

Sales prediction using Regression Analysis

Problem Statement:

Regression is an important machine learning model for these kinds of problems. Predicting sales of a company needs time series data of that company and based on that data the model can predict the future sales of that company or product. So, in this research project we will analyze the time series sales data of a company and will predict the sales of the company for the coming quarter and for a specific product.

For this kind of project of sales predict, we will apply the linear regression and logistic regression and evaluate the result based on the training, testing and validation set of the data.

Background:

A sales analysis report shows the trends that occur in a company's sales volume over time. In its most basic form, a sales analysis report shows whether sales are increasing or declining. At any time during the fiscal year, sales managers may analyze the trends in the report to determine the best course of action. Managers often use sales analysis reports to identify market opportunities and areas where they could increase volume. For instance, a customer may show a history of increased sales during certain periods. This data can be used to ask for additional business during these peak periods. A sales analysis report shows a company's actual sales for a specified period a quarter, a year, or any time frame that managers feel is significant. In larger corporations, sales analysis reports may only contain data for a subsidiary, division or region. A small-business manager may be more interested in breaking sales down by location or product. Some small, specialized businesses with a single location are compact enough to use general sales data. A sales analysis report may compare actual sales to projected sales.

Linear regression and logistic regression is the best machine learning models for this kind of problem where we can easily fit a line of high sale and low sale product, quarters and zone for a product. Also we need huge amount of data for the training of the model which we can collect from the sales data of any product or company of last 1 or 2 years for any live project. However, for this research project, the description of the dataset which we are going to use for this project is provided in the dataset portion of experimental setup section.

Methodology:

In this research, linear regression and logistic regression model will be trained and tested for our dataset. For this we will download the sample dataset from the given link in dataset section. The raw data is then undergoes for feature selection and feature extraction. After that we will apply machine learning regression models for the training dataset to train the model. This train model will be then tested on test dataset and validation dataset for checking the accuracy of the model.

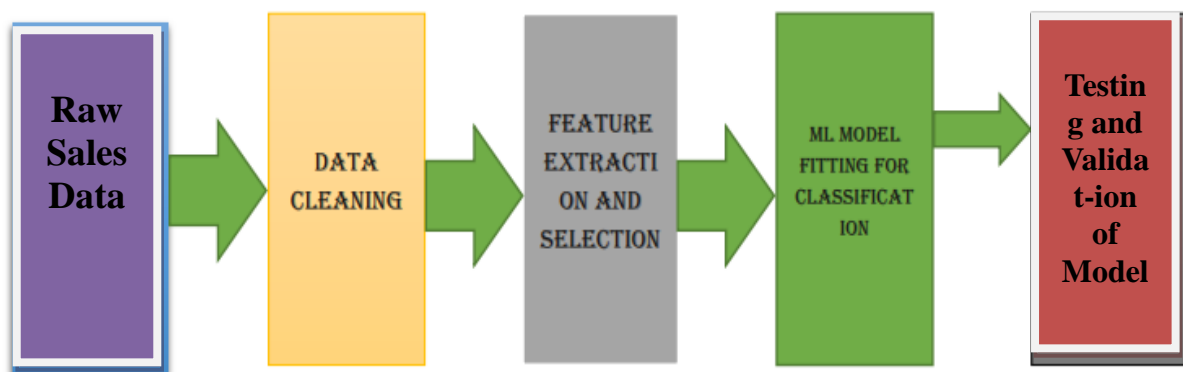


Fig-1 Methodology for fitting machine learning model

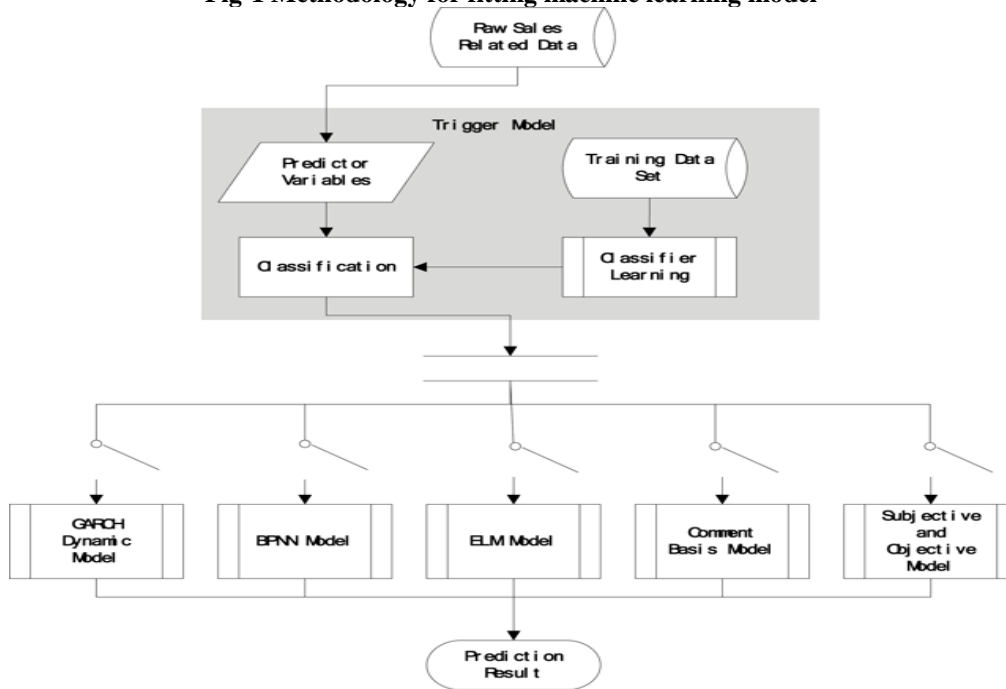


Fig-2 The Framework of trigger model system [Huang, Wenjie, et al. "A Novel Trigger Model for Sales Prediction with Data Mining Techniques." Data Science Journal 14 (2015).]

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. While fig.2 explains the trigger model system framework with some classifiers but for this research regression is proposed. 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 some machine learning model and compare the results.

Experimental Design:

Dataset

We are using the superstore sales data for sales prediction. Sample data that appears in the December Tableau User Group presentation. Note: Geographic locations have been altered to include Canadian locations (provinces / regions). The link for sample dataset is

<https://community.tableau.com/docs/DOC-1236>

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.