Ahmedabad University

CSE523 Machine Learning

Weekly Project Report

Date: 25-03-2023

Project title: Big Mart Sales Prediction

Group 10

Name	Enrolment no.
Meet Patel	AU2040010
Dev Patel	AU2040056
Kush Patel	AU2040137
Vatsal Shah	AU2040019

1. **Task performed and outcomes of task performed this week**
    - We implemented the Random Forest regression model.
    - We computed the accuracy of the models and compared the results.
    - The linear regression gave accuracy of 0.51 on training set and 0.49 on test set.
    - The accuracy of XGBoost was 0.86 on training set and 0.54 on test set.
    - The accuracy of Random forest model in 0.93 on training set and 0.55 on test set.
    -

```
[57] from sklearn.ensemble import RandomForestRegressor
     rf= RandomForestRegressor()

     rf.fit(X_train,Y_train)

     ▾ RandomForestRegressor
     RandomForestRegressor()

[60] # prediction on training data
     training_data_prediction_RF = rf.predict(X_train)

[61] # R squared Value
     r2_train_RF = metrics.r2_score(Y_train, training_data_prediction_RF)

     # Mean Absolute Error
     MAE_train_RF = metrics.mean_absolute_error(Y_train, training_data_prediction_RF)

     # Mean Squared Error
     MSE_train_RF = metrics.mean_squared_error(Y_train, training_data_prediction_RF)

[62] # XGBoost regression model performance on training set
     print('R Squared value = ', r2_train_RF)
     print("Mean Absolute Error = ", MAE_train_RF)
     print("Mean Squared Error = ", MSE_train_RF)

     R Squared value =  0.9386572019592816
     Mean Absolute Error =  290.9191635811088
     Mean Squared Error =  175871.55436649767
```

```
[63] # prediction on test data
     test_data_prediction_RF = rf.predict(X_test)

[64] # R squared Value
     r2_test_RF = metrics.r2_score(Y_test, test_data_prediction_RF)

     # Mean Absolute Error
     MAE_test_RF = metrics.mean_absolute_error(Y_test, test_data_prediction_RF)

     # Mean Squared Error
     MSE_test_RF = metrics.mean_squared_error(Y_test, test_data_prediction_RF)

[65] # XGBoost regression model performance on test set
     print('R Squared value = ', r2_test_RF)
     print("Mean Absolute Error = ", MAE_test_RF)
     print("Mean Squared Eror = ", MSE_test_RF)

     R Squared value =  0.5504677830337481
     Mean Absolute Error =  825.1441216117303
     Mean Squared Eror =  1387739.5385357765
```

- From the accuracy of XGboost and Random Forest on the training and test sets, it can be concluded that the model overfits the data and we need to do hyperparameter tuning to avoid such problems.

2. **Tasks to be performed in the upcoming week**
    - We will perform hyperparameter tuning to improve our models' accuracy.