ALTRON KARABINA

Forecasting Timeseries Model Development Documentation

Document Prepared By: Richmore Dzanza

Contents

Document Version

Document Location	https://github.com/dzrich/Richmore_Project
Filename	Forecasting Timeseries framework.docs
Author	Richmore Dzanza
Reviewed By	To be Reviewed

Version	Developer	Date	Comments
1.0	Richmore Dzanza	28 November 2022	Creation and documentation

1. Introduction

This report encompasses the developmental evidence relating to the forecasting Timeseries Models developed for Altron Karabina. The Cross-Industry Standard Process for Data Mining (CRISP-DM) was followed as a base to develop a prediction model for the forecasting of sales of a superstore furniture shop. The superstore's data is used in order to validate the use of the various machine learning algorithms. Two Machine Learning Algorithms were built, in comparison, which are Random Forest Regression and Decision Tree. The regression Tree performs better than the random forest in this analysis with a Root Mean Squared Error (RMSE) of 15,7.

2. Model Developer

The Data Scientist Responsible for the Development of the Forecasting Timeseries Models is Richmore Dzanza.

3. Data Summary

The furniture store data was extracted from 2014 to 2017 and was used for the development. Identifier variables were dropped out and only correlated variables to the target variable was used for the development. Since all the predictor variables were weakly correlated to the target variable, the developer decided to use all the remaining variables for model training, and some experimentation would have been performed by dropping the variables with lowest correlation values in their order. But this was not done because of time constraint.

The list of variables which was used for the development is fount in the Superstore.ipynb file as provided.

4. Model Building

Both forecasting models were developed using Python Python in a Jupyter Notebook. Various Machine Learning packages were used and the final model was obtained by performing some Hyper Parameter tuning in comparison. Two different Machine Learning models were build which are Decision Tree Regression and Random Forest Regression to predict the furniture sales.

5. Model Evaluation.

Root Mean Squared Error (RMSE) was used to measure the performance of the models. A odel with the lowest RMSE was chosen and the decision tree had the lowest RMSE as compared to the Random Forest.

Model	RMSE
Random Forest Regression	15.7
Decision Tree Regression	28.6