

Stevens Institute of Technology  
MIS637- Fall 2020  
Initial Project Proposal  
**Predicting Walmart Sales Using Random Forest Regressor**

By- Soham Sanjay Shinde

CWID- 10452638

**A. About Walmart:**

Walmart is multinational departmental store which operates a chain of hypermarket, discount department stores and grocery stores. The company was founded in 1962 and incorporated on 31<sup>st</sup> October 1969. It also owns the Sam's Club retail warehouses. Walmart is the world's largest company by revenue with US\$514.405 billion, according to Fortune Global in 2019.

**B. Problem Definition:**

Walmart wants to predict the **Weekly Sales** of their stores. They have some of the features for 45 stores. They want to precise calculation of prediction of sales so that they can maintain their inventory at each store. Walmart has the dataset with Store Number, Weekly Sales, Temperature, CPI (Consumer Price Index), Fuel Price of week, Holiday Flag, Unemployment rate. All the data is available from February 2010 to December 2012.

- a) Make Exploratory Data Analysis (EDA) and plot time series for weekly sales.
- b) Some holidays have a negative impact on sales. Find out holidays which have higher sales than the mean sales in non-holiday season for all stores together
- c) Walmart wants to build a model to **predict target variable weekly sales** using machine learning algorithm. You are working as Data Scientist for Walmart and your manager approaches you to solve this problem. You have to select model which has high accuracy to predict sales of the store considering holiday, CPI, fuel price etc.

**C. Information of Dataset:**

Historical sales data for 45 Walmart stores located in different regions are available with us.

○ **Columns/Independent variables:**

▪ Start Date of week (dd- mm- yyyy)
▪ Temperature of week (degree Fahrenheit)
▪ CPI (Consumer Price Index)
▪ Fuel Price of week (\$)
▪ Holiday Flag (1: Not Holiday, 0: Holiday)
▪ Unemployment rate

○ **Target variable:**

▪ Weekly Sales
----------------

#### **D. Modeling Phase:**

- Selection of Model.
- Importing Libraries in python.
- Building selected model and predict sales.
- Evaluation: Check for accuracy and mean squared error.
- If accuracy okay, then done.
- If not, independent variables and evaluate.
- Evaluate till we get good accuracy.

Thank you!