**WiDS ’22 – ’23 Final Documentation**

**Project UID = 6, Credit Card Fraud Detection**

**Mentor - Meet Joshi**

|  |  |  |
| --- | --- | --- |
| Team Member Name | Roll Number | E-mail-ID |
| Pranit Chute | 22M0209 | pranit11ngu@gmail.com |

**Introduction to problem Statement:**

Accorgding to data from RTI and RBI,

1) A typical banking organization loses 5% of their yearly revenues to fraud.

2) 2480 cases of fraud in public sector banks Involving Rs. 31,898.63 crore.

3) In 2017-18, total of 911 credit card frauds amounting to Rs 65.26 crore.

The purpose of banks is, to keep highly profitable consumers under risk from banking frauds. This is a problem that both banks and customers are concerned about because of the significant financial losses, credibility, and trust involved. Machine learning-based credit card fraud detection in the banking sector is not only popular, but also essential for them to implement proactive monitoring and fraud detection systems. Therefore, using machine learning models, we will identify fraudulent credit card transactions in this project.

**Existing Resources:**

1) Course on Machine Learning on coursera by Andrew NG

2) Coding basics from You tube channel – Krish Naik

3) Project basic ideas from You tube channel – Data Magic

**Proposed Solution:**

Credit Card fraud detection model is proposed which is made by using Machine Learning Algorithms.

Solution is proposed using different machine learning models like,

1) Logistic Regression

2) Dimensionality Reduction followed by Logistic Regression and,

3) Decision trees

Completely unbalanced dataset is converted into balanced dataset using under sampling method and after that this models are applied. In the proposed solution Decision tree model is giving the highest accuracy while Dimensionality reduction model using Principle Component analysis is giving little less accuracy. Logistic Regression is giving lowest accuracy among all three. So, we can use Decision tree model to detect credit card frauds.

**Proposed Solution:**

**Methodology and Progress:**

* Week 1 : (16 Dec to 24 Dec )

Brushing up basics of python, pandas and matplotlib libraries. Learning of Machine Learning algorithm which are related to our project. Coding practice from sources like you tube, Hacker rank and to know how machine learning algorithm works.

* Week 2:   (24 Dec to 31 Dec )

Data analysis and finding insights in data by data visualization. Completion of coding part of importing libraries, reading data, and visualizing it by exploratory data analysis. Unbalanced data is converted into balanced data by using under sampling method. Checking missing values in data, data cleaning, splitting the data into feature and target data, data preprocessing are done in second week of the project timeline.

* Week 3:  (31 Dec to 7 Jan )

Feature scaling and separating the training and test data. Started applying different machine learning algorithm to the model like Logistic regression, Dimensionality reduction using Principle Component Analysis followed by use of logistic regression, Decision tree model.

* Week 4: (7 Jan to 14 Jan)

Comparing accuracy of different models and tried to increase accuracy and coming to the conclusion. Making report for the project and checking for the mistakes in project.

**Results:**

**GitHub page link:** https://github.com/pranit11ngu/Credit-card-fraud-detection-project-for-WiDs/blob/1ed3873ec58e20154cf76fd5a2b9125188470ead/Credit%20card%20fraud%20detection%20project.ipynb

**Learning Value:**

I learned how to actually implement different machine learning models for actual projects. Learned use of logistic regression, dimensionality reduction and decision tree models. Learned us of plotting libraries like matplolib. Learned the workflow for ML project and different data analysis and data preprocessing techniques. Balancing the data and use of PCA are the biggest learning for me from this project. Realized the importance Machine Learning in future which will help me to increase my machine learning skills.

**Tech-stack used:**

* Python programming language
* Machine learning libraries like Numpy, pandas, matplotlib, sklearn.
* Jupyter notebook

**Reference and Citations:**

1*) Frauds – Classification and reporting (no date) Reserve Bank of India. Available at: https://www.rbi.org.in/commonman/English/Scripts/Notification.aspx?Id=578#42 (Accessed: January 21, 2023).*

*2) Krish Naik (no date) YouTube. YouTube. Available at: https://www.youtube.com/@krishnaik06 (Accessed: January 21, 2023).*

*3) https://www.youtube.com/@DataMagicAI*