WiDS '22 - '23 Final Documentation

6 – Credit card fraud detection

Joshi Meet Anilkumar

Team Member Name	Roll Number	Email-Id
Ponaveni Jai Shyam Yadav	22N0265	22N0265@iitb.ac.in

Introduction to Problem Statement

We have been provided with a dataset consisting of details of the transaction of the European Bank. We need to use multiple Machine Learning algorithms to detect the fraudulent transactions and the non-fraudulent transaction.

Existing Resources

Kaggle
YouTube videos python libraries
Geeks for Geeks

Proposed Solution

Predicted fraud transactions of bank data using Logistic Regression and Random Forests
- Used resampling techniques such as SMOTE 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 0.8705

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 sampling and SMOTE), and data cleaning.

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

GitHub link:
https://github.com/ijaishyam/Credit-card-fraud-detection

Learning Value

Learned new Machine learning algorithms that can be applied in real-world data to improve the accuracy of models to give a precise decisions 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 to python and its libraries which helped me in building logical thinking in python

This project acted as an initiator for me in the field of Data Science to gain some hands-on experience with a real-world dataset

Tech-stack Used

Jupyter Notebook
Python
Pandas
NumPy
Matplotlib
Seaborn
Sklearn

Suggestions for others

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