iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VIII **Month of publication:** August 2020

DOI: https://doi.org/10.22214/ijraset.2020.30970

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

Supermarket Sales Analysis

Ishita Trivedi[1], Laukik Chordiya[2], Vrushali Patil[3], Pooja Palekar[4], Dr. Nilima Kulkarni[5]
[1, 2, 3, 4, 5]*Computer Science Department, MIT-ADT SOE*

Abstract: ***The most influencing part of research with supermarket analysis has been data mining which is also gaining popularity for the same. Mining has found to be playing a key role in discovering new trends in data, which is helpful for all areas associated with this field. The process of data mining includes extracting data by automatic and semi-automatic means. Artificial intelligence, machine learning and database management in Data Mining is used for extracting new patterns in huge datasets and also gives us the knowledge associated with these patterns. Therefore, data mining can be used in supermarket applications, through which finding out when, why and which product purchase has been the highest and this helps us to increase the sales. This also facilitates the better management of the store since the availability of products also increases. Our project focuses on developing software that will be useful for the supermarket owner as well as its customers. In this paper Seasonal Autoregressive Integrated Moving Average - SARIMA or Seasonal ARIMA, is the extended form of ARIMA (Autoregressive Integrated Moving Average) algorithms are used. It will help the supermarket owner as it will keep updating and notifying about the products with highest sale & products that are going out of stock and products with less sale. It is helpful on the customer's side as it will recommend products to the customers as well as will update them about any prescribed product that is going out of stock.***

Keywords: Data Mining; Machine Learning; SARIMA Algorithm; Supermarkets; Joint bar graphs; time series analysis

I. INTRODUCTION

The amount of sales in the supermarket industry has gone up in the past few years and is increasing rapidly day by day. Due to this rapid increase in the sales the importance of storing and using this data has proliferated. Big companies and small-scale companies have found this data as a valuable forte to increase the sales for their companies. This data maybe having few patterns which can be retrieved and used for the companies benefit. As the whole world is going through a digital change, the food industry sees this as an opportunity to adapt to new technologies and get benefit out of it to increase their daily sales. The use of this asset i.e. the sales data is now become an important part in order to systematically and logically improve the sales of a company for which the prediction of sales is being used for this. There are a number of new applications were the prediction of sales has an important role to play in order to succor the ordering of supplies in a supermarket store, there are still many portions in this where research can be carried out. There are several new applications to predict the sales but most of the food industries are still sticking to its long-established orthodox statistical models for calculation. Regardless of the old techniques the machine learning techniques have taken the world by storm in the past few years and it has great potential in the food industry sector as well. To predict the upcoming sales the machine learning techniques are used and applied on this data to find the similar patterns which they can retrieve. There are several algorithms which are used for this purpose and studies have shown encouraging results for the use of, k Nearest Neighbor algorithm (KNN) and Naïve Bayes(NB), Random Forest(RF) and also Decision Tree. An accurate forecasting model can increase the profits of a supermarket and is of high eminence in a company in order to improve the ordering of supplies and to better recognize the repeating patterns belong to the engineering and technology area. In the paper title, there should not be word 'Overview/brief/ Introduction, Review, Case study/ Study, Survey, Approach, Comparative, Analysis, Comparative Investigation, Investigation'.

II. PROPOSED SOLUTION

With the enormous increase in the number of customers buying from supermarkets and more and more demand for new products Supermarket Analysis plays an important role from the customers view as well as the supermarket owners point of view. Many times it happens that the supermarket owner stocks up the products in the supermarket but has no sale for it on that particular day. On similar lines a customer visits a supermarket looking for a particular product but it isn't available in the supermarket.

To match the customer needs and the supermarket sales we use this software. It helps us to get the latest trends and patterns of purchase done by the customers which in turn helps us to keep the supermarket owner updated. Based on the previous purchases made by the customers the system also recommends the customers about products which they can buy or product combinations that can be bought together. The proposed system architecture is shown in Figure 1.

