"DEEPGUARD": HIGH ACCURACY CREDIT CARD FRAUD DETECTION

Presented by Piyush Gupta Under the Guidance of Prof. Krystn Gutu

Seidenberg School of Computer Science & Information Systems

INTRODUCTION

The rapid technological advancement and proliferation of digital payments have amplified the usage of credit cards extensively. Consequently, this has also led to a surge in fraudulent activities associated with credit card transactions, necessitating the development of robust and precise fraud detection mechanisms. Credit card fraud can manifest in various forms, occurring both online and offline. Numerous methods are available for detecting such fraudulent activities, with research suggesting that machine learning algorithms are particularly effective in identifying and mitigating such risks. Notable algorithms such as Logistic Regression, KNN, Decision Tree, Random Forest & CNN have been instrumental in enhancing the accuracy and efficiency of fraud detection processes. In the project "DeepGuard", we aim to achieve a high level of accuracy in credit card fraud detection by leveraging sophisticated machine learning and deep learning techniques.

OBJECTIVES

The "DeepGuard" project aims to devise two predictive models using machine and deep learning algorithms to identify fraudulent transactions. By analyzing regular patterns, these models can differentiate genuine transactions from fraudulent ones. Our focus is on maximizing fraud detection with minimal false alerts, treating this as a classification challenge. Data preprocessing and PCA transformations are integral to our approach.

MODELS

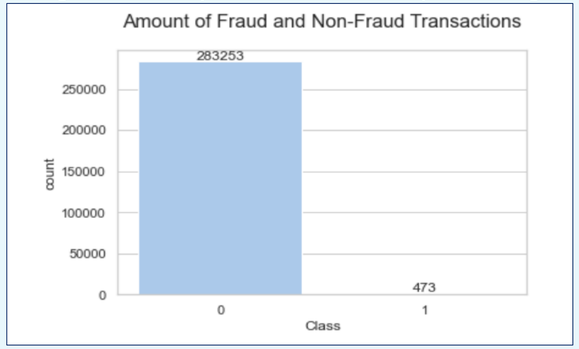
1.Logistic Regression: This technique is versatile, suited for regression and classification tasks, but is predominantly used for the latter. Logistic regression leverages dependent variables to forecast categorical outcomes.

2.Decision Tree: This tool addresses both classification and regression challenges by branching decisions based on data attributes.

3.Random Forest: This classifier harnesses multiple decision trees, each analyzing a data subset, and then consolidates their insights to enhance prediction accuracy for the entire dataset.

4.KNN (K-Nearest Neighbors): KNN operates by classifying data points based on how their neighbors are classified, making it effective for pattern recognition tasks.

5.CNN (Convolutional Neural Network): CNN is a deep learning algorithm primarily utilized for image analysis. It processes data with a grid-like topology, like an image, identifying patterns through various layers of information processing.



RESULTS

```
---------- classification report of logistic regression---------
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     84984
           1       0.70      0.64      0.67       134

    accuracy                           1.00     85118
   macro avg       0.85      0.82      0.83     85118
weighted avg       1.00      1.00      1.00     85118

--------------- Accuracy of logistic regression ----------------
Accuracy:- 0.9990013863107686
F1-Score:- 0.6692607003891051
Precision:- 0.6991869918699187
```

```
---------- classification report of decision tree---------
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     84984
           1       0.70      0.64      0.67       134

    accuracy                           1.00     85118
   macro avg       0.85      0.82      0.83     85118
weighted avg       1.00      1.00      1.00     85118

----------------Accuracy of decision tree----------------
Accuracy:- 0.9991306186705514
F1-Score:- 0.7375886524822697
Precision:- 0.7027027027027027
```

```
---------- classification report of random forest---------
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     84984
           1       0.70      0.64      0.67       134

    accuracy                           1.00     85118
   macro avg       0.85      0.82      0.83     85118
weighted avg       1.00      1.00      1.00     85118

----------------Accuracy of random forest----------------
Accuracy:- 0.9995183157499001
F1-Score:- 0.8270042194092827
```

```
---------- classification report of KNN---------
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     84984
           1       0.70      0.64      0.67       134

    accuracy                           1.00     85118
   macro avg       0.85      0.82      0.83     85118
weighted avg       1.00      1.00      1.00     85118

-----------------Accuracy of KNN----------------
Accuracy:- 0.9984609600789492
F1-Score:- 0.05755395683453237
Precision:- 0.8
```

```
---------- classification report of CNN---------
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     84984
           1       0.70      0.64      0.67       134

    accuracy                           1.00     85118
   macro avg       0.85      0.82      0.83     85118
weighted avg       1.00      1.00      1.00     85118

----------------Accuracy of CNN----------------
Accuracy:- 0.9993773349937733
```

CONCLUSIONS

The "DeepGuard" project successfully deployed machine learning and deep learning models for credit card fraud detection. Among the models tested, Random Forest achieved the highest accuracy at 0.9995. However, all models demonstrated impressive accuracy rates, underscoring their potential for real-world applications

S.No.	Model Name	Accuracy
1	Logistic Regression	0.9990
2	Decision Tree	0.9991
3	Random Forest	0.9995
4	KNN	0.9985
5	CNN	0.9994

REFERENCES

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.
- Jagtap, A., & Shirbahadurkar, S. (2020). Detection of Credit Card Fraud Using Machine Learning Algorithms: A Survey. Procedia Computer Science, 167, 191–201.
- Hassan, M.M., Huda, S., Uddin, M.Z., Almogren, A., & Alamri, A. (2019). Secure and Robust Cloud-Based Large-Scale IoT Infrastructure for Smart Cities. IEEE Access, 7, 50724-50734.