TIME-SERIES FORECASTING IN RETAIL SALES OF PHARMACEUTICAL PRODUCTS BASED ON DAY TEMPERATURE AND HOLIDAYS

Anuradha Nanayakkara

A dissertation submitted in partial fulfilment of the requirement for
Master of Science (Honors) degree in Big Data Analytics

Department of Computing

Informatics Institute of Technology, Sri Lanka
in collaboration with
Robert Gordon University, Scotland

2021

Abstract

This thesis joins the vivid details about pharmaceutical sales predictions using predictions models, feature selections along with their respective analysis. Collectively, existing thesis works from different researchers measures the accuracy using different algorithms in different domains in terms of identifying ways to maximize the sales using machine-learning processes. Distinctively, the process of finding the accurate model by understanding the features that relatively affect for the accurate model will be discussed throughout this paper.

This paper consists of factual pharmaceutical data in the area of Saskatchewan Canada, which applied to the models discussed in the latter part of the study. As the methodology for a successive implementation of accurate prediction
Comparing the model which were already implemented with time series analysis for the retails, interpret their results with the previous studies and building the multivariate time series model doing the feature engineering. External factors were taken into consideration for the pharmaceutical sales while finding their effectiveness and feature importance. Moreover, to analyse the correlation of the unit sales with other features and provide the meaningful graph to visualize the data for the end user is an additional important objective of the research.

Mainly this project concentrated on three time series models XGBoost, Regression, ARIMA ,LSTM , to experiment pharmaceutical retail analysis. Analysing output result of XGBoost, Regression model performance were thrived comparing to other models by achieving 13% of MAPE while others achieved 22% and 53% of MAPE respectively. This model has strength to predict the nearly 14 days of the future values, hence this would be a beneficial for the decision makers to organizing their financial plan and positioning the marketing campaigns effectively.