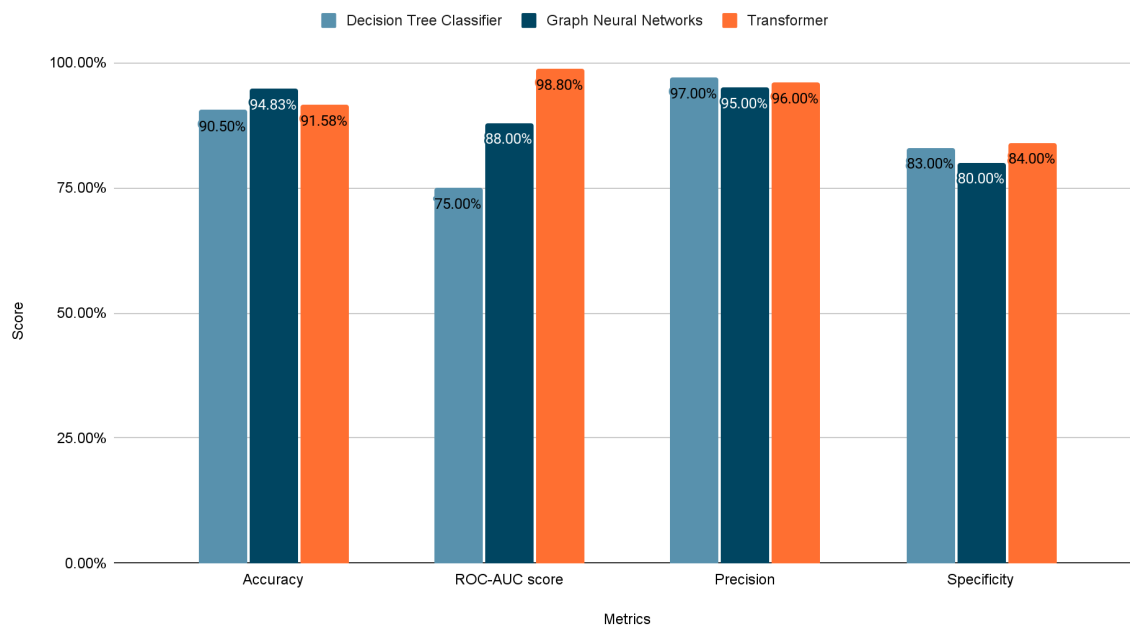


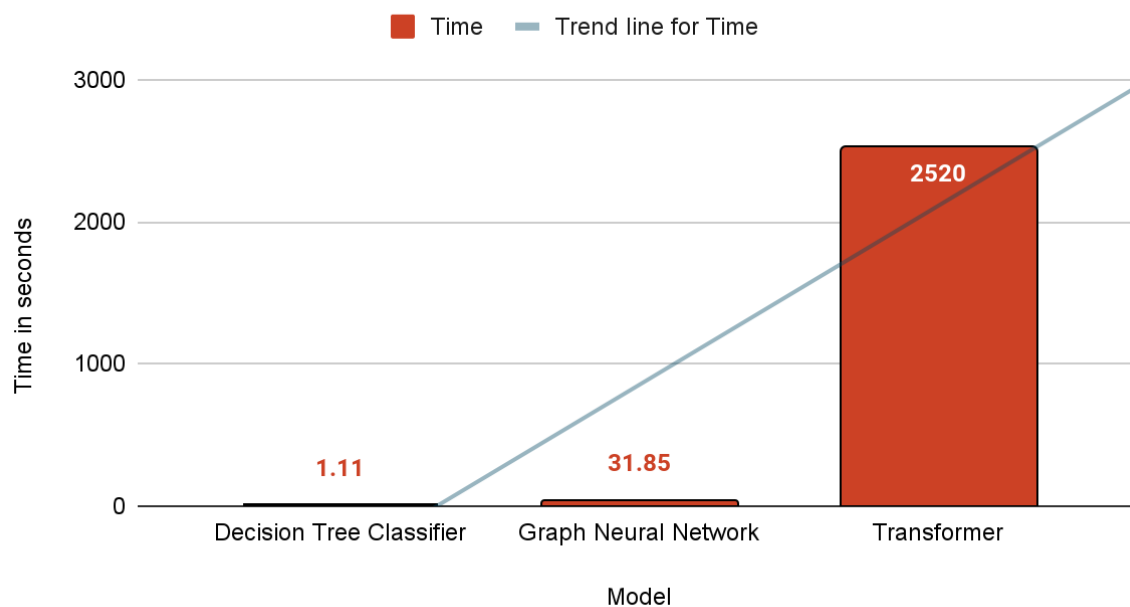
Final Insights

The benchmarks chart shows all the metrics considered for evaluating the model, while the Time vs Model chart evaluates the time taken to train and predict the outputs.

