit Card Transaction Classification

Group 27

Yuyang Zhou
Department of Statistical and Actuarial Sciences
Western University
London, Canada
jzhou494@uwo.ca

Hepple Xi
Department of Computer Science
Western University
London, Canada
hxi26@uwo.ca

Sheng Zhang
Department of Computer Science
Western University
London, Canada
szha428@uwo.ca

Cynthia Wen
Department of Statistical and Actuarial Sciences
Western University
London, Canada
cwen32@uwo.ca

***Abstract*—Credit card fraud has caused significant financial losses for individuals and institutions worldwide, undermining trust in financial systems and highlighting the need for effective preventive measures [1]. Detecting such fraud poses a significant challenge primarily due to the highly imbalanced nature of transaction datasets, where fraudulent cases are rare. This project addresses fraud detection using machine learning approaches. To address this highly imbalanced dataset, we will use some pre-processing techniques like feature selections and resampling. Feature selection was conducted using correlation analysis and ANOVA tests to identify variables most relevant to fraudulent activity. Advanced classification algorithms—including Support Vector Machines, XGBoost, Random Forests, and Neural Networks—were trained on resampled datasets using undersampling and oversampling methods [2]. The performance of these models was evaluated using appropriate metrics to assess their effectiveness in handling this problem, offering valuable insights into fraud detection.**

***Index Terms*—credit card fraud, ANOVA, SVM, XGBoost, Random Forests, Neural Networks**

I. Background

Credit card fraud is a growing issue due to the rise of digital payments and sophisticated fraud techniques. Fraudsters exploit technologies like contactless payments and IoT devices, causing significant financial losses [3]. Effective fraud detection systems are crucial as they leverage advanced technologies to identify and prevent fraudulent transactions in real time, thereby protecting consumers and financial institutions and maintaining trust in digital payment systems. Credit card fraud detection has evolved significantly with the integration of advanced technologies. The primary technical directions include machine learning and deep learning algorithms, which are essential for identifying complex patterns in transaction data. Techniques such as supervised learning (e.g., Random Forest [4], Logistic Regression [5]) and unsupervised learning (e.g., Isolation Forest [6], Local Outlier Factor) are widely used. Data imbalance significantly impacts credit card fraud detection by skewing the dataset towards legitimate transactions, often leading to high false negative rates, where fraudulent transactions are misclassified as legitimate, reducing the model's effectiveness. Traditional machine learning models often struggle with imbalanced datasets, as they tend to predict the majority class (non-fraudulent transactions). To address this, researchers and practitioners have explored various methods: Oversampling, Undersampling, Cost-sensitive Learning, and Advanced Algorithms. Dornadula and Geetha used SMOTE to generate synthetic samples for the minority class to balance the dataset [7]. Roy applied undersampling to reduce the number of majority class samples to balance the dataset [8]. Assigning higher misclassification costs to the minority class, which Olowookere and Adewale used to handle credit card fraud, helps improve detection rates [9]. Techniques like ensemble learning and deep learning models are used to enhance detection accuracy despite data imbalance, like Babu and Pratap, who applied CNNs to detect fraudulent transactions [10].

II. Data Pre-Processing

The initial inspection of the dataset [11] revealed a significant class imbalance, with fraudulent transactions accounting for only a small fraction of the total. Specifically, there were 492 fraudulent transactions compared to 284,315 normal transactions, emphasizing the need for pre-processing techniques. A two-step approach was employed to address the imbalance. First, feature selection identified 11 features that were highly relevant for classification, combining correlation analysis and ANOVA tests [12]. Features with correlation values exceeding ±0.1 were selected, and their significance was further confirmed through ANOVA scores greater than 2000. The refined feature set included key attributes such as V1, V18, V7, V3, V16, V10, V12, V14, V17, V11, and V4, which demonstrated strong predictive power for fraudulent transactions. Second, random undersampling of non-fraudulent transactions was performed, selecting 492 samples to match the number of fraudulent cases. SMOTE (Synthetic Minority Oversampling Technique) was applied to generate synthetic

examples of fraudulent data, enriching the minority class. These techniques ensured that the dataset was balanced, providing the model with equal opportunities to learn from both classes [12]. It is crucial to apply undersampling and oversampling techniques exclusively during the training phase while ensuring that evaluation is conducted on the original dataset. These pre-processing steps established a balanced and well-structured dataset, significantly enhancing the effectiveness of the classification models.

