International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-10 Issue-3, February 2021

Predicting the Demand for Fmcg using Machine Learning

Anish Mebal.P, Hema.S, Jothika.S.J, M.Manochitra

Abstract: ***Now-a-days the more accurate prediction of the demand for fast-moving consumer goods (FMCG) is a competitive factor for both the manufacturers and retailers, especially in the super markets, wholesale manufacturers and fresh food sectors and other consumable industries. This proposed system presents the benefits of Machine Learning in sales forecasting for short shelf-life and highly-perishable products, as it predict the statistical information as a result, improves inventory balancing throughout the chain, improving availability to consumers and increasing profitability. This performance is done with various classification algorithms and comparative study is done with some metrics like accuracy, precision, recall and f-score. So that it helps in finding customer need and to increase the profit of the manufacturers.***

Keywords: FMCG , Train set, Test set, Goods.

I. INTRODUCTION

Fast-moving consumer goods are goods or products that sell in market quickly at affordable cost. Such items are considered as "FAST MOVING" as they are quick to leave shelves of a store or super market because consumer use them on the regular basis. FMCGs have a short shelf life because of high consumer demand because they include soft drinks and confections or meat, dairy products, baked goods, grocery items etc. These are the goods that purchased very frequently at minimal price and are sold in large quantities. Fast-moving consumer goods are the largest segment of consumer goods. They fall into the nondurable category, which means goods have a shelf life of less than one year.So that they are consumed immediately. In the study it is found that more than half of all consumer spending very low-involvement purchases. Because the consumer need is to purchase goods with low cost and with high-quality. To satisfy the consumer needs and also to gain huge amount of profit FMCS is used.

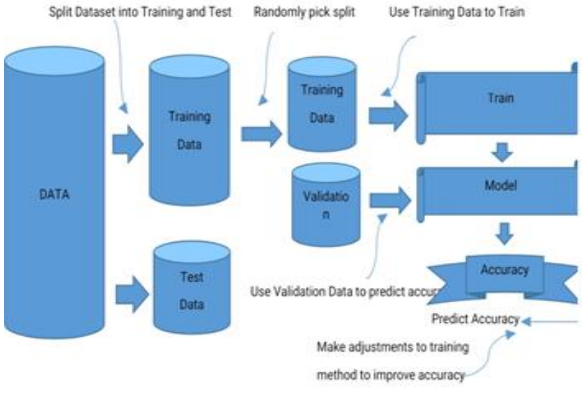
II. OBJECTIVE

1. To develop a system for increasing the sales activities in various business developments
2. To increase the classification accuracy using Machine learning classifier.
3. To Understand K-nearest neighbor (KNN) which is one of the most popular machine learning algorithm
4. To learn the working of KNN in python.
5. To choose the right value of K in simple terms.

III. PROPOSED SYSTEM

For finding the fast moving consumer goods (FMCD) we have used K-Nearest Neighbor (KNN) algorithm. K-Nearest Neighbor is a simple algorithm that stores all the data available in dataset and classifies the case based on similarity measures. It classifies a data point based on how its neighbor are classified, which means when new data appears then it can be easily classified into a nearest available point using KNN algorithm. First the data is fetched. Then during the time of processing the dataset is divided into two sets i.e. Training set and testing set. We will use 75% of data from dataset in training phase and remaining 25% of data in testing phase in the ratio of 75:25. Training set is a subset to train the module and the algorithm uses the training set to learn. Testing set is a subset to test trained module. After processing the training dataset, the model achieves almost 99% of precision on both the training set and the test set. After splitting the dataset into training and testing dataset, we have to instantiate K-nearest classifier. For that assume the value of K. The value of K may vary from one project to another. The most preferred value of k is 5 and greater than 5.So that the noise can be reduced. Then calculate the Euclidean distance between the existing points and new point. Then the algorithm searches the distance for the five (if the value of k is 5) closet nearest neighbor and predict where the new point fall for. Finally we can get the nearest neighbour and plot in graph based on the result.

IV. DIAGRAM



Manuscript received on February 17, 2021.
Revised Manuscript received on February 26, 2021.
Manuscript published on February 28, 2021.
* Correspondence Author
Anish Mebal.P*, CSE, National Engineering College, Kovilpatti, India. Email: anishmebalp@gmail.com
Hema.S, CSE, National Engineering College, Kovilpatt, Kovilpatti, India. Email: hemasankar1407@gmail.com
Jothika.S.J, CSE, National Engineering College, Kovilpatti, India. Email: jothikajo2300@gmail.com
Manochitra.M, Asst. Professor, CSE, National Engineering College, Kovilpatti, India. Email: :manochitra-cse@nec.edu.in

ALGORITHM

A. Import libraries

Start the program by importing the libraries required to implement the KNN algorithm in python. The following libraries are necessary for implementation

- ✓ numpy libraries for scientific calculation
- ✓ matplotlib. pyplot library for plotting the graph
- ✓ from sklearn.model_selection import train_test_split
- ✓ from sklearn.preprocessing import StandardScaler
- ✓ from sklearn.neighbors import KNeighborsClassifier

B. Data Fetching

Then we should fetch the data path from the respective folder and store it as 'dataset'. Then read the entire dataset using read function and then print the first five rows of dataset.