Benchmarks



Time vs Model



Out of all the models, Graph neural networks looks like the best model to predict molecular properties/ADME properties. This is because of a few things, the first being highest accuracy

and a decent score for ROC-AUC. The more closer to 1 ROC-AUC score the better the model, and graph neural networks has a score of 0.88. Although transformers model gives a better score for ROC-AUC, the time taken to complete the model is approximately 43 minutes, compared to GNN model, it is a massive increase in time, therefore would not be optimal. Although decision tree has the best model times compared to both the other models(GNN, transformers), comparing ROC-AUC score is the proven method to finding a good model, as it considers the trade-off between precision and recall.

Key Metrics For Sales Team

HIA (Human Intestinal Absorption), Hou et al.

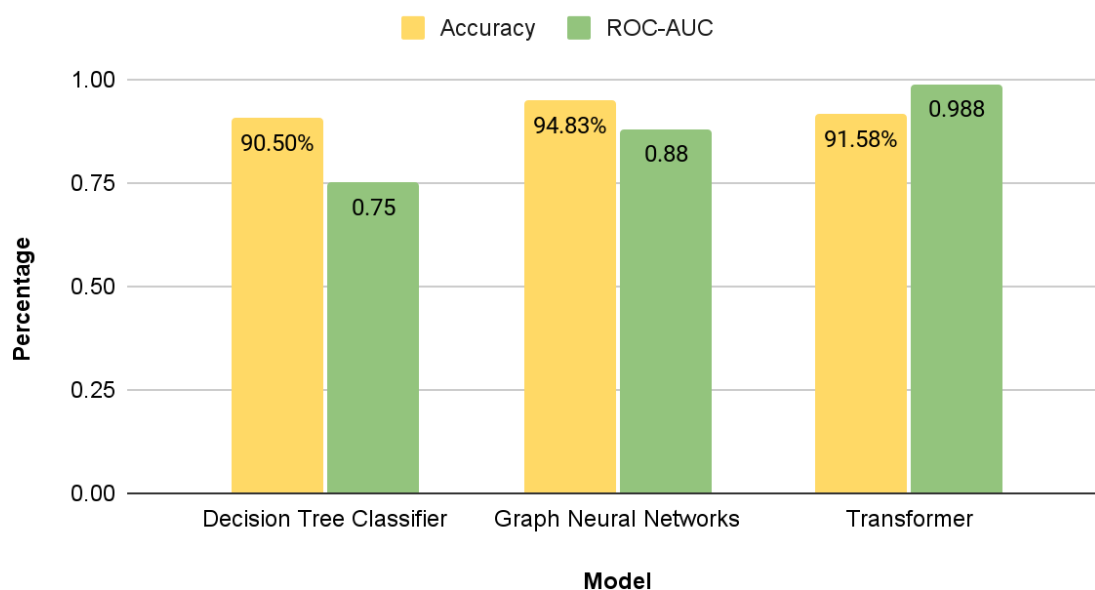
Dataset Description: When a drug is orally administered, it needs to be absorbed from the human gastrointestinal system into the bloodstream of the human body. This ability of absorption is called human intestinal absorption (HIA) and it is crucial for a drug to be delivered to the target.

Task Description: Binary classification. Given a drug SMILES string, predict the activity of HIA.

Dataset Statistics: 578 drugs.

HIA dataset was chosen because it is the most common method for humans to take drugs, which is orally, including tablets and syrups, and the absorption happens through the gastrointestinal system.

Key Metrics



Similarly the remaining properties, distribution, metabolism and excretion properties of a drug can be predicted.

- Accuracy shows how accurately the model predicts the output, in this case the HIA of a drug - Decision tree has the lowest accuracy and low ROC-AUC score
- Transformer model has the best ROC-AUC score, around 0.98 and decent accuracy but takes unoptimal amount of time
- GNN is the best model overall with the best metrics considering accuracy and ROC-AUC, the main trade-off made here is time, and ROC-AUC score with a threshold of .80