IMPLEMENTATION IDEA FOR SOLVING STRUCTURED DATA ASSIGNMENT

The implementation idea I used to solve the structured data assignment are as follows:

- **Data loading:** The input train and test files are shared as parquet files. These files are loaded by python-pandas package.
- **Data preprocessing:** The data contains historical data of 27k unique patients, out of which 9k patients are treated with the target drug. This problem can be considered as Random Forest Classifier.
- **Modeling:** Once the data has been preprocessed, a classification model can be applied to predict whether a patient will be treated with the target drug.
- **Evaluation:** The performance of the classification model can be evaluated using metrics such as accuracy, precision, and recall.

The challenges I faced to solve the structured data assignment are as follows:

- **Data quality:** The data quality was not perfect. There were some missing values. This required some data cleaning and pre-processing steps to be taken.
- **Model selection:** There are many different classification models that could have been used for this task. The choice of model can have a significant impact on the performance of the model.
- **Model tuning:** The hyperparameters of the model need to be tuned to get the best performance. This can be a time-consuming process.

With given time and resources, I am positive that a model can be developed for this problem.

REFERENCES:

1. https://medium.com/geekculture/drug-target-interaction-prediction-through-python-4af9e76fc90
2. https://academic.oup.com/bib/article/22/1/247/5681786