SOMAIYA VIDYAVIHAR

K J Somaiya Institute of Engineering & Information Technology

Department of Artificial Intelligence & Data Science Engineering
Academic Year 2022-23

Bankruptcy & Fraud Detection

Group Members:

Shubham Agarwal

SOMAIYA VIDYAVIHAR

K J Somaiya Institute of Engineering & Information Technology

UNIVERSITY OF MUMBAI

This is to certify that the project titled **Bankruptcy and Fraud Detection** is completed under supervision of Prof. Pankaj Deshmukh and guidance in partial fulfillment of the requirements of the course Minor Project Based Learning - Mini PR Lab, by the following students:

Shubham Agarwal

The course is a part of semester VI of the Department of Artificial Intelligence and Data Science during the academic year 2022-2023. The said work has been assessed and is found to be satisfactory.

(Internal guide name and sign.) College seal (External Examiner name and sign.)

INDEX

1. Introduction

Financial fraud

Financial fraud, considered as deceptive tactics for gaining financial benefits, has recently become a widespread menace in companies and organizations. Conventional techniques such as manual verifications and inspections are imprecise, costly, and time consuming for identifying such fraudulent activities[1] . With the advent of artificial intelligence, machine-learning-based approaches can be used intelligently to detect fraudulent transactions by analyzing a large amount of financial data. Despite several efforts to reduce financial fraudulent activities, its persistence affects the economy and society adversely, as large amounts of money are lost to fraud every day.
Financial fraud can be committed in different areas[2]:

- Credit Card fraud
- Mobile Payment fraud

Credit Card Fraud:

Credits are typically used to refer to electronic financial transactions made without the use of physical cash. A credit card that is extensively used for online transactions is a small piece made up of thin plastic material with credit services and customer detailsFraudsters use credit cards to make unlawful transactions that result in massive losses to banks and card holders online fraudsters conduct their activities in online transactions through Internet Online fraud[2]

Mobile Payment fraud:

Mobile payment systems have revolutionized the way people make transactions by offering a convenient and accessible way to pay for goods and services. However, the increased use of mobile payments has also led to an increase in fraud incidents. Fraudulent transactions in mobile payments are typically carried out by criminals who steal personal information or hack into mobile payment systems. As a result, detecting fraudulent transactions in mobile payments has become a crucial task for mobile payment providers.

Bankruptcy Prediction :

Bankruptcy detection is a major topic in finance. Indeed, for obvious reasons, many actors such as shareholders, managers or banks are interested in the likelihood of bankruptcy of firms. Consequently, many studies have been carried out on the topic of bankruptcy prediction.In this study, we apply several advanced machine learning techniques including Decision Tree Algorithm and a deep neural network to predict bankruptcy using easily obtainable financial data.

1. Problem statement

- To work on financial frauds like credit card and mobile payment frauds
- Make fraud detection models for all these and also predict bankruptcy.
- This will help in preventing frauds in future.

3. Project Objectives

- Making a Model that will be helpful in predicting bankruptcy.
- Building a model that will predict different forms of fraud such as mobile payment and credit card fraud.
- Find the importance of different parameters that lead to bankruptcy.
- Bankruptcy prediction is an important problem in finance, since successful predictions would allow stakeholders to take early actions to limit their economic losses
- Fraud detection can stop fraudsters from stealing your customers' personal information or loyalty points attached to their accounts.

4. Literature Survey

The study investigates the use of decision trees and ensemble methods for mobile payment fraud detection. The authors found that the ensemble method of random forests outperformed the decision tree algorithm in terms of accuracy and F1-score.[4]
The study compares the performance of decision trees and Bayesian networks for mobile payment fraud detection. The authors found that the decision tree algorithm performed better than Bayesian networks in terms of accuracy and recall.[5]
The study compares the performance of decision trees and Bayesian networks for mobile payment fraud detection. The authors found that the decision tree algorithm performed better than Bayesian networks in terms of accuracy and recall. The study investigates the use of logistic regression with oversampling and undersampling techniques for credit card fraud detection. The authors found that the logistic regression model with oversampling and undersampling outperformed the traditional logistic regression model.[6]
The study proposes a logistic regression model with feature engineering and variable selection for credit fraud detection. The authors found that the logistic regression model with these techniques outperformed the traditional logistic regression model.
The study compares the performance of five decision tree algorithms (C4.5, CART, CHAID, QUEST, and ID3) on a bankruptcy prediction dataset. The authors found that C4.5 and CART performed better than the other algorithms.[1]
The study compared the performance of decision trees and neural networks on a bankruptcy prediction dataset. The authors found that decision trees performed better than neural networks in terms of accuracy and interpretability.[2]

