Revised Plan:

Prepare the Dataset

Our first goal is to reduce the dataset size to avoid model overfitting. Since our target is to train the model to achieve at least 51% accuracy, we do not need a dataset of nearly a million games. A maximum of 50,000 games should suffice. This step includes cleaning and preparing the dataset for training. Time required: 3 hours per student.

Find the Best Algorithms

This step involves researching and selecting the most suitable machine learning algorithms for our model. We will test these algorithms on a smaller dataset to determine their viability. Time required: 4 hours per student.

Train the Model

Once the dataset is ready and algorithms are selected, we will train the model. This involves applying the algorithms to the dataset and iterating until we achieve stable initial results. Time required: 10 hours per student.

Test the Model

As the model begins to provide results, it will undergo rigorous testing to optimize its performance. We will iterate through training and testing phases until we meet our business goals. Time required: 12 hours per student.

Parallel Reporting

While training and testing the model, we will document progress, challenges, and insights. These reports will help guide further iterations and improve subsequent stages. Time required: 4 hours per student.

Prepare for Presentation

We will create a poster or slides summarizing our research, methodology, and results. This step includes designing visual aids and structuring the presentation content. Time required: 4 hours per student.

Present the Project

Both students will deliver the project presentation at the Delta building. This includes rehearsing and actual delivery of the presentation. Time required: 3 hours per student.

Methods and Tools:

Dataset Preparation: Python, Pandas, NumPy.

Algorithm Selection: Scikit-learn, research papers, online resources.

Model Training and Testing: Python, TensorFlow/Keras, Scikit-learn.