Hacettepe University

Computer Engineering Department

BBM480 Term Study Plan

Project Details

Title	Credit Card Fraud Detection Using ML Techniques & Comparing
Supervisor	Dilmurod Vahabdjanov

Group Members

	Full Name	**Student ID**
1	Diren Boran Sezen	21946553
2	İlkim İclal Aydoğan	21992814
3	Hasan Çağrı Sarıkaya	2200356852
4		

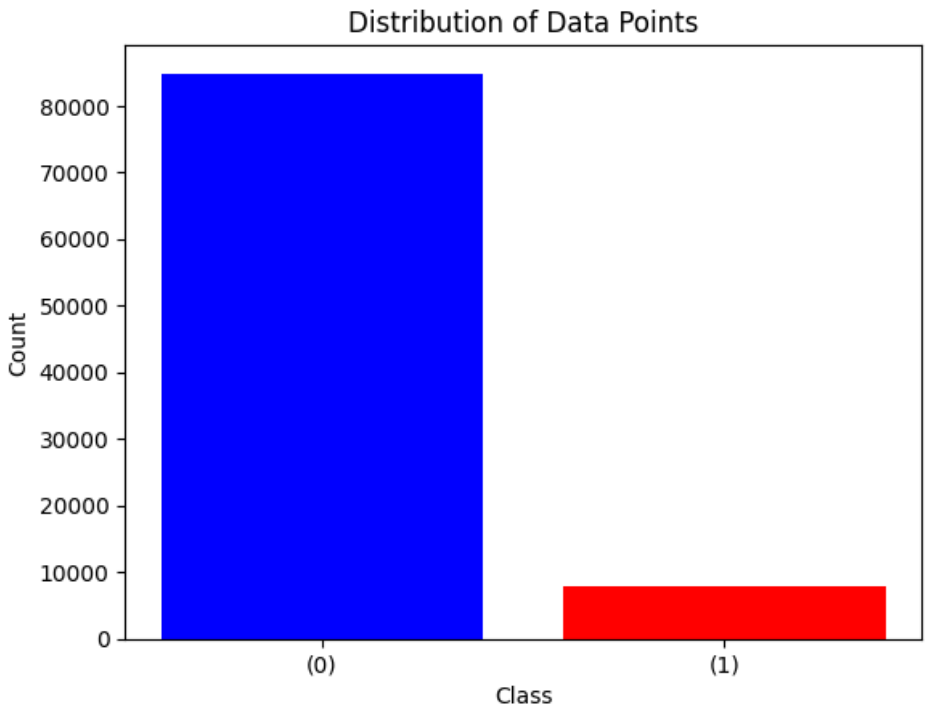
Current State (/ 50 Points)

Explain the current state of the project at the beginning of the term. Especially, emphasize the changes and development progress since the End of Term Development Report of BBM479. At this point you are expected to include solid evidence that you are making progress, such as screenshots, proofs, experiment results, data outcomes etc.

The project has progressed very well, successfully reaching all planned milestones for the last semester. As a team, we conducted thorough research into the problem, meticulously selected and preprocessed the dataset, and implemented and trained various models including Logistic Regression, KNN, and SVM, employing different sampling techniques. Subsequently, we meticulously analyzed the results and developed a demo.

The current status of the project closely resembles the status in the End of Term Development Report for BBM479. But there are some changes and developments. Following the identification of limitations within the Kaggle Credit Card Fraud Dataset, we made the decision to revisit the initial stages and conduct further research. After some exploration, we determined that the IEEE Fraud Detection competition on Kaggle presented a promising alternative dataset. We then embarked on preprocessing tasks for the selected alternative, including handling missing values, normalization, and feature engineering, with the aim of enhancing the dataset's relevance and effectiveness. These steps were crucial in ensuring data quality and preparing it for subsequent model training and evaluation. In this semester we aim to incorporate Decision Tree and Random Forest Classifier models to the project for both datasets.

We added some of the results we got from the new dataset to the figures.




```
(92712, 418) (92712,)
```

