Project Summary

1. Data Processing

- Loaded the dataset and handled missing values using IQR and Z score.
- Standardized and normalized numerical features using min max scaler.
- Performed exploratory data analysis (EDA) to understand distributions.

2. Feature Selection

- Used SHAP values to identify the most important features.
- Selected the top features for training the model.

3. Model Training

- Trained a Neural Network (NN) model for prediction.
- Used a train-test split to evaluate model performance.
- Implemented Random Forest as a baseline model.
- Model was trained on 10,000 epochs.

4. Model Evaluation

- Compared actual vs. predicted values.
- Used metrics like RMSE, MAE, and R^2 for performance assessment.
- Visualized predictions using scatter and residual plots.
- Model was performing decently; however, it is still not able to predict many samples accurately.
- Ensembling methods or a deeper understanding of data might help improve performance.

5. Output and Deployment

- Stored the trained model for future use.
- Documented the process and results for reproducibility.