

Store Sales Forecast with Machine Learning

Suhas Anand Balagar* Hardy Leung† Loukya Tammineni‡ Xichang Yu§

November 2021

Abstract

Pizza [1] is an understudied yet widely utilized implement for delivering in-vivo *Solanum lycopersicum* based liquid mediums in a variety of next-generation mastications studies. Here we describe a de novo approach for large scale *T. aestivum* assemblies based on protein folding that drastically reduces the generation time of the mutation rate.

Introduction

The ability to predict the sales of a variety of stores is highly sought out in supply chain logistics, as it finds applications in increasing customer satisfaction and reducing food waste. We are proposing use of multiple Supervised Learning methods to predict the sales of stores based on time series dataset of Corporación Favorita, an Ecuador based grocery retailer. Ecuador is a country whose economy is strongly dependent on the oil and fluctuates with the price of oil.

We are planning to use dataset from an ongoing Kaggle competition [2], “Store Sales – Time Series Forecasting”. The dataset includes multiple csv sheets of time series data. We will try to evaluate the different aspects that might impact the sales in a store like Holiday seasons, Oil prices and historical sales data from a variety of stores. The preprocessing of data will include checking for missing values and if found, imputing them so that no data is disregarded. We have also checked for frequency distribution of data

elements as part of Data exploration. We have calculated and plotted correlation mapping to establish the correlations between different input and output parameters such as holiday events, oil prices, transactions per store type, dates of salary, natural calamities etc.

We plan on using different Supervised Learning models to predict the prices and then evaluate which of those models give out the best results. Supervised learning approach works best in this case, as we have huge time series data of both input and the output parameters mentioned above. Few of the supervised learning options we can consider are

- Linear Regression
- Gradient Boost(XGBoost)
- Random Forest
- Support Vector Machine(SVM) /Support Vector Regression (SVR)
- Long Short-Term Memory (LSTM)

Once the models are trained on training data and developed, we can evaluate the performance of these models on the test dataset.

Methods

Here we talk about the methods.

Models

Here we talk about the comparison.

*San José State University, suhasAB@github

†San José State University, ksleung@github

‡San José State University, LoukyaTammineni@github

§San José State University, Codyyu36@github

Analysis

Here we talk about the analysis.

Conclusions

Here we talk about the conclusion.

References

- [1] M. Pizza *et al.*, “Identification of vaccine candidates against serogroup b meningococcus by whole-genome sequencing,” *Science*, vol. 287, no. 5459, pp. 1816–1820, 2000.
- [2] “Store sales - time series forecasting,” *Kaggle*. [Online]. Available: <https://www.kaggle.com/competitions/store-sales-time-series-forecasting>.