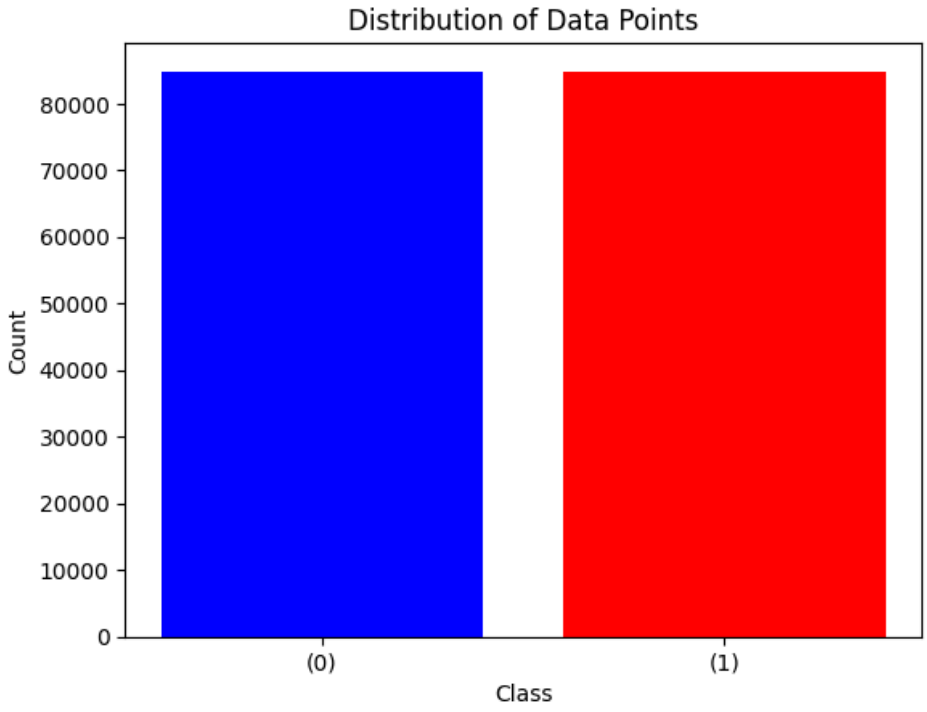


Figure - 1 New Data Distribution of Dataset

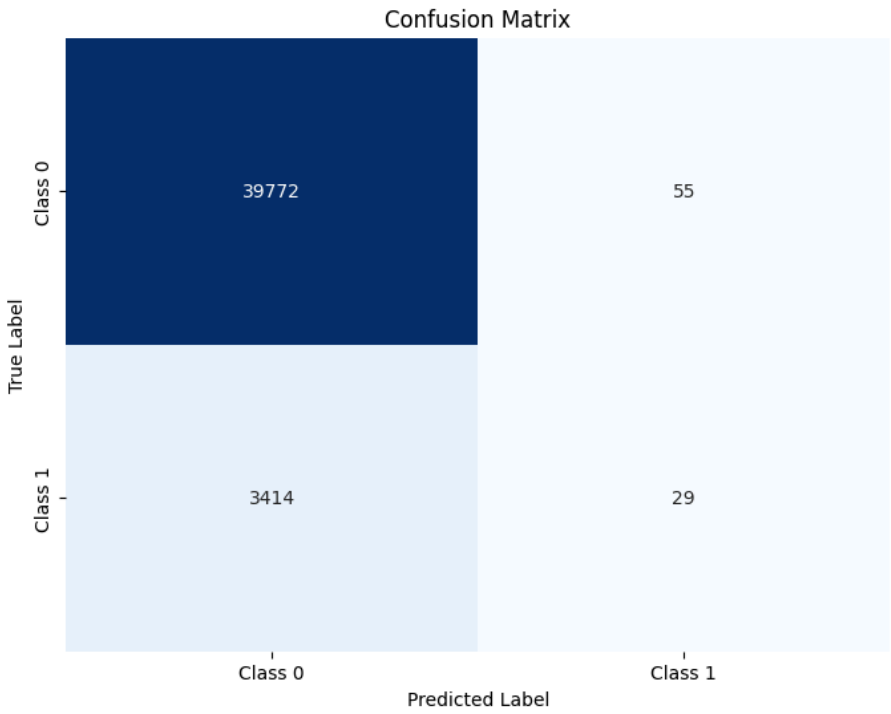


Figure 2 - Best performance in Logistic Regression (undersampling with one-sided-selection)

