

**WiDS ‘22 - ‘23 Final Documentation**

**< Project UID – 6: Credit Card Fraud Detection >**

**<Mentor: Joshi Meet Anilkumar >**

|  |  |  |
| --- | --- | --- |
| **Team Member Name** | **Roll Number** | **Email-Id** |
| Arnab Roy | 22N0082 | LDAP:22n0082@iitb.ac.in  PERSONAL:arnabroy7640@gmail.com |

**Introduction to Problem Statement**

|  |
| --- |
| Credit Card Fraud Detection using different Machine Learning Model. The aim of the project is to predict fraudulent credit card transactions using machine learning models. This is crucial from the bank’s as well as customer’s perspective. The banks cannot afford to lose their customers’ money to fraudsters. Every fraud is a loss to the bank as the bank is responsible for the fraud transactions.  The dataset contains transactions in credit cards of the customers. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. We need to take care of the data imbalance while building the model and come up with the best model by trying various algorithms. |

**Existing Resources**

|  |
| --- |
| 1. Some Youtube Videos link:   <https://www.youtube.com/watch?v=gmvvaobm7eQ&list=PLeo1K3hjS3uvCeTYTeyfe0-rN5r8zn9rw&index=1&t=0s>   1. Few articles like <https://www.geeksforgeeks.org/machine-learning/> |

**Proposed Solution**

|  |
| --- |
| At first clean the data. Then use different machine learning model and then check model score, precision, recall, F1 score for each of the models. Choose that model which one has least value of all these measures. |

**Methodology & Progress (Mention the work done week-wise)**

|  |
| --- |
| **Week 1:** In this week, I was focused to take a look on different machine learning methods and some basics of pandas.  **Week 2**: In this week dataset was provided to me. After fitting the dataset I started to visualize the dataset in python using some plots and measures. Then I normalize the dataset. Here the dataset was not balanced. So I balanced the dataset using SMOTE method.  **Week 3 & 4:** After normalizing the data I used different machine learning algorithm like logistic regression, decision tree, random forest. I also used PCA for each of the cases. |

**Results**

|  |
| --- |
| Here we use principal component analysis on the dataset, retaining 95% information. We see that Random Forest model (after applying PCA). F1 score for this model for class =1 is 1 which is phenomenal. Model accuracy is also 1 . So Random Forest model after using PCA is the best model for this dataset.  Github Link: <https://github.com/Arnab7640/CREDIT-CARD-FRAUD-DETECTION> |

**Learning Value**

|  |
| --- |
| This project helps me to involve within a real life machine learning problem. From this project I learn how to use different models in different situations, how to balance the dataset, how to increase efficiency of a model by changing the values of the parameters. |

**Tech-stack Used**

|  |
| --- |
| **Python** |

**Suggestions for others**

|  |
| --- |
| Here I have used only three machine leaning model. All of them are working very good. So one may try other machine learning models like SVM and can explore many other things. |

**References and Citations**

|  |
| --- |
| 1. [https://www.geeksforgeeks.org/dimensionality-reduction/](https://www.geeksforgeeks.org/dimensionality-reduction/%20) 2. <https://towardsdatascience.com/credit-card-fraud-detection-using-machine-learning-python-5b098d4a8edc> |