Mini Project - Individual Contribution

Perera M.D.B. - 210465P

The task at hand involved participating in a Kaggle competition focused on Credit Default Prediction. The objective was to develop a robust predictive model capable of accurately identifying clients likely to default on their loans. This task not only demanded proficiency in data preprocessing and model training but also necessitated innovative approaches to memory optimization for handling large datasets efficiently.

Personal Contribution:

As an individual contributor to the project, my primary focus was on memory optimization and ensuring efficient data handling throughout the model development process. My contributions can be summarized as follows:

- **Utilization of Polars LazyFrames**: Recognizing the memory-intensive nature of the task, I adopted Polars LazyFrames as a memory-efficient alternative to Pandas DataFrames. This allowed for the manipulation and transformation of large datasets while minimizing memory footprint.
- **Dynamic Memory Management**: I implemented dynamic memory management techniques to optimize memory usage during various stages of the project. This involved actively monitoring memory consumption and deallocating resources when necessary, thereby preventing memory leaks and ensuring optimal performance.
- **Datatype Optimization Function**: To further enhance memory efficiency, I devised a custom function to optimize the datatype of columns within the dataset. By selecting the most appropriate datatype for each column, I effectively reduced memory overhead without compromising data integrity or computational accuracy.
- **Memory Footprint Reduction**: Through diligent memory management practices and datatype optimization strategies, I successfully minimized the memory footprint of the entire training dataset. As a result, the dataset, comprising extensive records and features, was loaded into memory space totaling just 2659.8065 MB, demonstrating significant memory savings and enabling smoother execution of subsequent tasks.

In summary, my contributions focused on implementing memory optimization techniques, including the adoption of Polars LazyFrames, dynamic memory management, datatype optimization, and memory footprint reduction. These efforts were instrumental in ensuring efficient handling of large datasets and facilitating the development of a robust predictive model for Credit Default Prediction.