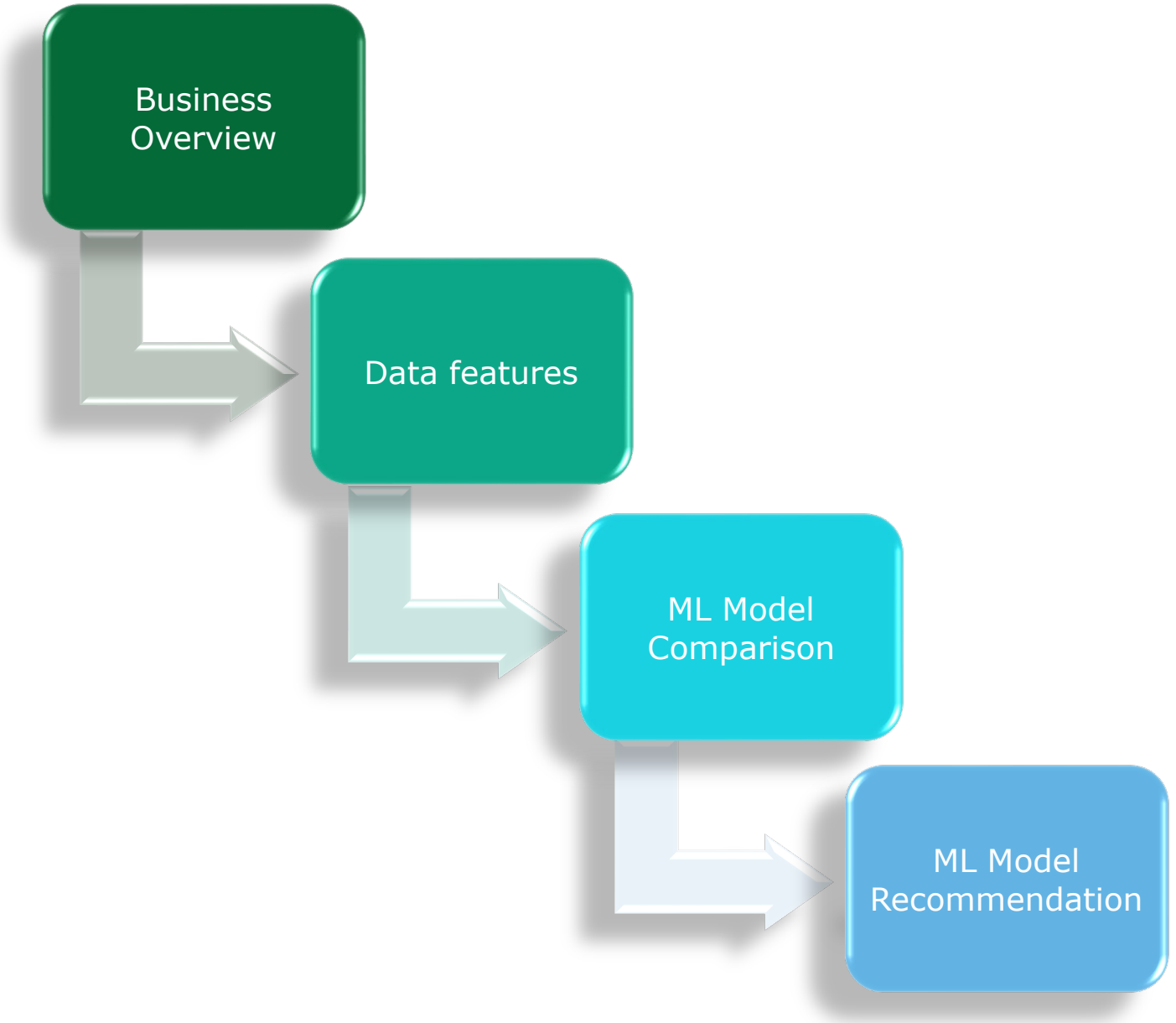
Deloitte.

Molecular Solubility Prediction

Project Scope



Business Overview and Goals

Business Need

- Production costs due to slow research
- Time for medication development
- Automating molecular solubility predictions