C. Define Predictor Value

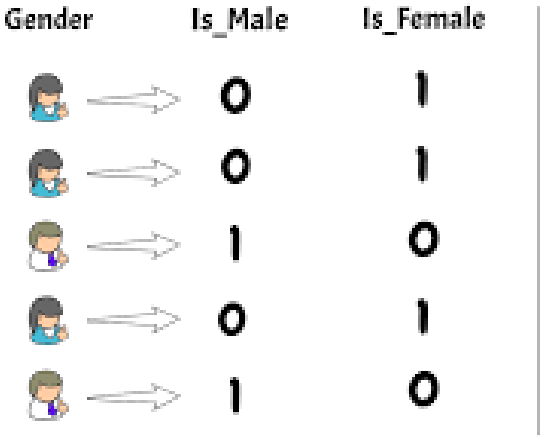
The Predictor Variable which is known as independent variable that is used to determine the value of the target variable. We use the values of 'product_id' and 'price'as a predictor variable. The store the predictor variable in 'X'.

D. Define Target value

The target variable is nothing but the dependent variable whose values are predicted by the predictor variable. Then store the dependent variable in 'Y' i.e. purchased which is 1 if customer purchases more than the target value and 0 if the target value is not reached.

E. Data Encoding

Gender- Some might say that Gender would play a role but that is really subjective to discuss. Moreover Since gender is a Categorical variable we would have to use Variable Encoder for it.

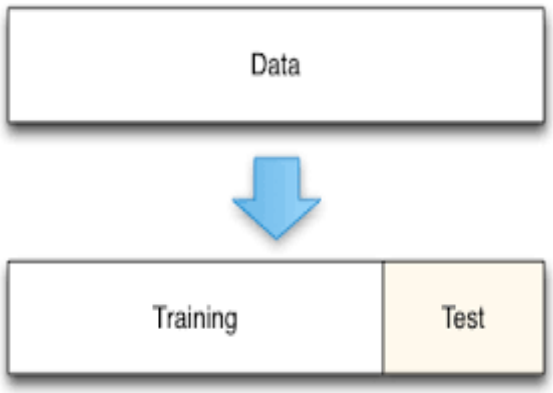


F. Feature of Scaling

Feature scaling is a method in which we scale the data into an accurate and scalable size for the purpose of increasing accuracy and reducing error. It basically prevents the large variance of data points to be used in the algorithm and allows us to achieve better results. Standard Scaler is a class imported from sklearn library. Here, standardization method has been used. We have standardized 'Price' and 'Product ID'.

G. Split the dataset

Now splitting the dataset into training dataset and testing dataset. 'X-train' and 'Y-train' are train dataset and 'X-test' and 'Y-test' are test dataset. Then divide the dataset in an 80:20 ratio. We will use first 80% of data for training set and remaining 20% of data for the testing set



V. PSEUDOCODE

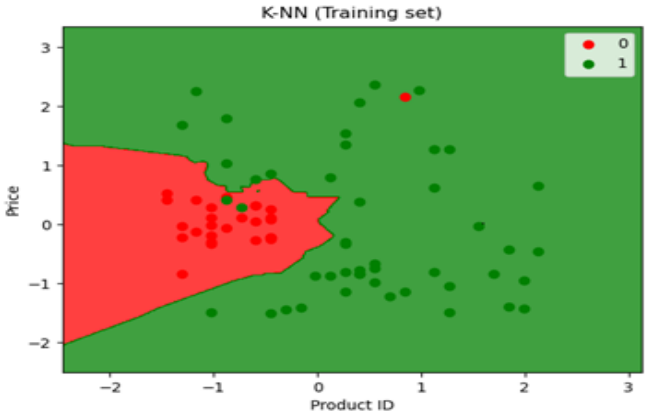
To Find Accuracy:
After finding values of x_test, x_train, y_pred, y_train
First assign c=0
Using for loop in range 0 to length of y_pred
If y_pred is equal to y_test
Increment c by 1.
To find accuracy, use the formula accuracy=c/len (y_pred)
Print accuracy.

```
c=0
for i in range (0, len (y_pred)):
   if (y_pred[i] ==y_test[i]):
      c=c+1
accuracy =c/len (y_pred)
Print ("accuracy is")
Print (accuracy)
```