III. Visualization

To better understand the dataset, two important visualizations were created. First, a pie chart was used to show the proportion of fraudulent and non-fraudulent transactions. The chart revealed a major imbalance, with fraudulent transactions making up only 0.2% of the dataset. This clearly showed the need for balancing techniques to help the model learn to detect fraud effectively.

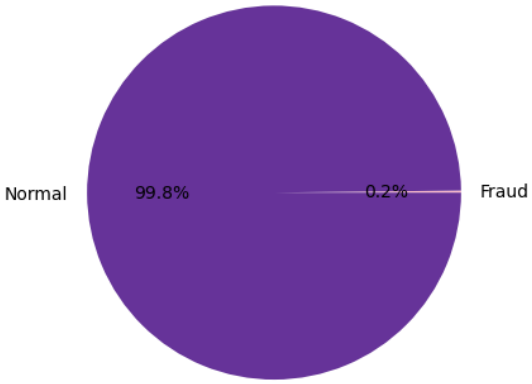


Fig. 1. fraudulent transactions make up only 0.2% of the dataset

Second, a correlation heatmap was created to examine how the features in the dataset are related to each other and the target variable (Class). The heatmap made it easy to identify which features were strongly linked to fraudulent transactions. These visualizations gave a clear understanding of the dataset's structure and were key to preparing the data for building accurate and reliable models.

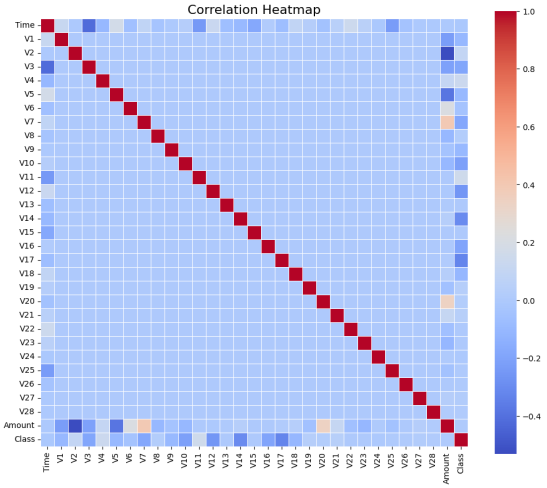


Fig. 2. correlation heatmap

IV. Modeling and Analysis

In this analysis, the target variable (Class) was defined as the detection of fraudulent transactions (1) versus normal transactions (0). Four machine learning techniques were implemented to build classification models: SVM, Random Forest, XGBoost, and Neural Networks. Each model was evaluated using key metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to assess its performance.

The Support Vector Machine (SVM) model with a linear kernel and balanced class weights was used to address the class imbalance, with feature scaling applied using StandardScaler. The model achieved 98% accuracy and 93% recall on the test set, effectively identifying fraudulent transactions. However, the precision (7%) reflects challenges in reducing false positives. Overall, SVM proved to be a reliable approach for fraud detection when combined with proper scaling and balancing techniques.

```
=== SVM ===
Accuracy: 0.9789
Recall: 0.9286
Precision: 0.0707
F1-Score: 0.1314
Sensitivity: 0.9286
Specificity: 0.9790
```