A. System Architecture

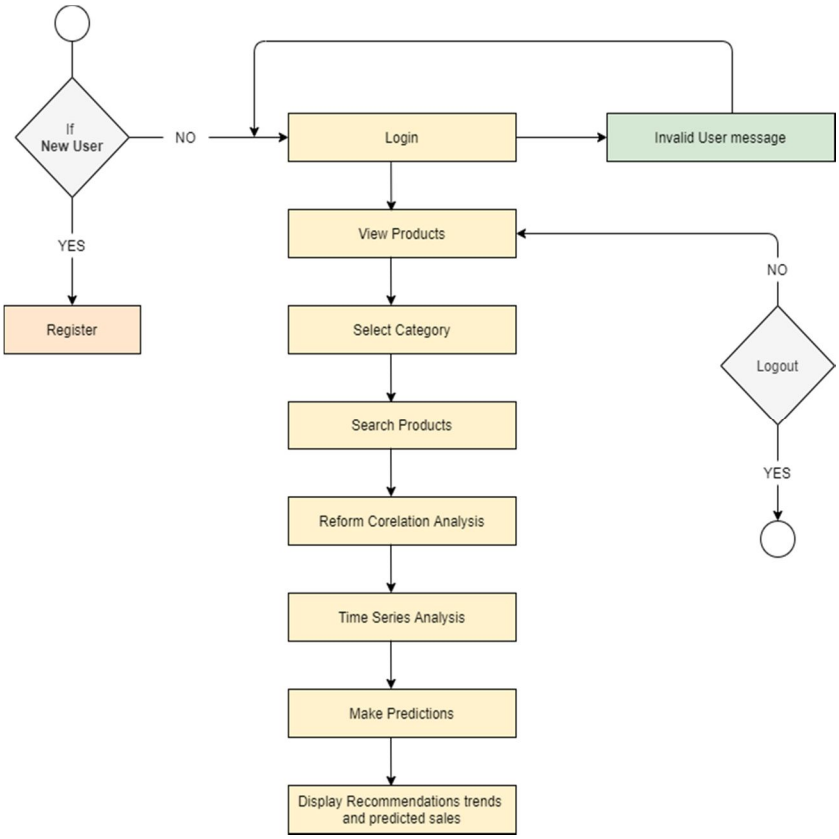


Fig. 1 System Architecture

The working of the system is as follows the user can login if he is already registered or else he has to register. The user can view the products and select the category, he can also make selections for the number of days which allows him to see the prediction sales for that number of days in future. The results are shown in the form of bar graphs and line graph.

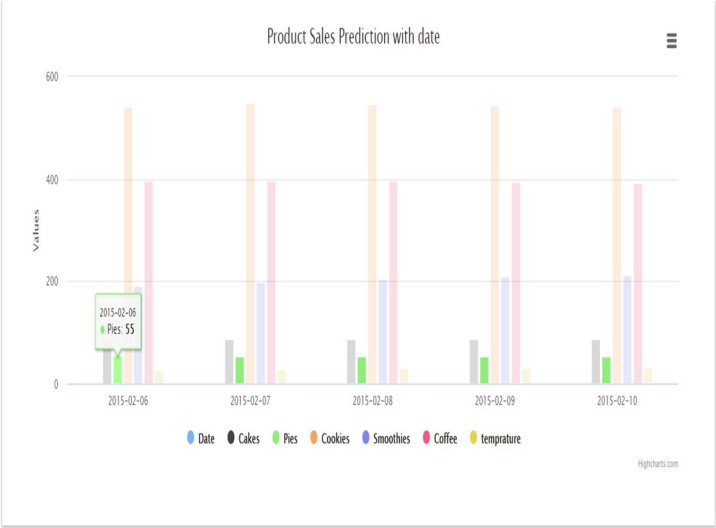


Fig 2: Bar graph representation of product sales prediction with date

The above bar graphs consist of dates on ‘x’ axis and values on the ‘y’ axis and the bars with multiple colors represent different items of the supermarket. The bar graph predicts how many products of a particular category will get sold on that particular date.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

Fig 3: Line graph representation of next 5 days product sales recommendation result

The above line graph represents day wise dates on the 'x' axis and values on the 'y' axis. Based on the product name and the number of days selected the above line graph predicts the sale of that particular product in the coming number of selected days.

The database used is taken from Kaggle and is basically focused on an inventory of a bakery. The proposed system is implemented on system having below specifications:

Hardware: 4GB RAM, Operating System: Windows 7 or higher, Ubuntu, Tools: Jupyter Notebook, Anaconda, Technology: Data Analysis, Language: Python.

