MSCI 718 – Statistics for Data Analytics

Project Proposal

Weekly sales prediction using ARIMA and Linear Regression

Abstract

Modeling retail data to make decisions is of great importance in modern competitive retail world. Accurate predictions are very important to handle big events such as Black Friday, Thanksgiving, and Christmas. Not only that, daily/weekly sales prediction can have huge impact on the smooth operation of the day to day business. Many critical decisions such as manpower arrangement, inventory management etc. lies upon such prediction and can affect the overall sales and revenue of the business. In this project, an approach will be made to predict the weekly sales of one of the largest retailer in the world "Walmart".

Overview

Data presented in form of time series as its analysis and applications recently have become increasingly important in different areas and domains. Predicting the sales of any business is one of them. Big fishes in the business industry always try to have prior knowledge about their sales and demand of the products. For this project a data set will be will be used which consists data of 45 different stores located in different regions and 10 different features to analyze and predict the department wise weekly sales of Walmart. The features of the data set are temperature, fuel price, consumer price index, unemployment etc. In addition, Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labor Day, Thanksgiving, and Christmas. The historical data set has two parts training component and test component. We will train our data on the training data set to build time series prediction models later the models will be tested to measure the efficiency on test data set. Both ARIMA and Linear Regression model will be implemented on the dataset to find out the better one for prediction.

Methods

The goal is to build a robust model that will be able to withstand uncertainty. To achieve that following methods will be used.

1. Auto-regressive Integrated Moving Average (ARIMA)
2. Linear Regression

Tools

Following tools will be used to achieve our goal.

1. Python
2. R

Data Source

https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data

Project Proposal

Evaluation

The result will be evaluated based on the weighted mean absolute error (WMAE):

$$WMAE = \frac{1}{\sum w_i} \sum_{i=1}^{n} w_i \, |y_i - \hat{y}_i|$$

Where

- n is the number of rows
- $\hat{y}_i$ is the predicted sales
- y_i is the actual sales
- w_i are weights. $w = 5$ if the week is a holiday week, 1 otherwise.

Group Members

1. Samina Islam Eva- 20704504
2. Souradip Dey- 20741206
3. Md Rubayatur Bhuyian- 20743420