SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

DATA MINING TECHNIQUE TO FIND LOW-COST GROCERIES

Amrutha B[*1], Muthuramalingam B[*2]
[*1]*Dept. of MCA, Sir M Visvesvaraya Institute of Technology, Bengaluru, India.*
[*2]*Dept. of MCA, Faculty of MCA, Sir M Visvesvaraya Institute of Technology, Bengaluru, India.*

ABSTRACT
An e-store, internet shop, web shop, online store, or virtual store is the physical equivalent of purchasing goods or services from a real business or shopping center. This is known as business-to-consumer (B2C) online shopping. The shopping cart project must design a shopping cart system to organize product data and other consumer information. When a customer adds a product to their cart, the algorithm compares data from other stores to find low-cost alternatives. Sentimental analysis is used to determine the pricing of the items in the cart. It compares the prices of each product across different stores and delivers a result that allows you to acquire everything for a low price. The data is evaluated using the Naive Bayes classifier, which looks at each product in the dataset and assesses if it is being purchased at a low cost.
KEYWORDS: *Naive Bayes classifier, Sentimental analysis, Low-cost prediction.*

I. INTRODUCTION

Online grocery shopping has greatly increased in recent years. In the current system, we can purchase products from a variety of sellers, such as e-bay, Amazon, and Flipkart, which are all associated marketing websites. Customers can purchase merchandise from a variety of vendors. However, there is no effective suggestion system in place to locate a low-cost product in the current system. Everyone wants to get a good deal on a product. People want to save money therefore they can't check each store to see what things are available at a low price. The primary goal of online grocery is to create and develop new models, as well as to optimize relationships between grocery stores and their customers. Changing from supermarket to online grocery shopping can increase the productivity of an online grocery retailer by shortening supply chains, lowering overhead costs, and enabling "just in time" service. Most grocery store chains provide online shopping with free in-store pick up, and a few even provide home delivery for free. In this study, the developers create an online grocery system that will reduce the customer's effort and time spent wandering around, as well as the proposed system will be used to identify the product that can be brought for a low cost.

The project is being designed to eliminate a time-consuming and inconvenient system in which customers must go to a store or shop and select products at a lower price in a neighboring store. This approach is more cost-effective and saves a lot of money for the user. The program evaluates the prices of various things in various stores and determines the store where all of the items are available for a lower price.

To locate a low-cost grocery store, sentimental analysis is being applied. This research is based on a dataset that includes all of the stores in the area. The Naive Bayes classifier algorithm is used to evaluate the product's low cost. This analyses each product in the various shops and provides an accurate result indicating which store has the product with the lowest cost.

II. METHODOLOGY

A. SENTIMENTAL ANALYSIS

Sentiment analysis is textual mining that discovers and extracts subjective information from the source material, allowing an organization to understand the social sentiment of its brand, product, or service while monitoring online conversations. In this designed system the analysis is made on the set of stores. With the sentimental analysis, the nearest available store which contains the grocery can be found.

B. NAIVE BAYES CLASSIFIER

The Bayes theorem encompasses the Naive Bayes classification technique. It is one of the most basic supervised learning algorithms. The Naive Bayes classifier algorithm is fast, accurate, and consistent. On large datasets, Naive Bayes classifiers perform well in terms of accuracy and speed.

The Naive Bayes classifier is based on the assumption that the effect of one feature in a class is independent of the effects of other features. The Naive Bayes classifier calculates the probability of an event in the following steps:

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

1. Calculate the prior probability for given class labels
2. Find the Likelihood probability with each attribute for each class
3. Put these values in Bayes Formula and calculate posterior probability.
4. See which class has a higher probability, given the input belongs to the higher probability class.

III. MODELING AND ANALYSIS

1. SYSTEM ARCHITECTURE

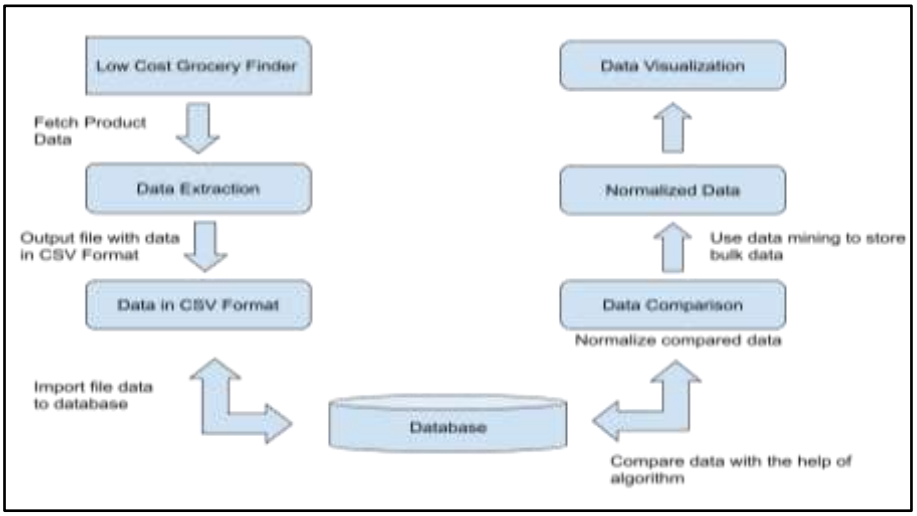

Figure 1: System Architecture

Product prices are set by retailers depending on production and transport expenses. This tool picks a few of the competitor's product prices and uses data extraction to retrieve product information such as titles, categories, and prices. There are several ways to store the data that has been retrieved. To retrieve the data and store it in a database, an open-source browser-based extension is employed in this study. The data were then normalized by importing them into a MySQL database. The normalizing procedure separates the relevant or required data from the redundant data and stores it in a separate table. The data is normalized and stored in a database table by comparing the product title and category. Users can use the saved information to compare the prices of different products. To evaluate the product, a comparison technique such as the Naive Bayes algorithm can be used to visualize the data.

