

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

**<Project UID –6>**

**Meet Joshi**

|  |  |  |
| --- | --- | --- |
| **Team Member Name** | **Roll Number** | **Email-Id** |
| Tushar Khatri | 22N0066 | 22N0066@iitb.ac.in |
|  |  |  |
|  |  |  |
|  |  |  |

**Introduction to Problem Statement**

|  |
| --- |
| A dataset was given to us consisting of details of the transaction of the European Bank . We have to predict the Fraud Transaction using various Machine learning algorithms and find the optimal 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, SVM ,KNN.  • 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 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 on our project.  Week 2: From this I started data analysis of the data and found the insight (only using visualization) from the data and I made a report in a python file regarding the data.  Normalizing data, data balance (using under sampling, over sampling and SMOTE) and cleaning (for data cleaning search on google that what are the different factors are checked).  Week 3: Start applying different machine learning algorithms on the data.  Week 4:·Cleared doubt regarding the algorithm and its application Started  Making report of your work while doing all the application |

**Results**

|  |
| --- |
| <https://github.com/TusharKhatri127/Projects.git>  File named as Bank Data Fraud Detection on GitHub |

**Learning Value**

|  |
| --- |
| **Learned many new Machine learning algorithm that can be applied in real world data to improve accuracy of models to give precise decision to solutions related to business.**  **Developed intuition of which algorithm to apply along with techniques to handle data that are not very nice (dirty data).**  **Was exposed much to python and its libraries which helped me in building logical thinking in python**  **This project acted as initiator for me in the field of Data science to gain some hands-on experience with real world dataset** |

**Tech-stack Used**

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

**Suggestions for others**

|  |
| --- |
| What I felt during the entire time is that our interaction with our mentor was quite less. If the time span 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 group was very less active due to new year and other fest at IIT Bombay during session so i got very few 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 |