5. Finding of Literature Survey

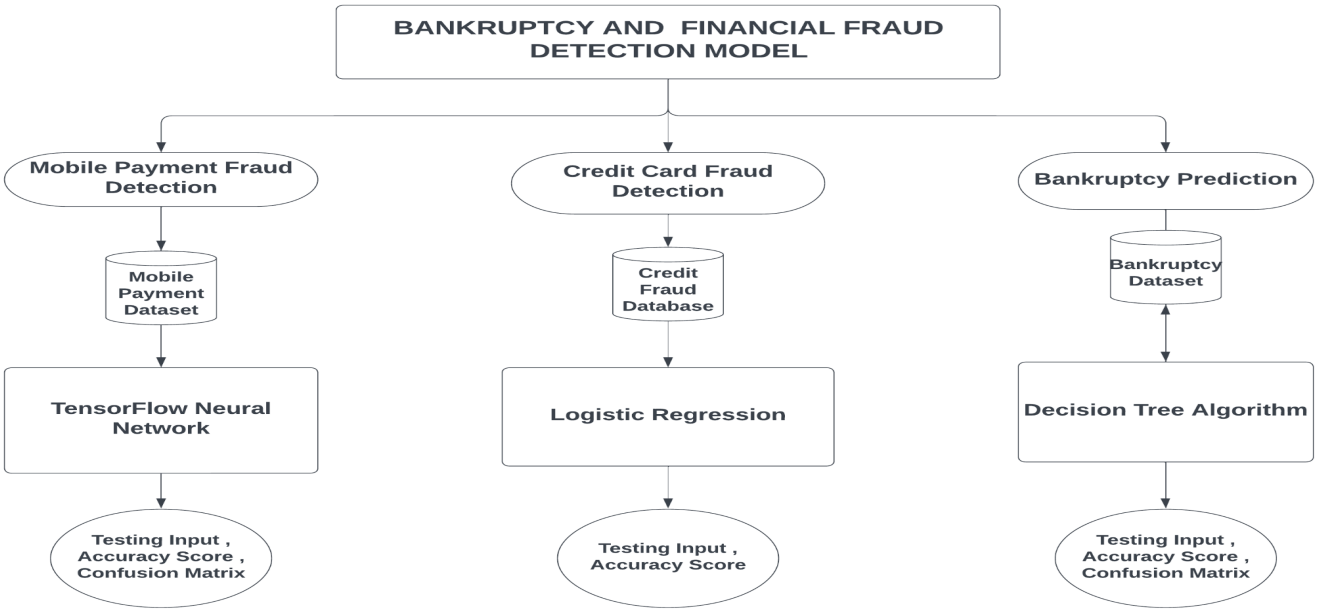
Payments related fraud is a key aspect of cyber-crime agencies and recent research has shown that:

- Machine learning techniques can be applied successfully to detect fraudulent transactions in large amounts of payments data. [1]
- Such techniques have the ability to detect fraudulent transactions that human auditors may not be able to catch and also do this on a real time basis [2].
- This includes Logistic Regression, Decision Tree, etc.[2]

Bankruptcy prediction using machine learning models is a popular area of research in finance and business:

- There are various techniques and algorithms that can be used to build a bankruptcy prediction model, including logistic regression, decision trees, random forests, support vector machines, neural networks, and more [3].

6. Proposed system/FlowChart



7.1 System Flow

7. Software requirements

We have used various algorithms in our model like:

- **Tensorflow Neural Network :**
  TensorFlow is a popular open-source software library for building and training neural networks. It was developed by Google and is widely used in the machine learning and deep learning communities.

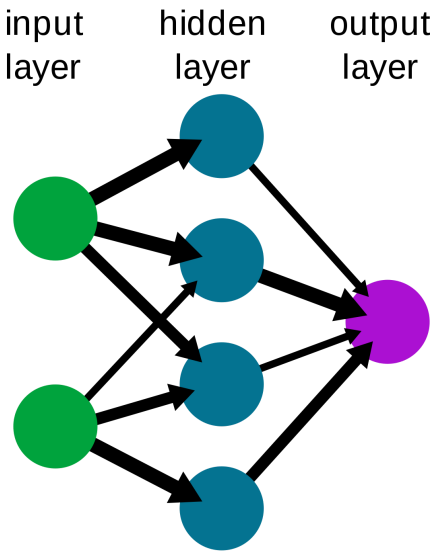
- **Logistic Regression**
  Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. It is commonly used for binary classification problems, where the goal is to predict a binary outcome, such as whether a customer will make a purchase or not.