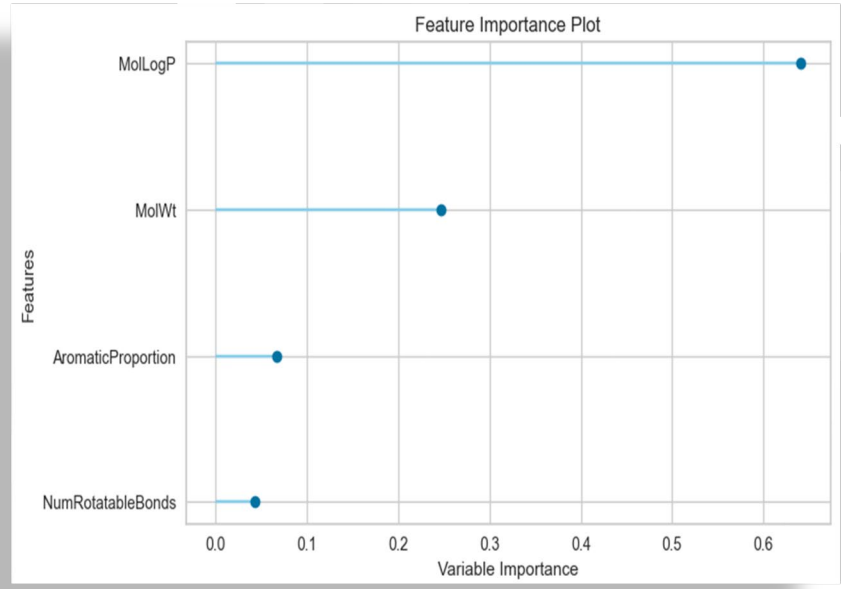
Goals

- Identify Key Molecular/Compound Features
- Deliver Accurate ML Model
- Ability for business to Accurately predict molecular solubility

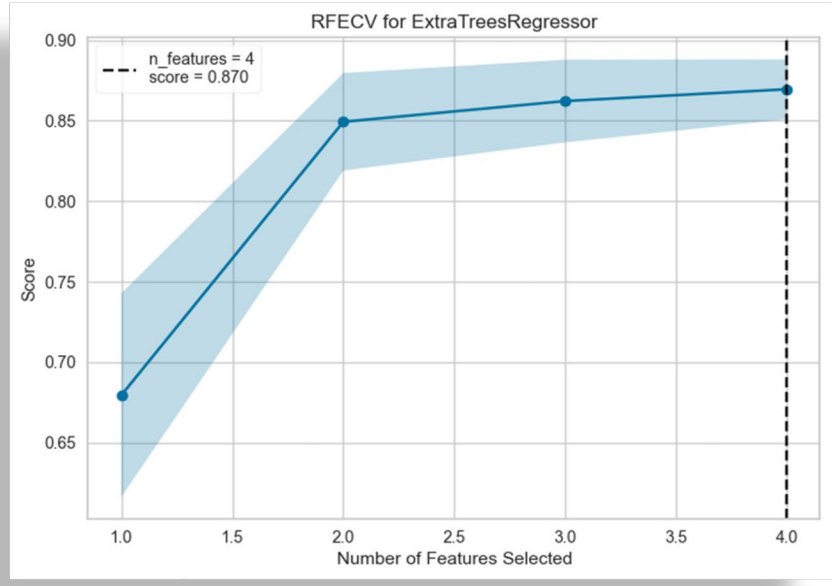
Up Front Suggestion

- Extra Trees Regressor Model provided the best results

Model Feature Analysis & Plotting



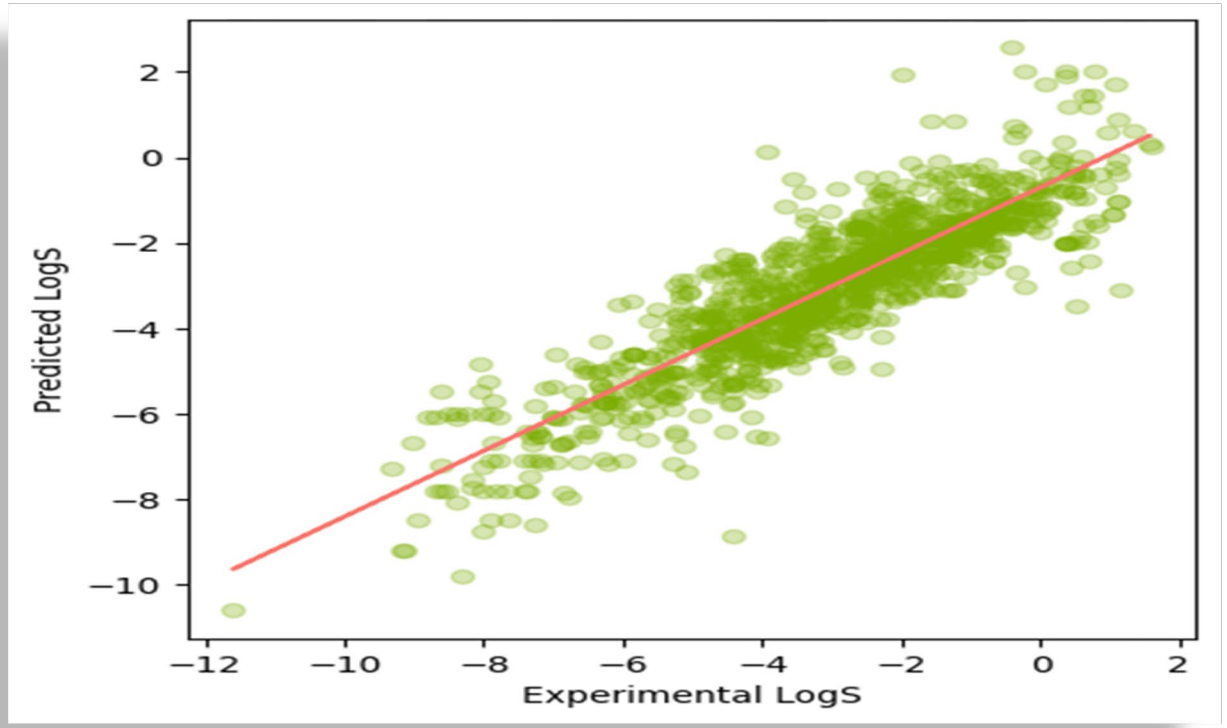
- MolLogP and Molecular Weight are the two variables with the most importance.
- The importance of these two variables impacts the score of the model



- Out of the 4 molecular descriptors, using only 2 features could provide an excess of 0.85 for the R^2.
- The use of the remaining two descriptors, will slightly improve the performance of the prediction.