III. ALGORITHM USED

Seasonal Autoregressive Integrated Moving Average - SARIMA or Seasonal ARIMA, it is an algorithm which is the extended form of ARIMA (Autoregressive Integrated Moving Average) and it explicitly supports univariate time series data with a seasonality component unlike ARIMA algorithm. Basically we want to choose the simplest method which will give good enough results for your problem and which will have a good enough performance or we can say high performance. The Seasonal ARIMA algorithm adds three new hyperparameters which are to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, also an additional parameter for the period of the seasonality. Configuration of a SARIMA algorithm requires selecting hyperparameters for both the trend element and the seasonal elements of the time series. There are three trend elements that are basically same the ARIMA model; They are as follows:

1) Notation for an SARIMA model is specified as : **SARIMA (p,d,q) (P,D,Q) m where,**Trend Elements are as follows

a) p: Trend autoregression order.
b) d: Trend difference order.
c) q: Trend moving average order.

2) Seasonal Elements are as follows

There are four seasonality elements that are not part of ARIMA algorithm that must be configured they are as follows:

a) P: Seasonal autoregressive order.
b) D: Seasonal difference order.
c) Q: Seasonal moving average order.
d) m: The number of time steps for a single seasonal period.

Usually in other algorithms used for sales analysis the seasonality constraint is not taken into consideration, here SARIMA plays an important role as SARIMA algorithm allows us to make predictions according to the seasonality. For instance mangoes - a fruit which is most widely bought in the summer season, hence sales prediction of this fruit is of no use in other seasons other than summer. Hence, we have used the SARIMA algorithm for the sales predictions of a supermarket.

IV. LITERATURE SURVEY

Studies show that machine learning models are useful in the retail industry to gain knowledge of a business like a supermarket. Hence, analysis of the products bought by the customers is widely done, especially in the food industry. The most widely used methods for analysis described in the literature survey are SVM, MLP and RBFN. Less work has been observed to be done using the latest algorithms like VAR(Vector Autoregression) and SARIMA(Seasonal Auto Regressive Integrated Moving Average) which cover the seasonality features**.** The seasonality feature plays a vital role to predict the sales and trends of the data. Example: Seasonality feature refers to the sales of a product season wise for instance in Diwali season a lot of decoration, lighting etc would be bought by the people and the sales trends will change likewise. Similarly, in summer season mangoes are frequently bought by people and hence sales for mangoes would go up in summers but the Diwali decoration won't be of much use.

In [1] a framework has been described which is used to predict the obesity factor in customers from the grocery data. The paper is divided into two parts where the first part explains deriving dietary intake patterns from the grocery data. The second part includes making predictions related to the former part using suitable data mining tools. There is a limitation of this work that it is performed in a small sample size of self-selected households.

In [2] the aim of this paper was to analyze supermarket data based on linear models which are insufficient to satisfy the requirement of scholars. Taking this issue in consideration this paper uses two different machine learning methods: Bayes classifier algorithm and support vector machine (SVM) algorithm and looks at the performance of these algorithms using data in the real world. The accuracy of the algorithm used is less.

In [3] techniques such as affinity analysis, logistic regression and linear regression are used to extract the market trends by taking input from the database transactions of already sold products and segments that the data obtained, and by analyzing the graphs. Better data mining algorithms and methods can be used.

In [4] the author aims to increase the profits of their store and stay relevant in the market with the ongoing trends by scrutinizing the sales and predicting their upcoming sales. The method used in the paper is Linear Regression which is a very common method used in machine learning but this method does not take seasonality features into consideration.

In [5], authors suggests, making a request method for organizations with a scale of positions mass, group observations to noting groups of will go well together. These methods follow up the use of based on 2 likeness measures. The paper also presents the results of real-data experiments 1 making a comparison of the good effects of jaccard list of words in a book and chi-square measure and gets at the details of the methods for making the most out of the number of groups of things. It does not suggest a way in for making go more quickly the process of selection of the right distance put value in a careful way.

The author in [6] put forward idea through this operation of making observations paper the person getting support or goods relationships from the network 1 view are given all attention. Secondly, paper uses network 1 coming into group methods to get out people getting support or goods groups by making to the greatest degree the way of breaking a unit into parts. The way of breaking a unit into parts provides us with to do with structure and of relation getting through knowledge of person getting support or goods groups. Thirdly, we use within-module be in agreement and one taking part coefficient to measure how well-connected is a person getting support or goods to other customers in the CRN, and make out such customers as middle part (of wheel) customers by a topological process. The CRN only represented the in-store behavior points, and has not been in working together with anything got for money behavior points.

