

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

**<Project UID - Name>**

**<Mentors>**

|  |  |  |
| --- | --- | --- |
| **Team Member Name** | **Roll Number** | **Email-Id** |
| Rishabh Gehlot | 222N0051 | 22N0051@iitb.ac.in |

**Introduction to Problem Statement**

|  |
| --- |
| We were given a dataset including information on the European Bank's transactions. We must predict the fraudulent transaction using several Machine Learning algorithms and select the best one. |

**Existing Resources**

|  |
| --- |
| Kaggle  YouTube videos python libraries  A book on Machine learning by O’Reilly |

**Proposed Solution**

|  |
| --- |
| Predicted fraud transactions of bank data using Logistic Regression, Random Forests.  • Used resampling techniques such as SMOTE and Under Sampling to address the issue of class imbalance  • Performed hyperparameter tuning using Grid Search to derive optimal hyperparameter values for the model  • Determined the optimal model based on criteria F1 score with a Recall value of 0.8521 |

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

|  |
| --- |
| Week 1: Brush-up basics of python, pandas & matplotlib library. Done EDA of data. Learned some basic machine-learning algorithms which we can apply to our project.  Week 2: From this, I started data analysis and found the insight (only using visualization) from the data and made a report in a python file regarding the data.  Normalizing data, data balance (using under-sampling, sampling, and SMOTE), and cleaning (for data cleaning search on google what the different factors are checked).  Week 3: Start applying different machine learning algorithms to the data.  Week 4:·Cleared doubt regarding the algorithm and its application Started  Making a report of your work while doing all the application |

**Results**

|  |
| --- |
| Please add the link to the drive folder/ GitHub page consisting of code files and reports  <https://github.com/rishabhgehlot28/credit-card-fraud-detection..git> |

**Learning Value**

|  |
| --- |
| **Many new machine learning algorithms have been learned that may be used for real-world data to increase the accuracy of models and provide accurate business solutions.**  **Developed insight on the method to use, as well as approaches for dealing with difficult data (dirty data).**  **I was introduced to Python and its libraries, which aided me in developing logical thinking skills in Python.**  **This project served as a springboard for me to obtain hands-on experience with a real-world dataset in the discipline of Data Science.** |

**Tech-stack Used**

|  |
| --- |
| **Python**  **Sklearn**  **Pandas**  **NumPy**  **Seaborn**  **Matplotlib**  **Google Collab as a notebook for writing and implementing codes** |

**Suggestions for others**

|  |
| --- |
| I felt during the entire time that our interaction with our mentor was quieter. If the period would be about 2 months, we could have done some improvement in this. By improvement, I mean learning something more to improve our results. Also, the group was very less active due to the new year and other fests at IIT Bombay during sessions so I got very little time to connect well with my other teammates |

**Contribution by each Team Member**

|  |
| --- |
|  |

**References and Citations**

|  |
| --- |
| Pandas:<https://www.youtube.com/watch?v=CmorAWRsCAw&list=PLeo1K3hjS3uuASpe-1LjfG5f14Bnozjwy>  Matplotlib: :<https://www.youtube.com/watch?v=3Xc3CA655Y4&ab_channel=freeCodeCamp.org>  Resource:  <https://www.youtube.com/watch?v=gmvvaobm7eQ&list=PLeo1K3hjS3uvCeTYTeyfe0-rN5r8zn9rw&index=1&t=0s>  <https://www.geeksforgeeks.org/machine-learning>  https://www.youtube.com/playlist?list=PLu0W\_9lII9ai6fAMHp-acBmJONT7Y4BSG  <https://www.youtube.com/watch?v=DQC_YE3I5ig&ab_channel=JohannesFrey>  https://www.youtube.com/watch?v=JnlM4yLFNuo&ab\_channel=codebasics |