Title: Predicting sales and gaining useful insights

Abhijeet Pradhan

Abstract

We live in a country where small scale and agriculture contribute more to the economy than other sectors. Therefore, in this project I have proposed an idea which will boost the profit of small shopkeepers, only by analysing the purchased goods data on a computer.

The basis of the system is a statistical model that combines effects of prices, promotion schedules and other factors that could influence demands, and produces a flexible nonparametric predictive demand function that can be optimized simultaneously in prices and promotion schedules

1) Problem statement

Shopkeepers always use various promotional and pricing schemes to improve their sales and revenue. Due to competition, various discounts are applied on similar products which leads to the complexity in the market. It can also lead to difficulty in understanding the customers response and effectiveness of a scheme and discount.

Hence, I want to create a model which will take historical data of products and identify patterns to aid in prediction of sales. This can use the scanner data which is available from past months.

2) Market/Customer/Business Assessment

After analysing the sales shopkeepers can predict the sales of products which help in buying the products. Due to this there will be no wastage or shortage of products. Because of analysing the data, shopkeepers can predict the demand of particular product and prices in the market of particular product. Also, customers buying preferences has been changed, nowadays customers buy in huge quantities. Therefore, by using this model, shopkeepers can use a different perception to sell their products.

3) Target Specification

The suggested model will be beneficial to the shopkeepers to increase their revenue and understanding the customer behaviour using historical data. This model can show some insights in data which can be helpful. After using this model shopkeepers can identify strong products which are selling more and weaker products which are declining sales. Also we can predict the behaviour of customers, like when they buy, what they buy and how much they buy. It can also predict the demand of products in particular time.

4) External Search

The sources used as reference for analysing data to improve sales are listed below.

https://www.sales-i.com/blog/businesses-using-analytics-improve-sales

i) Benchmarking

To build this model we need statistical model which combines various features and conditions to analyse data. Machine Learning makes easy in using statistical tools which can be time efficient, big E-commerce industries like Amazon, Flipkart and also franchise like Starbucks, McDonalds uses Machine Learning to predict sales and to understand customers behaviour.

Applicable Patents

- Patent 1: System and method for profit maximization in retail industry
  Link:
  https://patents.google.com/patent/US7379890B2/en?q=data+analysis+to+predict+sales&oq=data+analysis+to+predict+sales

- Patent 2: Computer, prediction method, and prediction program
  Link:
  https://patents.google.com/patent/JP5963709B2/en?q=data+analysis+to+predict+sales&oq=data+analysis+to+predict+sales

There several patents on predicting the sales, since these two represent the problem, I have mentioned them.

The first patent describes the importance of data mining to boost revenue by predicting the future sales. By applying data mining procedures to historical data that contains sales data and various sales condition which can result in customized reports with results of particular prediction schemes.

The second patent describes the point of sales which are predicted using past data. But it mainly focus on new products which do not have past data they are called as unhandled data. It also describes customers behaviour on purchasing products.

Hence, these will be considered while building the model.

Applicable Standards

- Data privacy
- Cybersecurity
- Government regulation on small business
- Laws regarding data

Applicable Constraints

- Collecting data
- Data privacy and maintaining data input
- Lack of technical knowledge
- New products with no historical data
- Implementation of systems

Business Opportunity

Since the model is new for business it will make great impact in small business as it is doing in big business. It can be used by shopkeepers, vegetables vendors, goods distributer, small online business. Every business who wants to increase their sales and understand customer can use this model.

If your business offers many products, using this model you can conduct regular sales analysis and predict the items overcrowding your products. You can see bar charts and product sales in time framework.

Concept Generation

In current time customers are accurate about purchasing products, its owner's duty to provide with appropriate products. To make process effortless we need a model which can do this within few seconds by using historical data.

This model needs Machine Learning tools and coding to make it user friendly. It needs historical data, on which it can be trained to predict further sales. Model can use various tools to generate more accuracy of model.

Concept Development

For development of model, we can use API and framework and same for the deployment of the model.