VI. RESULT

A. Training Set

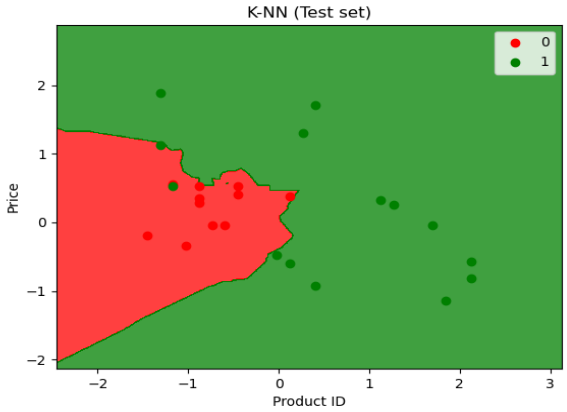
This section presents some results obtained by applying the kNN algorithm on the dataset



B. Testing Set

Retrieval Number: 100.1/ijeat.C22530210321
DOI:10.35940/ijeat.C2253.0210321
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication



The results show that our Training KNN algorithm was able to classify all the records in the training set with 99% accuracy, which is excellent accuracy. Our model does not affect Over-fitting and Under-fitting. Therefore from the results, KNN algorithm was able to classify all the records in the test set with 97% accuracy, which is excellent. Although the algorithm performed very well with this dataset, don't expect the same results with all applications. As noted earlier, KNN doesn't always perform as well with high-dimensionality or categorical features.

VII. CONCLUSION

Sales and demand forecasting has always been one of the main issues of the FMCG and retail Industry. Having an accurate prediction of the amount of sale helps all the supply chain actors to plan and operate accordingly. This leads to a more efficient, robust, effective, and sustainable supply chain operation. At the same time, the amount of data generated and stored by supply chain actors are becoming enormous. These data are collected from heterogeneous sources and satisfies the characteristics of Big data by having six main Vs of Big data, being, Volume, Velocity, Variety, Variability, Veracity, and Value. Therefore, it is crucial to formulate a roadmap towards utilization of this data. This formulation should be in such a way that, it considers the specifics of FMCG and retail industry, being demand uncertainties. On the other hand, machine leaning techniques have shown a great potential in providing a solution for this type of problem. The result of this research show that Machine Learning techniques, including the latest methods such as Deep Learning as well as utilizing a combination of various techniques applied in demand forecasting models, can provide benefits to manufacturers and retailers of fast-moving consumer goods. Thus, this study contributes to the identification of the Machine Learning benefits and characteristics, applied in improving demand forecasting accuracy in the FMCG industry, which is a key element of competitiveness.

REFERENCES

1. Singh Manpreet, Bhawick Ghutla, Reuben Lilo Jnr, Aesaan FS Mohammed, and Mahmood A. Rashid. "Walmart's Sales Data Analysis-A Big Data Analytics Perspective." In 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), pp. 114-119. IEEE, 2017.
2. Sekban, Judi. "Applying machine learning algorithms in sales prediction." (2019).
3. Panjwani, Mansi, Rahul Ramrakhiani, Hitesh Jumnani, Krishna Zanwar, and Rupali Hande. Sales Prediction System Using Machine Learning. No. 3243. EasyChair, 2020.
4. Cheriyan, Sunitha, Shaniba Ibrahim, Saju Mohanan, and Susan Treesa. "Intelligent Sales Prediction Using Machine Learning Techniques." In 2018 International Conference on Computing, Electronics & Communications Engineering (ICCECE), pp. 53-58. IEEE, 2018.
5. Giering, Michael. "Retail sales prediction and item recommendations using customer demographics at store level." ACM SIGKDD Explorations Newsletter 10, no. 2 (2008): 84-89.
6. Baba, Norio, and Hidetsugu Suto. "Utilization of artificial neural networks and GAs for constructing an intelligent sales prediction system." In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, vol. 6, pp. 565-570. IEEE, 2000.
7. Ragg, Thomas, Wolfram Menzel, Walter Baum, and Michael Wigbe. "Bayesian learning for sales rate prediction for thousands of retailers." Neuro computing 43, no. 1-4 (2002): 127-144.
8. Fawcett, Tom, and Foster J. Provost. "Combining Data Mining and Machine Learning for Effective User Profiling." In KDD, pp. 8-13. 1996.
9. N. S., Ahrens, D. (2017). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. International Journal of Production Economics, Vol.170, 321-335

AUTHORS PROFILE

Anish Mebal.P is currently pursuing B.E(CSE) final year in National Engineering College , Kovilpatti, India.
Area of Interest : Machine Learning

Hema. S is currently pursuing B.E(CSE) final year in National Engineering College , Kovilpatti, India. Area of Interest : Machine Learning

Jothika.S.J is currently pursuing B.E(CSE) final year in National Engineering College , Kovilpatti, India.
Area of Interest : Machine Learning
.

M.Manochitra is working as an Assistant Professor at National Engineering College, Kovilpatti, India. She has received M.E from, Anna University; Chennai. Her area of interest in machine learning concepts. She has published papers in national/International journals and conferences.