2. DATA FLOW DIAGRAM

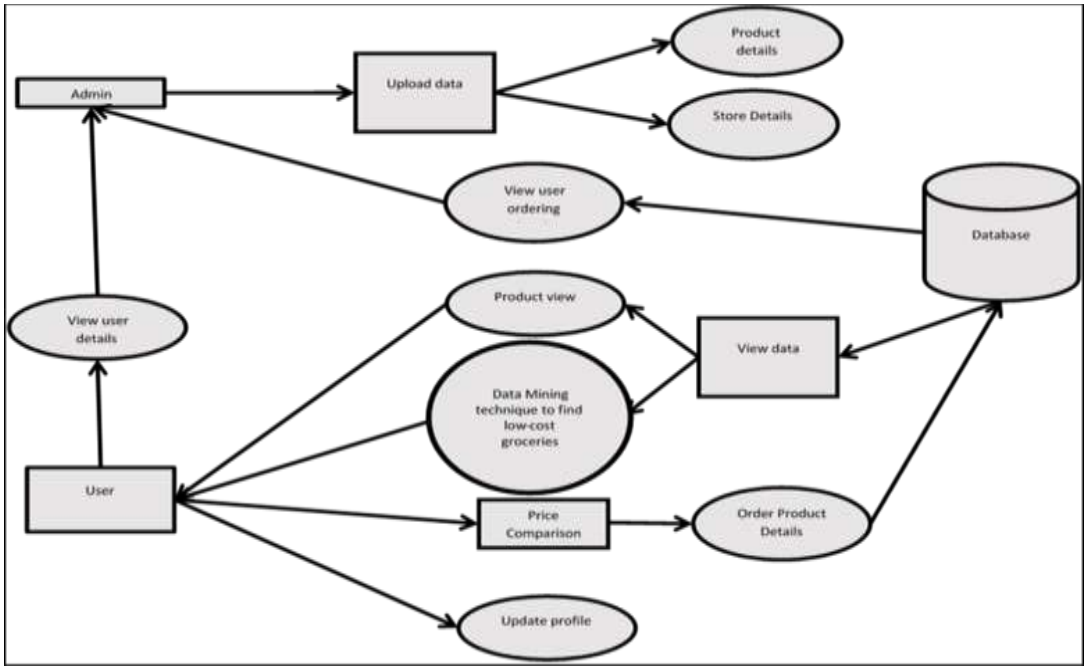

Figure 2: Data flow diagram

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

The project is in its base case is divided into two modules the administrator and the user the process that takes place in the system is the by fetching the data from the database and applying data mining techniques to find the low-cost grocery. The admin is responsible to update the data and the user views the information of the product after the data mining technique is applied.

IV. RESULTS

A. Home page

Figure 3: Home page

The admin and user login links are located on the home page.

B. Admin login page

Figure 4: Admin Login

Following the home page, the admin login page is accessed. To access the admin page, the administrator enters his user name and password.

C. Admin management page

Figure 5: Admin Management page

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

This page contains all of the products that have been entered by the administrator as well as the registration information of the users. The admin can also view the user's order details.

D. User login page

Figure 6: User Login

User login page to access the user required functions.

E. User price prediction page

Figure 7: Admin viewing order details of the customer

This page displays the orders placed by users from various stores, which are added to the total amount. When the user clicks the predict button, the software analyses the items in the list and estimates the lowest cost.

F. User page after the prediction of the lowest cost

Figure 8: Predicted lowest price

This shows the result page that predicts the total lowest cost predicted by the software.

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

G. Final order details in user page

Figure 9: Final order details in user page

Following the results page, which includes a button to view the order details page includes the location of the nearest store as well as the total cost of all the products in the cart.

V. CONCLUSION

Online grocery services satisfy a variety of customer demands, such as selling products for niche markets or assisting the time-crunched consumer in shopping for humdrum weekly groceries. The homebound elderly and handicapped can join in the shopping experience by having things delivered to their homes. Despite a significant decrease in the number of pure-play online retailers, there appears to be a strong market for online shopping. The fundamental business model that is now in place necessitates the assistance of existing brick-and-mortar supermarkets. Since of the perishable nature of many supermarket products, this strategy is efficient because it produces distribution efficiency and leverages reputation, which is an essential concern for consumers.

This application is mostly used to identify the products that we obtain at a reasonable cost, and our application is primarily used to bridge the gap between the consumer and the seller from various sellers. This will mostly be used to identify the product via sentimental analysis, and it will be used to decrease time and boost performance while decreasing analyzing time and saving money for each customer.

VI. REFERENCES

1. *Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. Social networks, 25(3), 211-230.*
2. *Backstrom, L., & Leskovec, J. (2011, February). Supervised random walks: predicting and recommending links in social networks. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 635-644). ACM.*
3. *Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., & Suri, S. (2008, August). Feedback effects between similarity and social influence in online communities. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 160-168). ACM.*
4. *Dong, Y., Zhang, J., Tang, J., Chawla, N. V., & Wang, B. (2015, August). Coupledlp: Link prediction in coupled networks. In Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 199-208). ACM.*
5. *Golbeck, J. A. (2005). Computing and applying trust in web-based social networks (Doctoral dissertation).*