Example: The model will create columns and index according to the products, later we need to fill these columns with historical data to train and test the model. Once the model gets trained it will show the accuracy, charts of sales, patterns of sales. Then we can feed the data to predict the sales and identify the patterns init.

Final product prototype

Historical data   Machine Learning   Predictive mode

- The first image shows the layout of model that we need historical data which will be used by machine learning tools to predict the sales
- Business owner can create their own columns of products or group of products and predict the sales using historical data available to them.
- This model will show bar charts which will represent the sales of products in specific time period which can show the demand of product in specific time.
- This model will also show the number of products sold and purchased by the owner.

Product Details

How does it work
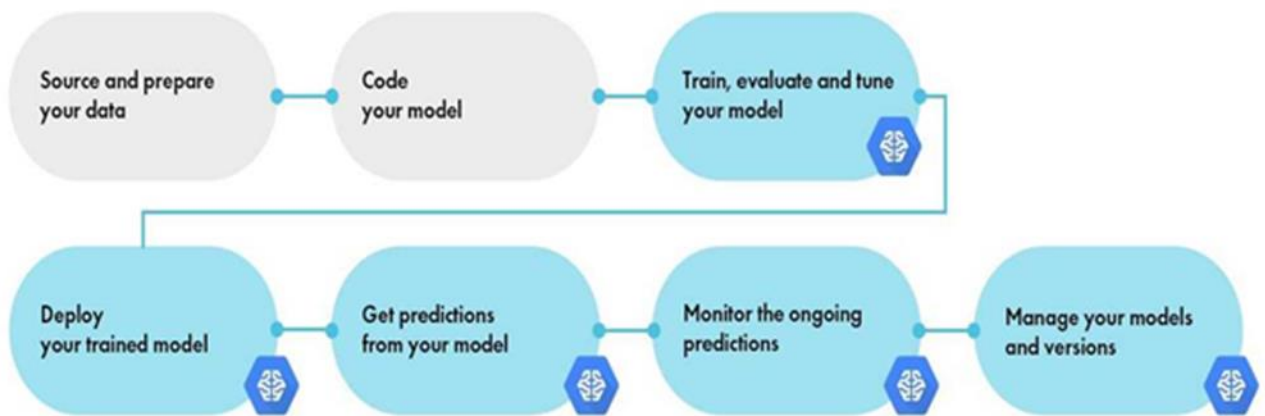
- Model takes the features which will affect the sales of products like effects of prices, promotional schedules and other factors.
- Then fit the historical data in which will it will train and test the accuracy.
- Now the model will display the accuracy and ready to predict the sales of products.

Example: A shopkeeper is predicting the sales of shoes in upcoming monsoon season. To predict the sales he need to input the data in specific features like types of shoes, size, colours, brand and prize. After predicting model will show shopkeeper prediction and patterns on sales of shoes.

Requirements

Algorithms to predict sales:

1. Logistic Regressor: A logistic regression model can take into consideration multiple input criteria. In the case of college acceptance, the logistic function could consider factors such as the student's grade point average, SAT score and number of extracurricular activities. Based on historical data about earlier outcomes involving the same input criteria, it then scores new cases on their probability of falling into one of two outcome categories.

2. Random Forest: Random Forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems.

3. Support Vector Machine: There are three different implementations of Support Vector Regression: SVR, NuSVR and LinearSVR. LinearSVR provides a faster implementation than SVR but only considers the linear kernel, while NuSVR implements a slightly different formulation than SVR and LinearSVR

Python Libraries required

1. Pandas: To read dataset and analyse them in a dataframe.

2. NumPy: It helps in any mathematical operation used in algorithms and codes.

3. Matplotlib: It is used to plot charts, label the charts, size of charts

4. Seaborn: Its is used for data visualization like plotting bar charts, count plot, use different colours for charts, plot various charts like histogram, bar plot, kde.

Conclusion

Machine Learning is a modern tool which can be in very effective ways to simplify the process. Big E-commerce companies are already using modern tools in their business to improve their productivity and gain useful insights in data. But this tools can also be beneficial in small business to improve their business with less time.

Hence, I conclude my project.