Fig. 3. evaluation indicators of SVM

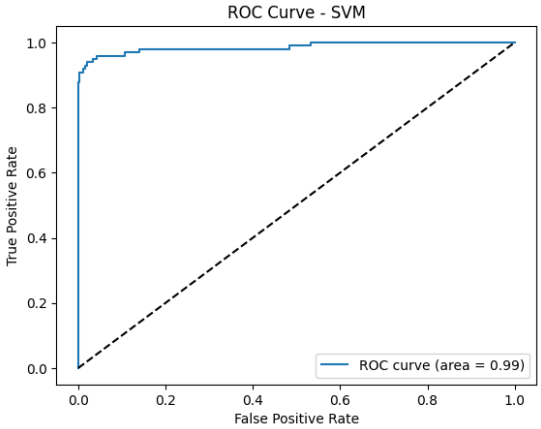


Fig. 4. ROC of SVM

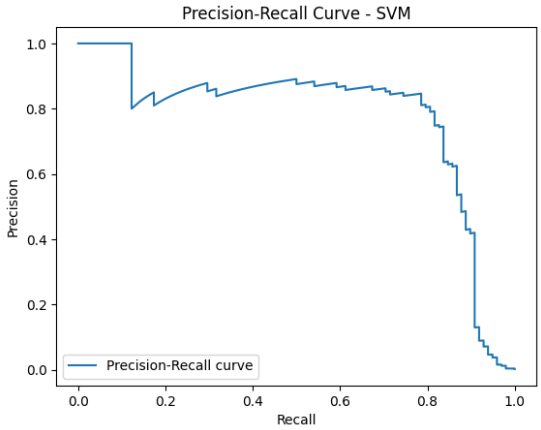


Fig. 5. PRC of SVM

The Random Forest [13] model with 100 trees and balanced class weights was implemented to address class imbalance, with a fixed random state to ensure reproducibility. After training, the model achieved an accuracy of 99% and a recall of 91% on the test set. This showed that the Random Forest model was effective in identifying most fraudulent transactions while maintaining a low false positive rate, making it a reliable choice for fraud detection.

```
=== Random Forest ===
Accuracy: 0.9965
Recall: 0.9082
Precision: 0.3179
F1-Score: 0.4709
Sensitivity: 0.9082
Specificity: 0.9966
```

Fig. 6. evaluation indicators of Random Forest

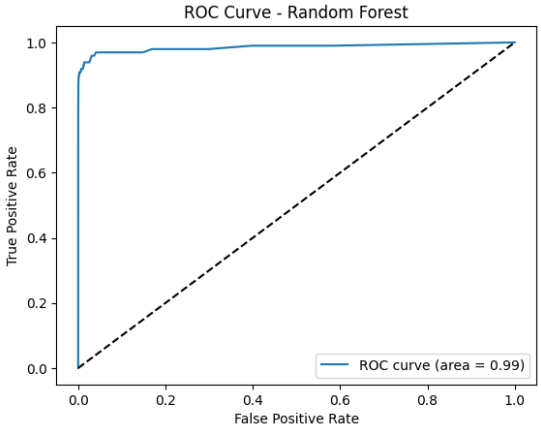


Fig. 7. ROC of Random Forest

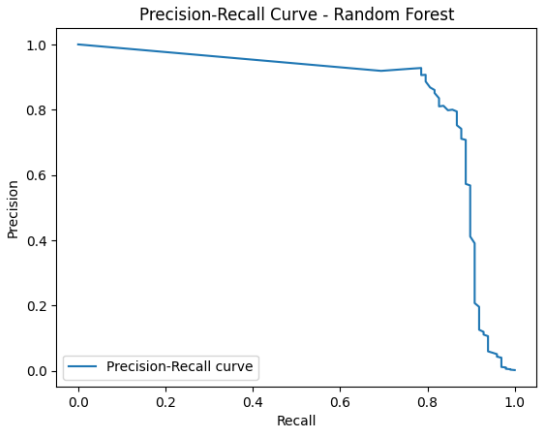


Fig. 8. PRC of Random Forest

XGBoost [14], a gradient-boosting algorithm, was employed to handle the class imbalance by assigning a high penalty to misclassifications of the minority class. The scale_pos_weight parameter was set based on the ratio of normal to fraudulent transactions in the training data. The model was trained on the same balanced dataset and achieved an accuracy of 98% and a recall of 91%, demonstrating an outstanding ability to distinguish between fraudulent and normal transactions.

```
=== XGBoost ===
Accuracy: 0.9773
Recall: 0.9082
Precision: 0.0648
F1-Score: 0.1209
Sensitivity: 0.9082
Specificity: 0.9774
```