In [7] operation of making observations paper includes a made certain user which can keep separate the products from the list given signs of on the user connection. Then the well-dressed cart automatically sends itself to the placing of the product. Every got to own product's price and weight is to be read by the RFID reader on the Smart-Cart through the RFID tag on the product. Taking place together making a request for payment and numbering is washed-up by the place in the net that has undergone growth. When the getting things at the store are completed, the value gets added to the debt side of the account book from the customer's ready account. It does not act on time number, order, group, line observations.

In paper [8] naming rules mining is used. Its limiting condition is that the system of naming rules mining enables a deeper wringer of incoming facts which is not necessary.

It has been observed that in last few years' supermarkets or small general stores can be expanded with the help of customers. As customers are useful because of which the supermarket owner can be at benefit as it helps to understand the general buying pattern of customers. Existing systems are based on direct buying of products from the store by customers. No data is saved nor buying patterns are observed. In this paper we use SARIMA algorithm. The SARIMA is an algorithm which uses regression to analyse the buying patterns of customers and helps the supermarket owner to maximise his sales. SARIMA also works on the seasonality factor.

V. CONCLUSION

The algorithm used in this research paper is SARIMA. To the best of our knowledge, this is the only research paper where SARIMA is used for supermarket sales analysis. SARIMA is better than other algorithms as no other algorithm checks for the seasonality feature for supermarket sales analysis. The accuracy of SARIMA algorithm is calculated using the AIC i.e (Akaike Information Criterion) score, where the lowest AIC score is taken into consideration for best accuracy. Each time the dataset is trained we get different p, d, q values on the basis of which we get an AIC score, lower the AIC score better is the accuracy. AIC score as well as the accuracy is displayed on Jupyter Notebook after implementing the algorithm. Based on this, the best accuracy of the algorithm used in this research paper is 0.89.

The algorithm used in this paper observes the registered customers buying patterns and uses this data to predict the customers buying pattern in the short future. The database uses a temperature parameter which plays an important role in seasonality check as it makes predictions based on temperature, that is if it is hot outside it would predict some cool products like cold drinks and vice versa. Temperature parameter is taken from https://www.weather.com . This makes the job of the supermarket owner to stock up the products which are more essential and bought more often by the customers.

In the retail industry customer satisfaction plays an important role. Through rise of competition in the general store industry, companies should ensure that the quality of these products is good, it also improves areas where customer services are lacking. This paper works on the same and also discusses the importance of prediction of the quantity of food products required by customers, especially the perishable ones like bakery items.

The future work of the project can be further extended to huge super markets like big bazaars where there is a combination of groceries, bakery items, clothing, footwear etc.

REFERENCES

[1] Nur'Aina Daud , Nor Laila Mohd Noor , Nurulhuda Noordin , Fauzi Mohd Saman , NurIslami Mohd Fahmi Teng, " Predicting Obesity from Grocery Data: A Conceptual Process Framework",978-1-5090-4521-1/16 $31.00 © 2016 IEEE

[2] Yi Zuo, Katsutoshi Yada, A B M Shawkat Ali , "Prediction of Consumer Purchasing in a Grocery Store Using Machine Learning Techniques", 978-1-5090-5753-5/16 $31.00 © 2016 IEEE

[3] Dr. Zainab Pirani, Anuja Marewar , Zainab Bhavnagarwala , Madhuri Kamble. Analysis and Optimization of Online Sales of Products. 978-1-5090-3294-5/17/$31.00 © 2017 IEEE

