



# Predict Future Sales

Aaryaman, Adithya, Anshul, Arun

Mr. Nithin Prince John

## Abstract

Predicting future sales is one of the most important aspect of the business world, Trillions of dollars are transacted every single day all over the world. Since a lot of money is involved, we need systems to get maximum profit from products sales. Machine Learning models are helpful in predicting future sales to get maximum profits. Multiple models are applied among which, we choose the one which gives the lowest RMSE Score. *Random Forest* was the one among many which gave the lowest RMSE value. LSTM is a deep learning model which was also implemented but RMSE score was poor. The dataset is relatively smaller for neural network to perform well.

## Introduction

This era of finance world is still growing because every day new technology is emerging out which gives better accuracy. Since reducing loss is the main aim in finance i.e. max profit sales, everyday new algorithm and model are implemented which improves the overall prediction.

Common challenge we face are the missing values and outliers which are taken care of by different regularization technique. There are many Machine / Deep Learning models which can be applied here based upon the problem statement. We applied Random Forest Regressor, LSTM, XGBoost, neural network out of which random forest regressor outshined them all with lowest RMSE score of 0.901.

XGBoost was second best with the RMSE score of 0.921 then neural network with 0.964 and LSTM with 1.03.

Starting with neural network, training it with the train dataset took a lot of time for a small dataset containing few thousand records, as time is a crucial factor, we just can't ignore it. If we scale it up to million shops in few countries, it won't be feasible.

Since data is non-sequential, we did transformation of dataset and later implanted LSTM on dataset. After running on test dataset, LSTM was not giving lower RMSE score hence the model was dropped. XGBoost is a powerful model used with different sets of hyperparameters. Altering hyperparameters was not significant as RMSE score had reached a saturation value. Random Forest regressor gave the lowest RMSE score. We altered the parameters and have specified the best in methodology.

## Conclusions

It was found under rigorous testing that Random forest regressor outshined all the other machine learning modes with a RMSE score of 0.901. Since the dataset was small for neural network to be trained and it couldn't give accurate predictions. Further improvement can be made by altering the hyperparameters. We learned that feature engineering is one of the most important aspect of improving model accuracy.

## References

- <https://www.kaggle.com/c/competitive-data-science-predict-future-sales/rules>
- <http://papers.nips.cc/paper/5955-convolutional-lstm-network-a-machine-learning-approach-for-precipitation->