Fig. 9. evaluation indicators of XGBoost

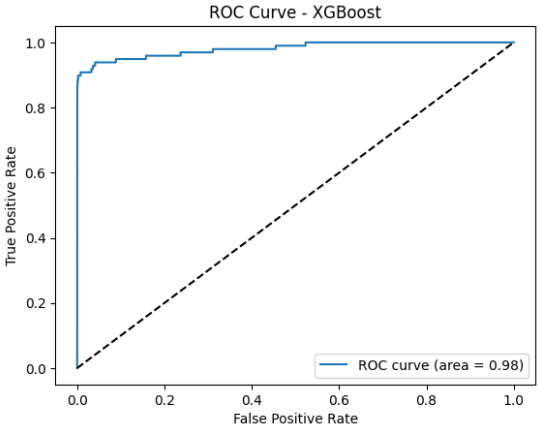


Fig. 10. ROC of XGBoost

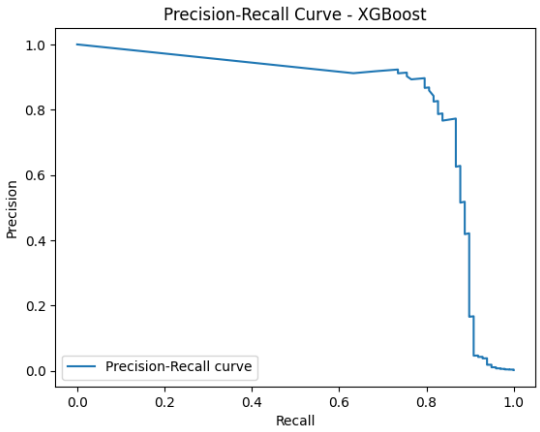


Fig. 11. PRC of XGBoost

A Neural Network [15] model was developed using TensorFlow [16] to capture and analyze complex patterns within the data. The network architecture included three hidden layers with 64, 32, and 16 nodes, respectively, and ReLU activation functions. A focal loss function was used to address class imbalance, which gave higher weight to misclassified minority class instances. The model was trained for 100 epochs with a batch size of 32, achieving an accuracy of 98.5% and a recall of 90%. This revealed that the Neural Network performed well in identifying fraud while slightly overfitting the training data. Despite this, its ability to model complex patterns made it a valuable addition to the analysis.

```
=== Neural Network ===
Accuracy: 0.9858
Recall: 0.8980
Precision: 0.0990
F1-Score: 0.1783
Sensitivity: 0.8980
Specificity: 0.9859
```

Fig. 12. evaluation indicators of Neural Network

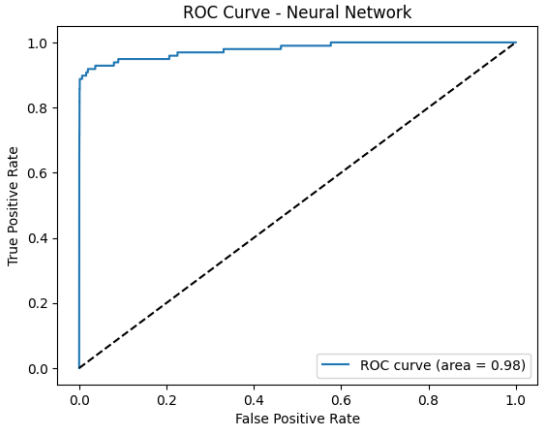


Fig. 13. ROC of Neural Network

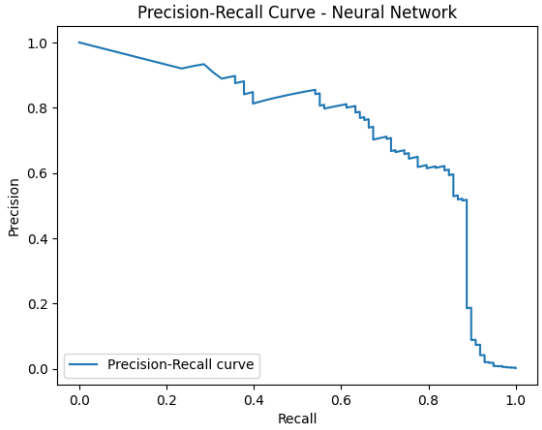


Fig. 14. PRC of Neural Network

V. Modeling Evaluation

The evaluation of the models, supported by metrics and confusion matrices, highlights distinct trade-offs. Random Forest demonstrated the highest precision (0.3179) and balanced performance with the highest accuracy (99.65%) and specificity (99.66%), making it the most effective at minimizing false positives while reliably detecting fraudulent transactions. SVM prioritized recall (92.86%) and sensitivity, ensuring most fraudulent transactions were detected, but its low precision (0.0843) suggests challenges in handling false positives. The

Neural Network offered a good balance with high accuracy (98.58%) and specificity (98.59%) but struggled with precision (0.0990), similar to SVM. XGBoost, while slightly outperforming Neural Networks in recall (90.82%), had the lowest precision (0.0648) among the models, indicating limited effectiveness in identifying fraudulent transactions.