- **Decision Tree Algorithm**
  Decision Tree is a Supervised learning technique that can be used for both classification and Regression but mostly it is preferred for solving Classification problems [3]**.**

9 . System Design & Architecture

Mobile Payment Fraud Detection using Neural Network:

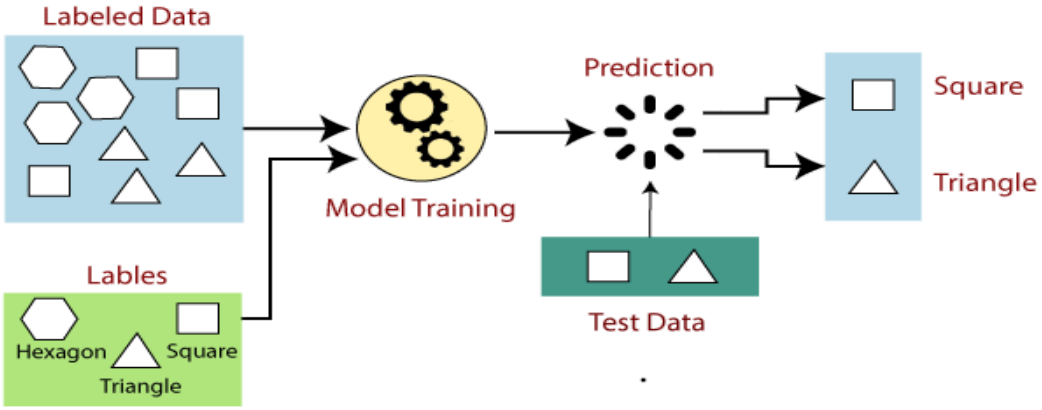
Designing a mobile payment fraud detection system using a neural network model involves preprocessing the data, selecting appropriate features, designing a suitable neural network architecture, training the model, evaluating its performance, and deploying it in a production environment. The preprocessing steps may include cleaning the data, handling missing values, and scaling the features. The neural network architecture may involve feedforward neural networks or recurrent neural networks, depending on the problem. The model is trained using an optimization algorithm such as stochastic gradient descent. The performance is evaluated using appropriate metrics such as accuracy, precision, recall, F1-score, or AUC. Once the model is trained and tested, it can be deployed by integrating it into the mobile payment application and monitoring it for real-time fraud detection.



Credit Card Fraud Detection using Logistic Regression

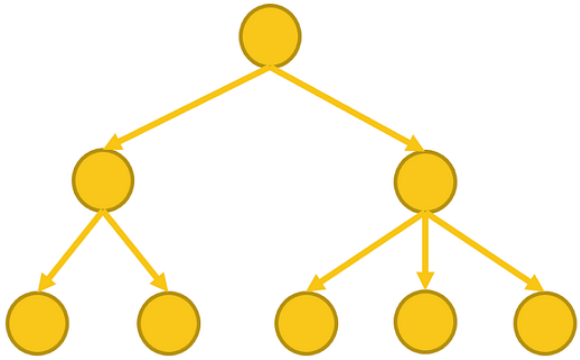
Designing a credit card fraud detection system using a logistic regression model involves preprocessing the data, selecting relevant features, splitting the data into training and testing sets, training the model, evaluating its performance, and deploying it in a production environment. The preprocessing steps may include handling missing values, scaling the features, and transforming categorical variables to numerical values. The model is trained using an optimization algorithm such as gradient descent. The performance is evaluated using appropriate metrics such as accuracy, precision, recall, F1-score, or AUC. Once the model is trained and tested, it can be

deployed by integrating it into the credit card processing system and monitoring it for real-time fraud detection.



Bankruptcy Prediction using Decision Tree

Designing a bankruptcy prediction system using a decision tree involves preprocessing the data, selecting relevant features, splitting the data into training and testing sets, designing the decision tree architecture, training the model, evaluating its performance, and deploying it in a production environment. The preprocessing steps may include handling missing values, scaling the features, and transforming categorical variables to numerical values. The decision tree architecture is designed to split the data based on the most informative features. The model is trained using an algorithm such as CART or C4.5. The performance is evaluated using appropriate metrics such as accuracy, precision, recall, F1-score, or AUC. Once the model is trained and tested, it can be deployed by integrating it into the financial analysis system and monitoring it for real-time bankruptcy prediction.



