**Walmart Sales Forecasting**

**Overview**

The main objective was to forecast weekly sales for each department in 45 stores located in different region.

we are going to predict which departments were largely affected by the holiday markdown events and what was the level of impact they had on the sales.

We would also like to create a data mining technique to find a specific value for Weekly Sales that we want to predict.

**Business understanding**

Business Problem here is to intimate the sales forecast and how much Walmart be able to analyze the effect of various factors affecting the sales of the products in the 45 stores.

The various factor includes Weather condition, temperature, store size, fuel price, markdown in prices, unemployment and CPI.

Data mining technique here helps to accurately predicts the sales forecast store wise, department wise and during markdown events. By this, the business problem can be easily controlled and managed.

**Scenario**

* The data has Store as primary key over which I have joined the table.
* I have 45 stores and 98 department.
* Weekly sales are the target variable.

**Data understanding**

The original dataset was downloaded from “kaggle.com”. it contains data of 45 Walmart stores and its various department.

Three data set are used:

**stores.csv**

This is historical training data which covers 2 yrs. (05/02/2010 to 10/26/2012)

This indicates types and size of store.

**train.csv**

This is the historical training data. Within this file you will find the following fields:

Store - the store number

Dept - the department number

Date - the week

Weekly\_Sales - sales for the given department in the given store

IsHoliday - whether the week is a special holiday week

**features.csv**

This file contains additional data related to the store, department, and regional activity for the given dates. It contains the following fields:

Store - the store number

Date - the week

Temperature - average temperature in the region

Fuel\_Price - cost of fuel in the region

MarkDown1-5 - anonymized data related to promotional markdowns that Walmart is running. MarkDown data is only available after Nov 2011 and is not available for all stores all the time. Any missing value is marked with an NA.

CPI - the consumer price index

Unemployment - the unemployment rate

IsHoliday - whether the week is a special holiday week

**Data Attributes**

* Store – the store number
* Dept – the department number
* Date – the week
* CPI – the consumer price index
* Weekly Sales – sales for the given department in the given store
* IsHoliday – whether the week is a special holiday week
* Temperature
* Fuel\_Price

**Supervised/Unsupervised**

Supervised data mining data technique is suitable when a specific “target value” was present to predict information about data. So here supervised technique is used.

Data Modelling: For this problem statement, since the outcome is a continuous variable (Number of sales), it is reasonable to build a Regression model. The Linear Regression model can be used to solve such problems since it is specifically used to predict continuous dependent variables.

**Target Variable**

* Weekly sales is the target variable.

**Data instance**

* Data instance here is data of 45 stores and 98 departments.

**How exactly would it add business value?**

* We could set pricing strategy
* We would attract the customers to the store and make profit to the maximum extent by them. Once the customers enter the stores, they are attracted then definitely they shop more by the special offers and obtain the desired items which are available in the favorable cost and satisfy them. If the products as per the needs of the customers, then it can make maximum profit the retailers can also make the changes in the operations.
* It can help in market basket analysis for upselling products.
* We could know purchase behavior of customer.
* We could perform better inventory management
* We could also make campaign sale more effective.