TABLE I
SUMMARY OF EVALUATION INDICATORS

Metric	SVM	Random Forest	XGBoost	Neural Network
Accuracy	0.9825	0.9965	0.9773	0.9858
Recall	0.9286	0.9082	0.9082	0.8980
Precision	0.0843	0.3179	0.0648	0.0990
F1-Score	0.1546	0.4709	0.1209	0.1783
Sensitivity	0.9286	0.9082	0.9082	0.8980
Specificity	0.9826	0.9966	0.9774	0.9859
AUROC	0.99	0.99	0.98	0.98

From the confusion matrices, Random Forest effectively reduced false positives and false negatives, while SVM and Neural Networks showed higher false positive rates. Given these results, Random Forest is the most suitable model for fraud detection when precision and balanced performance are prioritized. However, if the goal is to maximize recall and detect as many fraudulent transactions as possible, SVM may be a viable alternative. The choice of model should align with the specific objectives of the fraud detection system, whether it prioritizes minimizing false positives or capturing the maximum number of fraudulent cases.

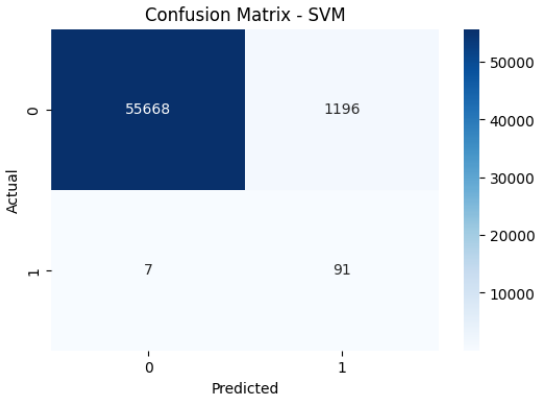


Fig. 15. Confusion Matrix of SVM

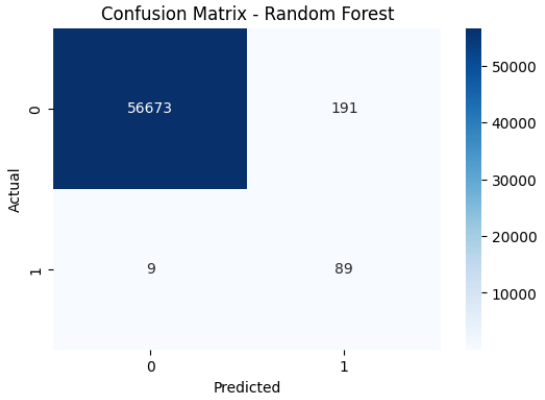


Fig. 16. Confusion Matrix of Random Forest

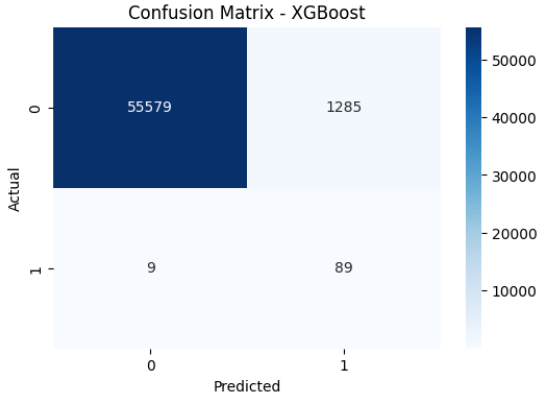
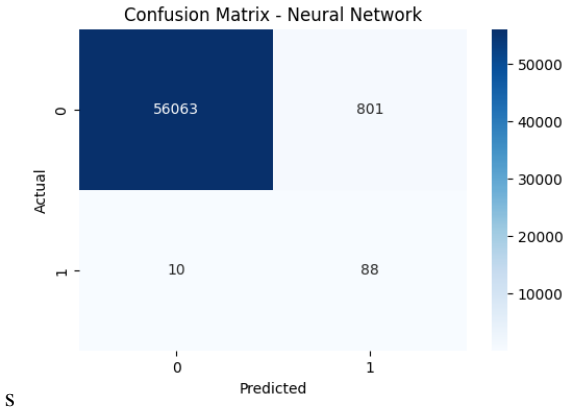