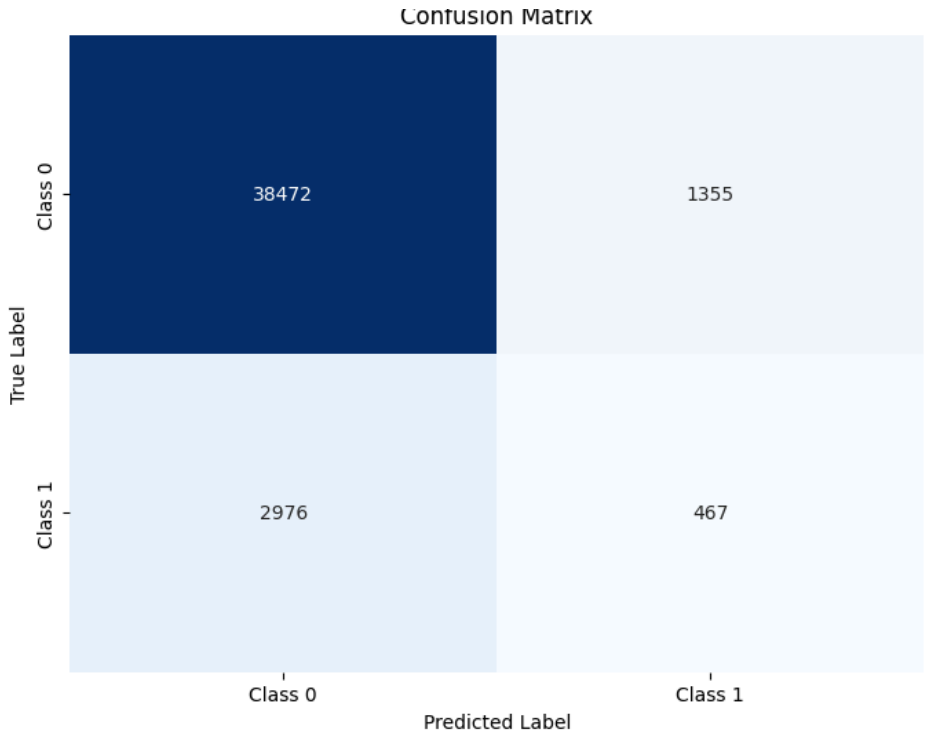


Figure 3 - Best performance in K-NN (undersampling with one-sided-selection)

Term Plan (/ 50 Points)

Outline your work plan for the second term of the project. Do you have any changes, worth mentioning, related to the time management of the project? Clearly show who is working on what and the personal responsibilities. Are there any changes in the workload distribution?

As we embarked on our credit card fraud detection project, our initial plan revolved around implementing the five basic machine learning models commonly used in classification tasks: Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest Classifier, and Decision Tree Classifier. However, as we delved deeper into the project and assessed our progress, we recognized the potential for further exploration and enhancement of our fraud detection methodologies. Also upon realizing that we had more time at our disposal and acknowledging the complexity of the fraud detection challenge, we decided to expand the scope of our project. We identified two additional techniques that could significantly bolster our fraud detection capabilities: anomaly detection and neural networks.

Since every group member has different knowledge levels about the subject and we want to finish one model before moving on to the other everybody works together for the implementation part. There are no changes in the workload distribution.

Term Plan:

1 - Implementing Random Forest Algorithm (WEEK 3 - 4)

During Weeks 3 to 4, our focus will be on implementing the Random Forest algorithm for credit card fraud detection. We will begin by understanding the principles behind Random Forest and its application in fraud detection scenarios. Then, we will proceed with the implementation phase, building and training the Random Forest model using our selected dataset. Through iterative testing and refinement, we will optimize the model parameters to enhance its performance in accurately identifying fraudulent transactions and compare the results with the prior implemented models.

2 - Implementing Decision Tree Classifier (WEEK 4 - 5)

In Weeks 4 to 5, our attention will shift towards implementing the Decision Tree Classifier, another fundamental machine learning algorithm commonly used in classification tasks. We will meticulously craft the decision tree model, considering various splitting criteria and pruning techniques to prevent overfitting. Through rigorous testing and validation, we will fine-tune the Decision Tree Classifier to achieve optimal performance in identifying fraudulent transactions within our dataset. We will again compare the results with the prior implemented models.

3 - Implementing Anomaly Detection (WEEK 6 - 7)

During Weeks 6 to 7, we will venture into the expanded version of our project and start to explore anomaly detection techniques to augment our fraud detection capabilities. Anomaly detection approaches focus on identifying unusual patterns or outliers within the data, which could indicate potentially fraudulent activity. We will explore methodologies such as Isolation Forest and One-Class SVM (Support Vector Machine) for anomaly detection, leveraging their ability to detect deviations from normal behavior. Then decide which model is best for us. The aim in this implementation is to see which action is better: underfitting the dataset or changing the model to predict.

4 - Implementing Neural Network Architecture (WEEK 8 - 9)

In Weeks 8 to 9, we will delve into the realm of neural networks to develop a sophisticated architecture for credit card fraud detection. Neural networks, particularly deep learning models, offer unparalleled flexibility and capacity for learning complex patterns from data. We will design and implement a multi-layer perceptron (MLP) neural network, leveraging its ability to capture nonlinear relationships and hierarchical features within the dataset. Through extensive training and optimization, we will endeavor to harness the full potential of neural networks in detecting fraudulent transactions with high accuracy and efficiency and test if a deep learning model is really necessary or simple models do just fine.

5 - Trying the models with new Dataset (WEEK 10)

In week 10, we aim to preprocess the new dataset and run all implemented models with it from https://github.com/IBM/TabFormer/tree/main/data/credit_card

6 - Comparing the results (WEEK 11)

Week 11 will mark the culmination of our implementation efforts, where we will meticulously compare the performance of the various algorithms and techniques employed throughout the project. We will evaluate the effectiveness of Random Forest, Decision Tree Classifier, Anomaly Detection methods, and Neural Network Architecture in accurately identifying and classifying fraudulent transactions. By analyzing metrics such as accuracy, precision, recall, and F1-score, we will gain valuable insights into the strengths and limitations of each approach, facilitating informed decision-making and further refinement of our fraud detection system.

7 - Finalization (WEEK 12)

In this final week, our focus will shift towards wrapping up loose ends and preparing the end-of-term report. We'll conduct code reviews to ensure cleanliness and efficiency, addressing any lingering issues or bugs. We'll also write the report for end of term.