Critic Review on Team 1
Predict The Future Sales using Machine Learning

Provided by Team 8
Naveena Kanderi
Sandeep Narayanasetty
Vaishnavi Valluri
Jaswanth Reddy Baggari

December 12, 2022

This evaluation report is for Team 1's Capstone Project project, "Predict the Future Sales Using Machine Learning." Our understanding of their presentation is that they want to examine and contrast two machine learning methods termed Random Forest (RF) and Decision Tree that could be used for prediction (DT). They developed two machine learning techniques and tested the effectiveness of the models on the training dataset utilized in this study. Based on performance accuracy, the best algorithm is chosen for the prediction.

Future events are what prediction is concerned with. Machine learning algorithms boost the system's intelligence without involving humans. Machine learning (ML) is used to optimize the performance criterion using test data or prior knowledge. They created a model that is exact, reliable, and accurate and can forecast product sales.

The sales of a company like 1C Company are predicted using machine learning techniques, and they selected to utilize Random Forest algorithms and Decision tree algorithms and compare their degrees of accuracy.

Machine learning and deep learning were utilized in this work to analyze the data and extract significant information from patterns and trends. Data preprocessing is used to transform the unstructured data into understandable information. For this, it is necessary to comprehend the data, deal with missing data, and get rid of duplicates. They have implemented all this steps to get a dataset that is proper and complete.

They trained the sales data using these two algorithms by employing a training technique to identify the relationship between the data. Data from the database that is appropriate for analysis is picked through feature selection.

outlier detection is used for model optimization. The technique for spotting outliers can be used to deploy the model, launch more iterations, and be beneficial for displaying more general data unrelated to the models. The main issue is the quality of the data, especially the quality of each data attribute.

These also consider deleting low-value data attributes. Although there were many entries in the initial dataset, the final dataset was only marginally less than the initial dataset after removing worthless, redundant, and unneeded data.

Random forest and decision tree algorithms are utilized to examine the historical sales data utilizing the two machine learning techniques. Following the application of the random forest regression technique to the sales dataset, we discovered training and validation losses of 79% and 91%, respectively. Decision tree regression is applied, yielding training loss of 80% and validation loss of 92%. Both models' training and validation RMSE values are computed and contrasted.

The main aspect we identified is that they could have implemented other algorithms such as LSTM and ARIMA which mostly deals with the time series data. These algorithms might provide them better results as they are highly efficient for data that has time series.

They could have also included details such as the highest sold item, least sold item and stores in which highest products are sold, least products are sold during the data cleaning phase.

For analysing the performance they have used validation loss and training losses. They could have used other metrics such as confusion matrix, f-score or classification reports.

They have presented the Project slides and descried all the aspects succulently and clearly which is very commendable. It should also be noted that their data has a foreign language i.e Russian. so, they cleverly omitted the description of the item product as its not relevant to the prediction, making it more easily.