Linear Regression Model

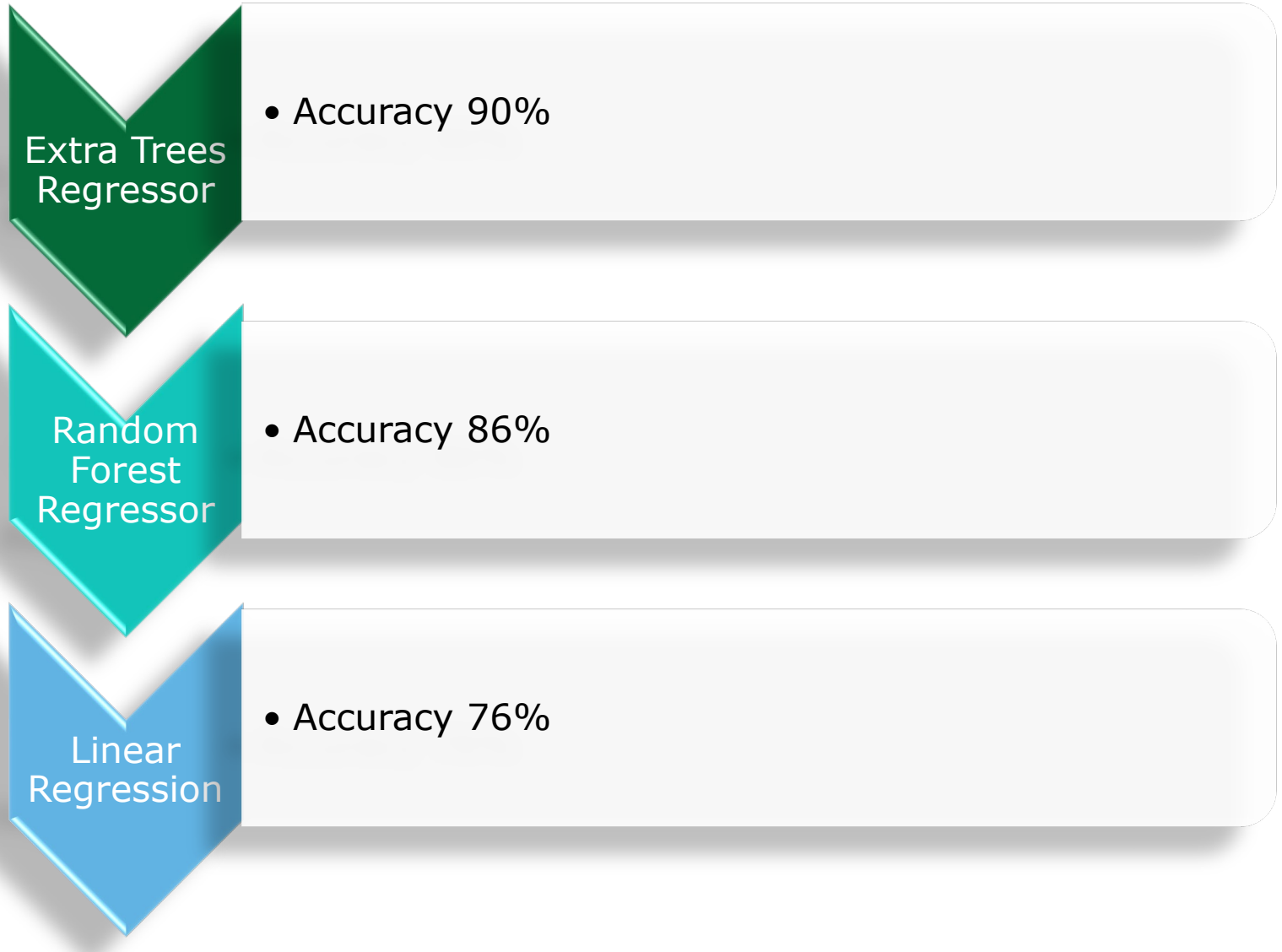
The Linear Regression model is used as the main prediction model to be compared to the other models.



Results

- R^2: 0.76
- MSE: 1.01
- RMSE: 1.0112

Model Results & Recommendation



Questions, Comments, Concerns?

Leo Muntaner

Solution Analyst

Lmuntaner@deloitte.com

Deloitte – Orlando GPS Strategy and Analytics