PACE UNIVERSITY

FALL 2023

Analytics Capstone Project (By Prof. Krystyn Gutu)

"DeepGuard": High-Accuracy Credit Card Fraud Detection

Analysis & Result

Submitted by :- Piyush Gupta

CRN: - 74233; CS 668 FALL 2023

Abstract:

In "DeepGuard," our innovative approach converges advanced analytical algorithms, cultivating a multi-faceted defense matrix against credit card fraud. By meticulously harnessing and integrating the distinctive capabilities of each selected model, we foster a synergistic environment where predictive precision and robustness are optimized. This fusion of methodologies allows for a more holistic view, ensuring nuanced detection mechanisms that adapt and respond to the dynamic landscape of financial fraud, thus bolstering the safety and trust in electronic transactions.

Model Overview:

1. **Logistic Regression:**
   A foundational model in our arsenal is Logistic Regression. It is instrumental due to its simplicity and effectiveness in dealing with classification problems. It offers a solid baseline in fraud detection tasks by differentiating between fraudulent and legitimate transactions based on selected features.

2. **Decision Tree:**
   The Decision Tree algorithm is incorporated for its interpretability and efficiency. By subdividing the data into subsets based on feature values, it renders a hierarchical decision-making structure that facilitates the identification of potential fraud instances. However, it requires meticulous tuning to avoid overfitting.

3. **Random Forest:**
   Building upon Decision Trees, Random Forests contribute depth and stability. By aggregating multiple decision trees to produce a more generalized model, Random Forests enhance predictive accuracy and resilience against overfitting, making it a pivotal component in our multifaceted approach.

4. **K-Nearest Neighbor (KNN):**
   KNN accentuates the power of the community in decision-making. It classifies transactions by aligning them with the categories of their nearest neighbors in the feature space, hence using the wisdom of proximate data points to discern the nature of transactions.

5. **Convolutional Neural Network (CNN):**
   CNN introduces a layer of neural network sophistication. Primarily revered in image and sequential data processing, its adaptation to fraud detection presents a novel perspective. CNN's capacity to unearth intricate patterns and dependencies in the input data enriches our model's discriminatory prowess.

Result & Analysis: -

1. **Logistic Regression:**

Result: -

```
---------- 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
```

Analysis

The Logistic Regression model demonstrates excellent performance with an overall accuracy of 99.9%. For class '0', both precision and recall are perfect at 1.00, indicating impeccable detection of true negatives. For class '1', the model has a precision of 0.70 and a recall of 0.64, indicating a good but not perfect detection of true positives. The F1-Score, which harmonizes precision and recall, stands at 0.67 for class '1', revealing a balanced yet slightly moderate performance for positive fraud detection.

2. **Decision Tree:**

Result: -

```
---------- 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
```

Analysis

The Decision Tree model showcases an outstanding overall accuracy of 99.91%. For class '0', the precision and recall are both optimal at 1.00, denoting flawless detection of true negatives. In contrast, for class '1', the precision is 0.70 and recall is 0.64, suggesting a commendable but not absolute detection of true positives. The F1-Score for class '1' is 0.67, pointing to a balanced performance, albeit with a slight leaning towards precision over recall. In essence, while the model excels in identifying valid transactions, its efficacy in pinpointing fraudulent activities has slight margins for enhancement.

3. **Random Forest:**

Result: -

```
---------- 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
```

Analysis

The Random Forest model boasts an exceptional accuracy rate of 99.95%. For class '0', both precision and recall stand at a perfect 1.00, indicating impeccable detection of genuine transactions. For class '1', the precision is at 0.70 and recall at 0.64, signifying a competent, though not perfect, detection of fraud. The F1-Score for class '1' is noticeably higher at 0.67, reflecting a well-rounded performance between precision and recall. Overall, the model demonstrates excellent capability in recognizing legitimate transactions, while its ability to detect fraudulent ones, though commendable, presents room for further optimization.

4. **K-Nearest Neighbor (KNN):**

Result: -

```
---------- 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
```

Analysis

The KNN model showcases an impressive accuracy of 99.84%. For class '0', precision and recall are both at an ideal 1.00, indicating flawless identification of legitimate transactions. When it comes to class '1', the precision is commendable at 0.8, suggesting fewer false positives, while the recall stands at 0.64, pointing to a moderate detection rate of fraudulent transactions. The F1-Score for class '1' is 0.67, denoting a balanced performance between precision and recall. In summary, the KNN model excels in spotting genuine transactions, and while it has a respectable performance in detecting frauds, there's potential for improvement in its recall for class '1'.

5. **Convolutional Neural Network (CNN):**

Result: -

```
---------- 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
```

Analysis

The CNN model presents a near-perfect accuracy of 99.94%. For class '0', both precision and recall are at an optimal 1.00, suggesting impeccable identification of legitimate transactions. In contrast, for class '1', the precision stands at 0.70, which means the model has a fair level of confidence in detecting fraudulent activities. The recall for class '1' is 0.64, indicating a moderate capability to identify actual frauds. With an F1-Score of 0.67 for class '1', the CNN model showcases a balanced relationship between precision and recall. In essence, while the CNN model is excellent at classifying genuine transactions, its detection of fraudulent ones can be enhanced, especially in terms of recall.

Conclusion: -

The diverse array of models tested, including Logistic Regression, Decision Tree, Random Forest, KNN, and CNN, showcased varying degrees of strengths in fraud detection. While all models demonstrated high accuracy rates for classifying legitimate transactions, there were noticeable differences in their ability to detect fraudulent activities. The Random Forest and CNN models emerged as top contenders with near-perfect overall accuracy. Notably, the CNN model, typically renowned for image classification, showed promising results in this financial domain, solidifying its adaptability and potential in various application scenarios. The F1-Scores, which balance precision and recall, indicated areas of improvement, especially in the fine-tuning of models to better detect frauds without increasing false positives. In conclusion, "DeepGuard" has paved the way for a robust and multifaceted approach to credit card fraud detection, emphasizing the importance of model diversity and continuous optimization in the ever-evolving landscape of financial security.