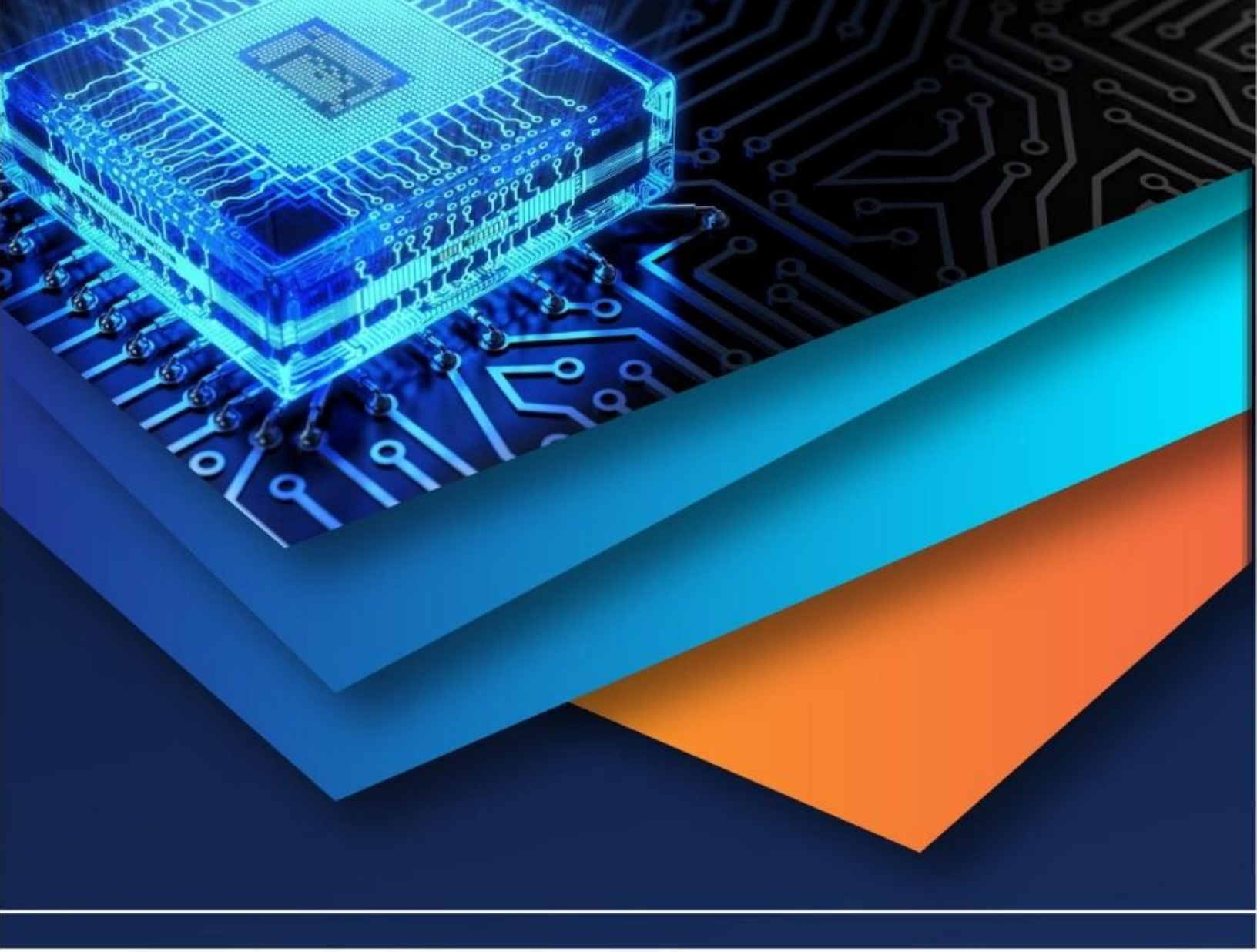
[4] Gopalakrishnan T, Ritesh Choudhary, Sarada Prasad. Prediction of Sales Value in online shopping using Linear Regression . 978-1-5386-6947-1/18/$31.00 ©2018 IEEE

[5] Yuriy Zhuravlev, Alexander Dokukin, Oleg Senko, Dmitry Stefanovskiy, "Use of Clusterization technique to highlight groups of related goods by digital traces in retail trade",978-1-7281-0450-8/19/$31.00 ©2019 IEEE

[6] Y. Zuo, Y. Kajikawa, and J. Mori," Application of Network Analysis Techniques for Customer In-store Behavior in Supermarket" 978-1-5090-3294-5/17/$31.00 ©2018 IEEE

[7] Prem Kumar G, Sushruth B Bangre, Kavya M, Varun M, Anupama R Smart-Cart for Smart-Cities$31.00 ©2018 IEEE

[8] Petro Pukach 1, Khrystyna Shakhovska,2The Mathematical Method Development of Decisions Supporting Concerning Products Placement Based on Analysis of Market Basket Content$31.00 © 2017 IEEE



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY