MACHINE LEARNING BASED SALES FORECASTING SYSTEM

Pathum R. Ranathunga

A dissertation submitted in partial fulfilment of the requirement for

Master of Science in Business Analytics

Department of Computing

Informatics Institute of Technology, Sri Lanka

In collaboration with

Robert Gordon University, Aberdeen, UK

2022

Abstract

Sales forecasting aims to predict demand for sales figures in the future and reserve the number of products and perform marketing strategies based on the forecasting results. An accurate and reliable forecasting system can be seen in sales demand patterns and avoid unnecessary overstocking, and maintenance costs and also impact a major role in decision-making operations in the areas corresponding to sales, production, purchasing, finance, and accounting. Many factors can impact the sales forecasting results. But researchers have only taken a few facts while their research. In this project, internal factors and external factors are analyzed, namely as; temperature, fuel prices, holidays, Consumer Price Index, Employment rate, and discount strategies, that can be assumed directly affect consumer sales demand in supermarkets and their departments, all related research done in the domain along with the inputs of experts in the field. Past research papers and publications are used to identify suitable methodologies and machine learning algorithms, and then based on the findings, an experiment process is initialized to evaluate the performances of machine learning algorithms. Also, a unique forecasting solution is proposed based on those major factors, which has been developed as an accurate machine learning-based sales forecasting system for regional supermarkets and the departments in Walmart USA supermarkets to fill the gaps in existing solutions.

Findings from literature reviews claim that different regression algorithm models such as Simple Linear Regression, Support Vector Machine Regression, Ridge Regression, Gradient Boosting Regression, Random Forest Regression, XGBoost Regression, Long Short Term Memory, and, ARIMA Time series forecasting in Python are suitable algorithms, and outcomes from the experiment, Extreme Gradient Boost Regression is performing good accurate than other machine learning algorithms. Since the chosen dataset is labeled set, supervised learning is the best fit method for machine learning.

Based on the results, it can be concluded that studies on the influence of external and internal factors over customer demand can be used to forecast the accuracy of sales, which can make a remarkable difference in profit, expenditures, and the stability of businesses.

Keywords: Sales forecasting, Time series forecasting, machine-learning