Project Diary
Data Mining and Machine Learning
F20DL - Group 2

Mohammad Kasra Sartaee

Week 1: Group Formation and Initial Meeting

During this week, we officially formed Group 2 and held our first team meeting. We discussed the overall coursework objectives and responsibilities. We also completed Lab 1, which gave us a hands-on introduction to some of the basic concepts in data mining and machine learning. This activity allowed us to familiarize ourselves with the tools and techniques we would be using throughout the project.

Week 2: Topic and Dataset Selection

In our second meeting, we explored several possible topics and datasets, evaluating their relevance to the coursework objectives and their feasibility for analysis. After some discussion, we finalized our project topic and selected a dataset that aligned with our research goals. We also outlined a plan for the project timeline, assigning tasks to ensure steady progress in the coming weeks.

Week 3: Repository Setup and Data Analysis

This week, we created a GitHub repository to store and manage our project files collaboratively. We finalized our dataset selection and began conducting initial data analysis. This included cleaning the data and performing exploratory data analysis (EDA) to understand the patterns and trends in the dataset. Additionally, we created visualizations to better represent the data, which will inform our next steps in model selection and training.

Week 4: Presentation Preparation

Our focus this week was on preparing for our project pitch presentation. We compiled a slide deck to explain our selected topic, dataset, and initial findings. This included outlining our objectives and presenting some of the visualizations created in Week 3. We rehearsed the presentation to ensure it would be clear and concise, ready for feedback from the instructors.

Week 5: Project Presentation

In Week 5, we delivered our project pitch presentation. Each team member presented different sections of the project, including our chosen topic, dataset, and the exploratory data analysis conducted so far. Feedback from the instructors and peers provided us with valuable insights and suggestions for improvement. We noted down action points to refine our approach and decided to incorporate their feedback into our ongoing analysis and model development.

Week 7: Task Assignment and Dataset Exploration

This week, our team divided tasks for the upcoming modeling work. Kasra and Mary began work on clustering, with the goal to complete this by early next week. Abdul, Prudhvi, and Thafheem started work on decision trees and linear/logistic regression models. We selected the CQI index dataset for these experiments and planned to try clustering on both datasets. We also set a team goal for each member to make at least one GitHub contribution this week, ensuring active involvement and collaborative progress.

Week 8: Clustering and Logistic Regression

This week, we continued our model development with a focus on clustering and logistic regression. Kasra and Mary worked on clustering algorithms, experimenting with different clustering techniques on the CQI index dataset to uncover any potential groupings within the data. Meanwhile, Abdul, Prudhvi, and Thafheem applied logistic regression to test the model's predictive power. The combined efforts helped us further our analysis and paved the way for evaluating other models in the following weeks.

Week 9: Decision Trees, Downloader Scripts, and Additional Clustering Methods

In Week 9, we made substantial progress in multiple areas. Abdul, Prudhvi, and Thafheem continued refining our decision tree models, exploring various parameter settings to improve accuracy. Additionally, downloader scripts were added to automate the acquisition of additional data sources, streamlining our workflow and data retrieval. Kasra and Mary expanded the clustering analysis by incorporating additional clustering methods, allowing us to compare different approaches and improve our understanding of the dataset's underlying structure.

Week 10: Implementation of CNN and MLP

This week, we implemented Convolutional Neural Networks (CNN) and Multilayer Perceptrons (MLP) to explore more advanced modeling techniques. The CNN was used to analyze data with spatial features, while the MLP was applied for general classification tasks. Both models were fine-tuned and evaluated against our dataset to ensure optimal performance. This allowed us to diversify our approach and gather insights from deep learning techniques.

Week 11: Final Compilation, Reporting, and Submission

In the final week, we compiled all results from our analyses and experiments. The team collaborated on writing the final report, ensuring it was comprehensive and well-structured. We also finalized the GitHub repository, organizing code, documentation, and results to create a polished submission package. Finally, we submitted our project files and report, marking the successful completion of our coursework project.