BUSINESS PROPOSAL

(Walmart Recruiting: Store Sales Forecasting)

Walmart is a well-known retailing corporation which operates as different types of hypermarket, departmental stores, grocery stores and garments buying house. For being a one of the largest retail company of the world, they often provide their datasets to public for forecasting or analyzing their information for taking of better decision about their sales.

Forecasting is about predicting the future as accurately as possible, given all of the information available including historical data and knowledge of any future events that might impact the forecasts.

OBJECTIVE

The aim of this project is to predict the weekly sales of 45 different stores/department of Walmart. Then, selected holiday markdown events are included which is known to affect sales, so, another objective is modelling the effects of markdowns on these holiday weeks in the absence of ideal historical data.

BENEFITS FOR THE CUSTOMER

Many retail businesses need accurate forecasting of the sales produced by each of their stores. These forecasts allow for planning, staffing optimization, as well as sure that each store has the necessary supply. Below are the benefits for the customer:

1. Predictable sales during holidays
2. Analysis of the effect of 4 different holiday markdown
3. This will help the Walmart for taking better decision on which inventory product they will focus and what kind of product they will need for that holiday
4. Also, to avoid overstocking in a store which is a waste money for the Walmart

DATASETS

The data has been taken from Kaggle, it contains data of 45 Walmart stores and its various departments. The original data files used for analysis were stores.csv, features.csv, train.csv and test.csv which contained the below mentioned fields:

Stores.csv → This file contains anonymized information about the 45 stores, indicating the type and size of store.

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

Train.csv → This is the historical training data, which covers to 2010-02-05 to 2012-11-01. 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

Test.csv → This file is identical to train.csv, except we have withheld the weekly sales. You must predict the sales for each triplet of store, department, and date in this file.

DELIVERABLE

The general deliverable will be sales analysis and forecasting which will help the Walmart to check their weekly sales of the department and the effect of markdown on the sales.

To build a model that can predict the weekly sales of different 45 stores having at least 60% of accuracy.

METHODOLOGY SELECTION

Like many projects, I can then proceed to three steps:

1. Data Cleaning: Cleaning process of our data and build some features. In this step, I need to analyze my dataset in order to clean it properly and to apply some features which is useful for the dataset
2. Predictive Modeling: Building and deploying a predictive model. In this part, I can use a lot of model like random forest, decision tree, extra trees, artificial neutral network etc., then, at last I will choose the best model.
3. Visualization: Creating a useful visualization of our predicted data

ETHICAL IMPLICATIONS

As a consultant, I don’t have the right to share the information provided by the client to their competitor. I need to secure and keep their data properly.