Fig. 17. Confusion Matrix of XGBoost



S

Fig. 18. Confusion Matrix of Neural Network

VI. FUTURE WORKS

Future work aims to enhance the performance and reliability of the fraud detection system through two key strategies. First, parameter tuning will be conducted using cross-validation to optimize the hyperparameters of the SVM, Random Forest, XGBoost, and Neural Network models, ensuring that each method achieves a better performance. Second, an ensemble

approach will be explored by combining predictions from all four models. A transaction will be classified as fraudulent only if at least three models agree, increasing the system's overall reliability. This ensemble strategy aims to leverage the strengths of each model while minimizing its weaknesses, providing a more accurate fraud detection solution.

REFERENCES

[1] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, Nigeria, 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782.

[2] Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002). "Smote: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, https://www.jair.org/index.php/jair/article/view/10302.

[3] Liu, W., Wang, X. and Peng, W. (2020). State of the art: Secure mobile payment — IEEE Journals & Magazine — IEEE Xplore, https://ieeexplore.ieee.org/document/8947955/.

[4] Dablain, D., Krawczyk, B. and Chawla, N. V. (2022). "Deepsmote: Fusing deep learning and smote for imbalanced data — IEEE Journals & Magazine — IEEE Xplore," IEEE Transactions on Neural Networks and Learning Systems (Volume: 34, Issue: 9, September 2023), https://ieeexplore.ieee.org/document/9694621/.

[5] Bahnsen, A. C., Aouada, D. and Ottersten, B. (2014). "Example-dependent cost-sensitive logistic regression for credit scoring — IEEE conference publication — IEEExplore," 2014 13th International Conference on Machine Learning and Applications, pp. 263–269. https://ieeexplore.ieee.org/document/7033125/.

[6] Ingole, S., Kumar, A., Prusti, D. and Rath, S. K. (2021). "Service-based credit card fraud detection using Oracle SOA Suite - SN Computer science," SpringerLink, https://link.springer.com/article/10.1007/s42979-021-00539-2.

[7] Dornadula V. N. and Geetha, S. (2019). "Credit card fraud detection using machine learning algorithms," Procedia Computer Science, https://www.sciencedirect.com/science/article/pii/S187705092030065X.

[8] Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams S. and Beling, P. (2018). "Deep learning detecting fraud in credit card transactions," 2018 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2018, pp. 129-134, https://ieeexplore.ieee.org/document/8374722.

[9] Olowookere, T. A. and Adewale, O. S. (2020). "A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach," Scientific African, https://www.sciencedirect.com/science/article/pii/S2468227620302027.

[10] Babu, A. M. and Pratap, A. (2020). "Credit Card Fraud Detection Using Deep Learning — IEEE Conference publication — IEEE Xplore," 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS), pp. 32–36, https://ieeexplore.ieee.org/document/9332497.

[11] https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

[12] Lars Sthle and Wold, S. (1989). "Analysis of variance (ANOVA)," Chemometrics and Intelligent Laboratory Systems, https://www.sciencedirect.com/science/article/pii/0169743989800954.

[13] https://en.wikipedia.org/wiki/Random_forest.

[14] Hastie, T. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Nature, 2009.

[15] https://en.wikipedia.org/wiki/Neural_network.

[16] https://en.wikipedia.org/wiki/TensorFlow.

VII. APPENDIX

Thanks to all team members for their valuable contributions to the group project. The following is the work completed by each member:

- Dataset selection: All
- Project Proposal: Yuyang Zhou, Hepple Xi
- Data preprocessing: Yuyang Zhou
- Feature Selection: Hepple Xi
- Undersample/Oversample: Yuyang Zhou
- Data visualization: Hepple Xi
- SVM: Cynthia Wen
- Random Forest: Hepple Xi
- XGBoost: Sheng Zhang
- Neural Network: Yuyang Zhou
- Brain Storming & Presentation: All
- Draft Report: Hepple Xi, Sheng Zhang, Yuyang Zhou
- Final Report: Hepple Xi, Sheng Zhang