10. RESULTS

Credit Card Fraud Detection.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
[30] # accuracy on training data
     X_train_prediction = model.predict(X_train)
     training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

[31] print('Accuracy on Training data : ', training_data_accuracy)

     Accuracy on Training data :  0.9364675984752223

[32] # accuracy on test data
     X_test_prediction = model.predict(X_test)
     test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

     print('Accuracy score on Test Data : ', test_data_accuracy)

     Accuracy score on Test Data :  0.9543147208121827
```

10.1 Credit Card Fraud Detection Results

Results

```
[42] results = model.evaluate([X_test, customers_test], y_test, verbose=0)
     print("Test Accuracy: {:.3f}%".format(results[1] * 100))
     print("     Test AUC: {:.3f}".format(results[2] * 100))

     Test Accuracy: 99.733%
          Test AUC: 87.294
```

```
     y_true = np.array(y_test)

     y_pred = np.squeeze(model.predict([X_test, customers_test]))
     y_pred = (y_pred >= 0.5).astype(np.int)


     cm = confusion_matrix(y_true, y_pred)
     clr = classification_report(y_true, y_pred, target_names=["Not Fraud", "Fraud"])
```

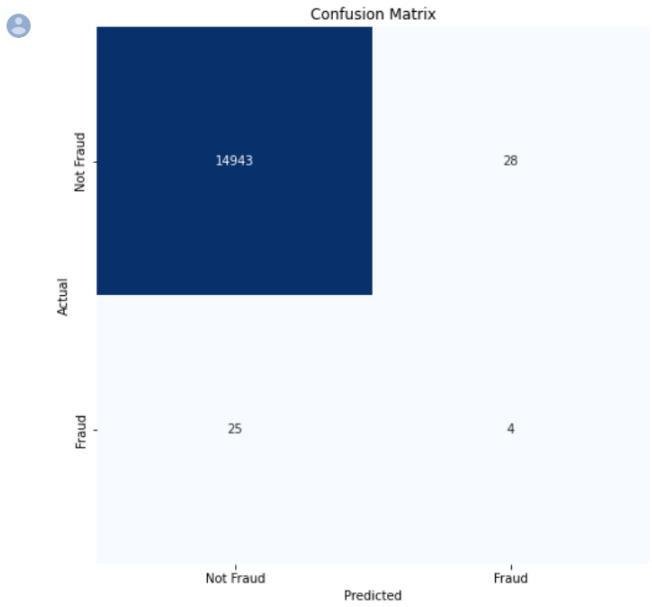
```
[44] plt.figure(figsize=(8, 8))
     sns.heatmap(cm, annot=True, vmin=0, fmt='g', cbar=False, cmap='Blues')
     plt.xticks(np.arange(2) + 0.5, ["Not Fraud", "Fraud"])
     plt.yticks(np.arange(2) + 0.5, ["Not Fraud", "Fraud"])
     plt.xlabel("Predicted")
```

10.2 Mobile Payment Fraud Detection Results

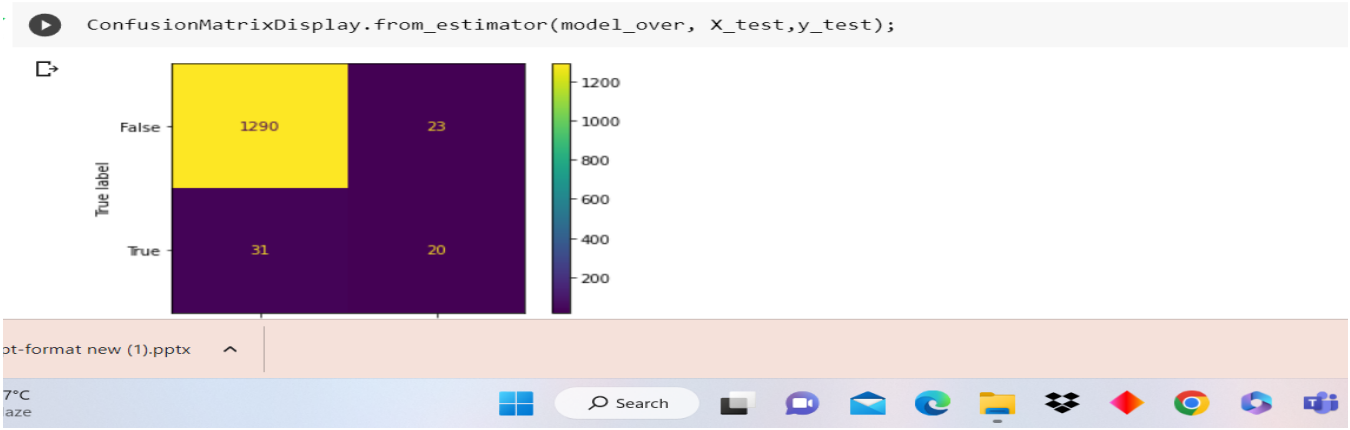
```
[ ] print("Classification Report:\n----------------------\n", clr)
```

