Project Development Phase
Model Performance Test

Date	9 November 2023
Team ID	PNT2023TMID591623
Project Name	Project - Walmart store sales forecasting
Maximum Marks	10 Marks

Model Performance Testing:

Project team shall fill the following information in the model performance testing template.

S.No.	Parameter	Values	Screenshot

1.	Metrics	**Regression Model:** MAE - , MSE - , RMSE - , R2 score -	LightGBM regression:
			
			 `[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was 0.067988 seconds.`
			 `You can set \`force_col_wise=true\` to remove the overhead.`
			 `[LightGBM] [Info] Total Bins 2668`
			 `[LightGBM] [Info] Number of data points in the train set: 301325, number of used features: 23`
			 `[LightGBM] [Info] Start training from score 0.179686`
			 `R-squared (R2) Value: 0.955149041617666`
			 `Mean Squared Error: 0.0020249160128211558`
			 `Mean Absolute Error: 0.0246452659355332106`
			 `Explained Variance: 0.955149091340892`
			
			 Linear regression:
			
			 `MAE 0.0246452659355532106`
			 `MSE 0.0020249160128211558`
			 `RMSE 0.04499906679945216`
			 `R2 0.955149091340892`
			
			 Extra tree regression:
			
			 `Mean Squared Error (MSE): 0.0008816651590708112`
			 `Mean Absolute Error (MAE): 0.014999600508864733`
			 `Root Mean Squared Error (RMSE): 0.029692846934418586`
			 `R-squared (R2) Value: 0.9803685869647454`
			
			 Adaboost regression:
			
			 `Mean Squared Error (MSE): 0.005677102799603772`
			 `Mean Absolute Error (MAE): 0.04545054862894153`
			 `Root Mean Squared Error (RMSE): 0.07534655134512643`
			 `R-squared (R2) Value: 0.8735919767771264`
			
			 Random Forest regression:
			
			 `MAE 0.015522536897538632`
			 `MSE 0.0009530632336469744`
			 `RMSE 0.030871722233250517`
			 `R2 0.9788909900125646`

			KNN regression:
			MAE 0.033122163743083126 MSE 0.003624289656000884 RMSE 0.060202073519114635 R2 0.9199211034808975
			XG boost regression:
			MAE 0.026771808878560288 MSE 0.0026134394830486384 RMSE 0.051121810248157665 R2 0.9421152350249367
2.	Tune the Model	Hyperparameter Tuning - Validation Method -	Number of trees in random forest regression: ```\n[115] y_pred = rf.predict(X_test)\n\n[118] from sklearn.ensemble import RandomForestRegressor\n n_estimators = 100 # can change this value to your desired number of trees\n rf_model = RandomForestRegressor(n_estimators=n_estimators)\n\n number_of_trees = rf_model.n_estimators\n print(f"Number of trees in the Random Forest: {number_of_trees}")\n print("MAE" , metrics.mean_absolute_error(y_test, y_pred))\n print("MSE" , metrics.mean_squared_error(y_test, y_pred))\n print("RMSE" , np.sqrt(metrics.mean_squared_error(y_test, y_pred)))\n print("R2" , metrics.explained_variance_score(y_test, y_pred))\n\n Number of trees in the Random Forest: 100\n MAE 0.015440129853658986\n MSE 0.00093403475199361\n RMSE 0.03056198213456812\n R2 0.978502531437008\n``` Train and test split validation method: