IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF ONLINE SALES USING LINEAR REGRESSION

[1]Mastan vali, [2]K.Sankeerthana, [3]B.Naveen, [4]N.Vishal

[1]Assistant professor, [2]student, [3]student, [4]student

Department of Information Technology

[1]JBIET, HYDERABAD, INDIA.

ABSTRACT: The aim of this paper is to analyze the sales of a big superstore, and predict their future sales for helping them to increase their profits and make their brand even better and competitive as per the market trends by generating customer satisfaction as well. The technique used for prediction of sales is the Linear Regression Algorithm, which is a famous algorithm in the field of Machine Learning. The sales data is from the year 2017-19 and prediction of data for the year 2019 is done. Then, real-time data of the year 2019 is also taken and the actual data of the year 2019 has been compared to the predicted data to calculate the accuracy of prediction. This is done so as to validate our results with the actual ones. This in turn would help them take necessary actions (which has been discussed later) for their increase their sales.

Index terms-Linear Regression, Machine Learning, accuracy, superstore

INTRODUCTION

We live in a world full of data. Data surrounds us everywhere. Right from handling monthly budgets, storing information on mobile phones, buying items from stores, all of it is stored in the form of data. In our everyday lives, we have to deal with a lot of data. This data could be as small as handling your monthly budgets to big ones like the data of a Multinational Company (often referred to as big data). One of the agents who work with a lot of data is a superstore, like ZMart. These big shopping complexes have a lot of data to work upon. Handling inventory, maintaining purchase from manufacturers, handling inventory costs, handling supplies data, handling with their sales, profits and quantity data, and many more. This is a tremendous task to work upon such a big dataset. Our ultimate task is generating profits and customer satisfaction and to maintain brand name. A lot of work has to be done on the dataset for its analysis and prediction.

This whole work is done so as to check the current position of sales and find out the future expected sales so that if any decline is found could be worked upon by doing proper market research and evaluating the trends of the market so that customer base could be increased. Also, within the store, what techniques (like putting discounts or updating inventory) could be applied so that our target customers increase their purchase and become a satisfied and a happy customer. All this is very important for any business to survive in this cut-throat competition and undoubtedly data science is very much required to fulfill this purpose. In this paper, we will describe the methodology to deal with such data along with predicting sales of the superstore for next years from the available tools like machine learning. A brief description about the processes involved in fulfilling our objective is discussed further. Also, the tools used for various processes would be even discussed.

LITERATURE SURVEY

In 1980's the big market sales is done through exchanging of items, not through money. The idea of those people was different when compared with the modern generation one. The value of money was not up to the mark so they go for exchange of items method. The first market developed in the year 1990 that too was in only few countries like Belgium, Florida etc. The E-commerce as is known today evolved as businesses (end to end process) started to shift from real time market to digital market. All of the business today as we see is done over the internet and anything which is not there is meant to be wiped off. Ecommerce, the online shopping system has brought down

political and physical barriers giving everyone in the world an equal playing ground for their market, everyone can put their products on sale through the e-stores (website dedicated to selling of product, a virtual store).Ecommerce, as is generally thought of, is not the birth child of "The Web" but it just got kick' started by "The Web".

SYSTEM ANALYSIS

To know about the existing and the proposed system we have to know about the web platforms and internet. The main reason for growth in online shopping is because of internet. The Internet, sometimes called simply "the Net," is a worldwide system of computer networks -- a network of networks in which users at any one computer can, if they have permission, get information from any other computer (and sometimes talk directly to users at other computers). It was conceived by the Advanced Research Projects Agency (ARPA) of the U.S. Ito create a network that would allow users of a research computer at one university to "talk to" research computers at other universities. A side benefit of ARPAnet's design was that, because messages could be routed or rerouted in more than one direction, the network could continue to function even if parts of it were destroyed in the event of a military attack or other disaster. Today, the Internet is a public, cooperative and self-sustaining facility accessible to hundreds of millions of people worldwide.

PROPOSED SYSTEM:

In order to overcome the defects of the existing system, we proposed a new system with new algorithm which has been extracted from machine learning. "LINEAR REGRESSION" an algorithm which is based on "SUPERVISED LEARNING". It performs a regression task. It is based on independent variables, used to find out the relationship between variables and forecasting. Linear regression is used to perform the task to predict the value of dependant variable(y) based on independent variable(x).so, this linear technique finds out a linear relationship between x(input) and y(output).

HYPOTHESIS FUNCTION FOR LINEAR REGRESSION:

$$Y=c+m1x1+m2x2+\ldots\ldots\ldots$$

Linear regression is a type of regression analysis where the number of independent variables is one and there is a linear variable [1]. The red line in the above graph is referred to as the best fit straight line. The line can be modelled based on the linear equation shown below.

$$y = a\_0 + a\_1 * x \quad (1)$$

The motive of the linear regression algorithm is to find the best values for a_0 and a_1. Before moving on to the algorithm, let's have a look at two important concepts that must be known to better understand linear regression.

SYSTEM ARCHITECTURE:

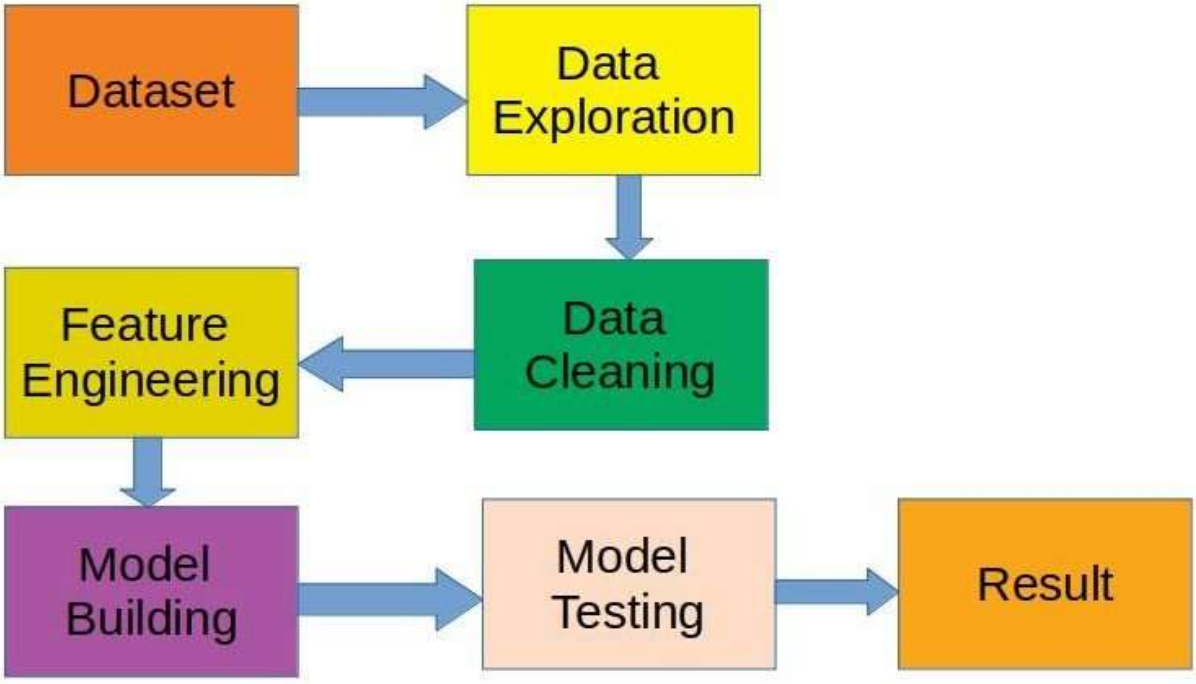
In this we will design a particular plan and sequence for our project, we go for an architecture which is of unique in style when compared with the other old architectures. There are different methods in which the data is trained.

DATA COLLECTION AND CLEANING:

In this technique the data is collected and undergoes for cleaning. The collected data will be saved in a separate file and it will be cleaned. Data cleaning is a method of detecting and clearing inaccurate data from the data records. Data cleansing may be performed interactively with data wrangling tools, or as batch processing through scripting. After cleansing, a data set should be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different stores.

MODEL BUILDING:

The processed data is used for predictive modelling so that appropriate results can be generated from it. This predictive modelling is done using a technique called Machine Learning [6]. It is defined as a "computer's ability to learn without being explicitly programmed". At its most basic, machine learning uses programmed algorithms that receive and analyse input data to predict output values within an acceptable range. As new data is fed to these algorithms, they learn and optimise their operations to improve performance, developing 'intelligence' over time. There are four types of machine learning algorithms: supervised, semi supervised, unsupervised and reinforcement. The most common and popular machine learning algorithms are- x Naïve Bayes Classifier Algorithm (Supervised Learning – Classification.



CONCLUSION:

This project entitled "Prediction of sales values Using Machine Learning Algorithm." is useful to predict a product, prices, and thereby to guide their customers accordingly. The proposed system is useful in achieving good prediction values for a product . This helps finally leads to the improvement of markets customers' satisfaction.

REFERENCES:

[1]. Xu-Ying Liu, Jianxin Wu and Zhi-Hua Zhou, "Exploratory Under sampling for Class-Imbalance Learning", IEEE Transactions on Systems, Man and Cybernetics, Vol. 39(2), pp. 539-550, April, 2009.

[2]. Deroski, Saso and Bernard Enko. "Is Combining Classifiers with Stacking Better Than Selecting the Best One?", Machine learning, Vol. 54(3), pp. 255-273, March, 2004.

[3]. Domingos Pedro, "A Few Useful Things to Know About Machine Learning", Communications of the ACM, Vol. 55, pp. 78-87, October, 2012.