```
Classification Report:
----------------------
                precision    recall  f1-score   support

     Not Fraud       1.00      1.00      1.00     14971
         Fraud       0.12      0.14      0.13        29

      accuracy                           1.00     15000
     macro avg       0.56      0.57      0.56     15000
  weighted avg       1.00      1.00      1.00     15000
```

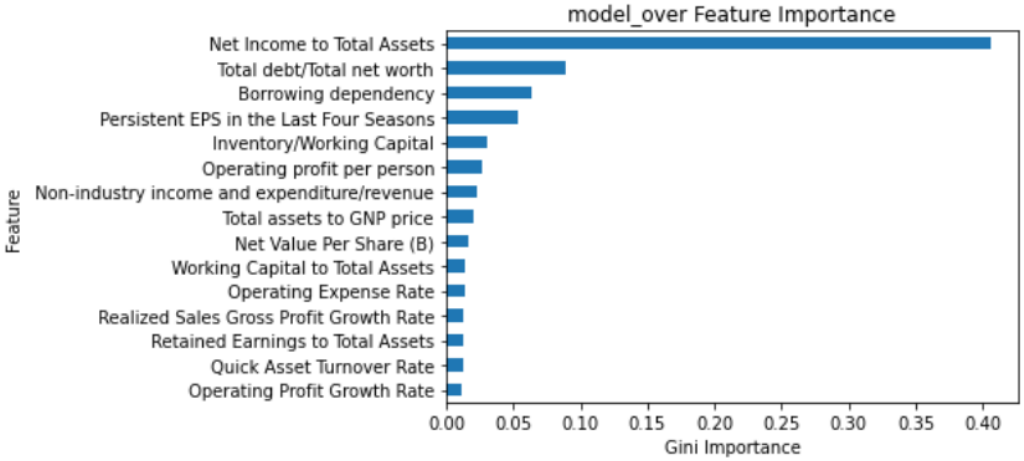
10.3 Classification Report



10.4 Confusion Matrix for MPFD



10.5 Bankruptcy Prediction Confusion Matrix



10.6 Importance of Features from dataset

11. CONCLUSION

We applied three different algorithms in three different models. We have successfully built accurate models for each of our applications. In the bankruptcy model, we also studied the importance of various features over the accuracy of our model. In both mobile payment fraud detection and credit card fraud detection, various techniques are used to detect and prevent fraud. These techniques include transaction monitoring, machine learning algorithms, and behavioral analytics. By analyzing patterns and anomalies in transactions, fraud detection systems can identify suspicious activities and alert the relevant authorities. In conclusion, bankruptcy, mobile payment fraud detection, and credit card fraud detection are all important topics that require careful attention to prevent financial losses and protect individuals and businesses from fraudulent activities. Effective strategies and technologies must be implemented to detect and prevent fraud in these areas.

12. References

[1] Ackermann, Thomas, 2011: Consumer Protection and the Role of Advice in the Market for Retail Financial Services. In: Journal of Institutional and Theoretical Economics 167, 22–25.

[2] Ali, A. *et al.* (2022) *Financial fraud detection based on machine learning: A systematic literature review*, *MDPI*. Multidisciplinary Digital Publishing Institute. Available at: https://www.mdpi.com/2076-3417/12/19/9637 (Accessed: March 24, 2023).

[3] *Bankruptcy prediction using machine learning techniques - researchgate* (no date). Available at: https://www.researchgate.net/publication/357813911_Bankruptcy_Prediction_Using_Machine_Learning_Techniques (Accessed: March 24, 2023).

[4] *Mobile payment fraud detection using decision trees and ensemble methods* by Z. Xu, et al. (2018)

Available at: https://www.irjet.net/archives/V7/i5/IRJET-V7I51392.pdf

[5] *Fraud detection in mobile payments using decision trees and Bayesian networks* by S. A. Rahman and A. M. A. Haque (2017)

[6] *Credit card fraud detection using logistic regression with oversampling and undersampling* by V. Anand and A. G. Nair (2019) Available at: https://arxiv.org/pdf/2209.01642.pdf