Comparative sales analysis of different stores, customers, demographics

Problem Statement

Store and market need analysis on regular basis to know the mood os customers. Now a day's all the store analyses their sales based on the data collected and predict the future demands. So there is big scope of these kinds of projects that accurately predicts and visualizes the sales and demands of product of different store.

So in this project we will collect the data from different store and apply machine learning techniques for the analysis of that data for their future demand and sales. Also we will track the customer behavior in different months and of different area.

Background

Performing sales analysis gives you valuable insight into the inner-workings of your business. Merchants use their data to make informed decisions like when to raise or lower prices on your products. These decisions shouldn't always be a "gut" feeling. Sometimes it can be reliable, but it shouldn't be your only decision-making tool. When looking for trends or patterns in your sales data, you can determine both opportunities and potential problems. You can track if a particular product is increasing or decreasing in sales. If it's declining, you can make timely decisions such as to cut prices, market more, or discontinue the product. If an item is selling off the shelves, you can be sure to stock inventory accurately across channels.

Sales trend analysis also helps you determine if you're meeting your sales goals by providing you an easy, measurable way to track your progress. You'll actually know if you increased sales from last year and by what percentage. If you didn't meet a goal, you can drill down to sales of a specific product or location to see what's stopping you. All retailers should have the ability to become data-driven businesses. With the right capabilities, you can have confidence in the decisions you make because they are backed by your own data.

Methodology

In this research, a machine learning model will be trained and tested on the collected dataset of different stores. For this kind of Analysis regression and logistic regression are two good approaches which can give better result in less effort. So in this analysis we will apply machine learning models for analysis i.e. linear regression, logistic regression.

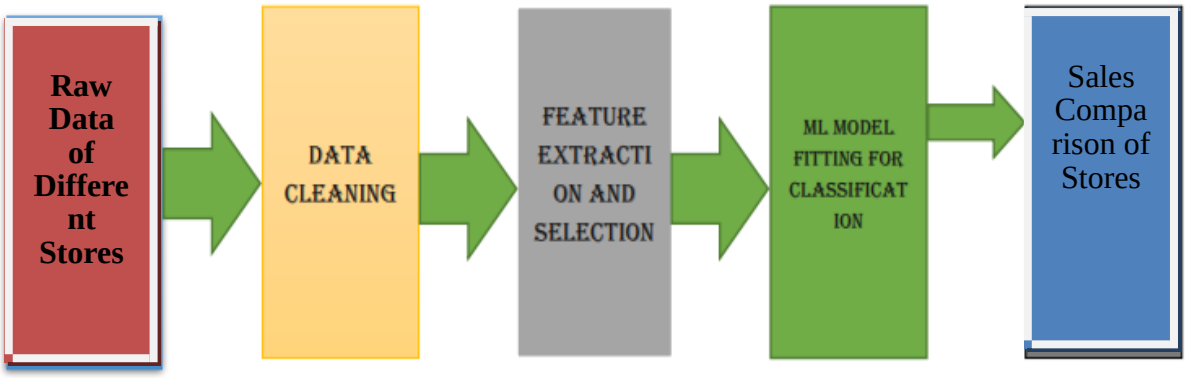

Fig-1 Methodology for fitting machine learning model

Fig-1 explains the method of experiment where we are taking the raw data from our source and will apply some data cleaning methods to make our data smooth. Then the most important step is feature extraction and selection will be applied to select best features out of available which are influencing the result more. Then we will apply Logistic regression of different dataset of stores and compare the results.

Experimental Design

Dataset

We are taking the historical sales data for 1,115 Rossmann stores. The task is to forecast the "Sales" column for the test set. Note that some stores in the dataset were temporarily closed for refurbishment. The link for data is:
https://www.kaggle.com/c/rossmann-store-sales/data
This dataset is standard dataset of sales, and provided for sales prediction of a famous Kaggle competition.

Evaluation Measures

Measures such as Classification error, Computational cost, Accuracy can be used for calculating the accuracy of drug discovery using neural network.

Software and Hardware Requirements

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such as, Numpy, Scipy, pandas, sklearn, will be utilized for this process. Training will be conducted on NVIDIA GPUs for training the Deep Q-learning technique for Drug Discovery.