Individual Learning Journal

Muskan Patel

With the Aviation Data Analysis and Fatal Accident Prediction project, my experience working on this project has been both enriching and enlightening. The primary goal of our project was to predict fatal accidents in aviation using machine learning algorithms. We utilized the Aviation Accident Database and implemented three different machine learning models: linear regression, random forest regression, and XGBoost regression. My main responsibilities included the selection and implementation of the machine learning algorithms, as well as model training, evaluation and explanatory data analysis representation. Throughout the project, I gained a deep understanding of the algorithms we employed. Linear regression provided a simple yet effective baseline model for our prediction task. Random forest regression and XGBoost regression, both more advanced models, allowed us to capture more complex relationships in the data. I learned the importance of choosing appropriate algorithms that not only suit the nature of the data but also offer the potential for improved predictive performance. Collaborating with my group partner Athulya, who was responsible for data pre-processing and feature engineering, was a valuable learning experience. Athulya's proficiency in handling missing values, removing outliers, and transforming the dataset into a format suitable for our models was truly impressive. Their work in creating new features and selecting the most relevant variables for our models played a crucial role in the success of our project. Their expertise in feature engineering and data pre-processing greatly improved the quality of our input data and subsequently, the performance of our models.

During the model training phase, I had the opportunity to explore different hyperparameter settings and optimization techniques to fine-tune the performance of our models. This experience taught me the importance of model tuning in achieving the best possible results in any machine learning project. Furthermore, I learned to evaluate our models using the Mean Absolute Error (MAE) score, which provided a clear and interpretable metric for comparing the performance of our models. In terms of my contribution to the project, I am proud of my work in selecting, implementing, and fine-tuning the machine learning algorithms. My efforts in model training and evaluation allowed us to identify the most effective model for predicting fatal accidents in aviation. I also appreciate Athulya's contributions, particularly their dedication to data quality and feature extraction. Their meticulous work in pre-processing the data and engineering meaningful features significantly impacted the success of our project. Additionally, I gained experience using Python libraries such as scikit-learn, matplotlib, seaborn, and plotly to implement our models and create visually appealing visualizations. These tools not only enhanced our understanding of the results but also improved the presentation of our project in the form of an 8-page report in IEEE format.