International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Grocery Recommendation System

Avik Ghosh[1], Pankaj Gupta[2], Mahavir Deora[3], Allan Lopes[4]

[1]Avik Ghosh, Information Technology, Mumbai University, Mumbai, India.
[2]PankajKumar Gupta, Information Technology, Mumbai University, Mumbai, India.
[3]Mahavir Deora, Information Technology, Mumbai University, Mumbai, India.
[4]Prof. Allan, Lopes Assistant Professor of Universal College of Engineering, Information Technology, Mumbai University Mumbai, India.

Abstract -*The online store is growing very fast. An online store helps people buy the product they want online. Competitive competition among Online Store providers is fueling technological advances. Many Online Store programs not only display a product but also require support for the selection of appropriate products to attract website visitors. As a result, many website visitors are confused when it comes to purchasing products in the Online Store. The amount of product variation that a customer provides when purchasing goods is sometimes more than one product. The problem leads to the idea of establishing a product recommendation system. Historical information from visitors and customers can be used to analyze user needs and product preferences. An organization rule using Apriori information will be able to capture user preferences. By targeting user preferences, a valid product recommendation can be created. This study will analyze the rules in the purchase history data from Online Store visitors to get product recommendations to be displayed. According to the test result, association law is able to produce specific recommendations with a confidence value of 76.92%.*

Key Words: Recommendation System, Apriori, Web Usage Mining, Association Rule.

1.INTRODUCTION

The development of information technology in the field of communication has at least two rapidly evolving technologies. The first is a mobile phone and the second is a computer in the network. Mobile phones are used not only for calling and text messaging but have been developed into detailed devices for whatever we know best as smartphones.

Business processes in the commercial sector cannot be different from the use of Android-based smartphones. is well separated from the operating system. Most of the leading mobile retailers currently use Android as an app (OS) for their smartphones. Android has evolved into the most widely used operating system on smartphones just a few years after its advent However, user privacy remains an important part of every mobile app[4].

Shoes, which are one of man's needs, have become increasingly used over time, and have become an obligation in various fields of work, education and fashion-ion practice that has increased the number of models and types of shoe year by year. you want to buy new shoes, they need to roam looking for a nearby shoe store to sell their favourite shoes. mobile app based on android shoes. Android is preferred because it has many advantages over platforms (desktop, website) and applications (OS) Smartphones (such as iOS, Windows Phone, Symbian)[7].

The advantages of Android include a large share of the market, open-source, easy to use compared to other platforms and OS, and easy to use anytime and anywhere. The mobile app with the redesigned algorithm will help the user to be selective. There are many algorithms that can be used to make recommendations, during this case, shoe removal. one among them is Apriori. This study uses Apriori due to its benefits in making selected materials and testing them on a daily basis[6].

Subsequent applications are often employed by customers to see product recommendations, prices, and store shoe details which can be customer redesign within the selection of desired shoes. A product recommendation system with the Apriori algorithm will increase the app. the merchandise redesign program allows applications to display products that users might not like, thus reducing product search time. Therefore, users can make their choice supported the practice of shoe users. consistent with the edition, the choice are often accelerated in order that they will maintain their resources[5].

1.1 Existing System.

Today, E-commerce sites use promotional programs on an outsized scale to grow their business. Products are often recommended supported the extent of total sales in reference to the location , in terms of buying strategies, or in terms of additional customer customer behavioural analysis, as a prediction of inauspicious shopping behaviour. generally , these methods are a mirrored image of the customization of a site that gives its customers. This method is employed by retailers round the world to work out what items are purchased together. and that they faced the primary cold problem, namely. 1) How are you

able to recommend a replacement user? 2) How are you able to recommend new things? It also provides recommendations supported location of user interests, customer searches and suggests products supported it. Amazon or Flip cart uses user view data eg if any customer or user searches for a product from a specific category system suggests a product forms an identical category. And supported current user search, the location recommends products. Every user visiting the location won't purchase the merchandise . they will bypass it and are supported those search results with a recommendation product.

1.2 Proposed System.

In our proposed plan, we decide to use a way that uses transaction history. As Apriori is meant to figure with data that contains transactions and compliance rules, while using the "bottom-up" method, where regular subsets are extended to at least one item at a time and groups of nominees (a candidate set contains all sets of frequency item) are tested against data. Algorithm terminates where no simpler extensions are found. The structural design of the proposed system is shown within the figure. This approach eliminates the initial problem of creating recommendations for brand spanking new customers. Since the proposed system uses transaction information, you'll not need a customer profile to recommend products[5].

The recommendation System uses the Apriori algorithm which may be used with advanced functionalities: Visual Interface, Data Extraction, Web Usage, and Pattern Visibility.

Before you start formatting your paper, first write and save the content as separate text. Complete all organizational content and layouts before formatting. Please note the sections A-D below for more information on grammar, spelling and grammar. [7]

Keep your text and files of different images until the text is formatted and styled. do not use difficult tabs, and restrict the use of a hard return to a single restore at the top of the section. do not add any paper extensions inside the paper. don't count the captions - the template will do this for you.

A. UI Components
This module communicates between users and therefore the web mining system. Allows the user to interact with the system by specifying a question or web function within the mines, providing details to assist specialise in usage, and performing web-based mining operations supported mineral web usage results[7].

B. Data Extraction.
Data extraction is that the act or process of retrieving data (usually unstructured or poorly structured data) out of knowledge sources for further processing or data storage (data migration). In our proposed system we are extracting data from database on which we are applying the Apriori algorithm.

C. Web Usage Mining.
Web mining is achieved first by reporting visitor details to log server files and another source of traffic data. blog file is differently to gather Web traffic data. After the acquisition of Web traffic data, it are often combined with other related data, where data processing techniques are used for specific data processing techniques like organization rules, mergers etc[6].

D. Pattern Recognition.
Pattern recognition or testing is employed to spot really interesting patterns that represent information supported specific steps of interest. This section usually uses interesting methods and links to Web Mining Modules to focus search on interesting patterns. the advice program should be ready to display patterns obtained in some ways , like domains, tables, pies or charts of bars, cubes or other visual representations. In our Recommendation Programs we offer product recommendations with product images. It can use interesting thresholds to filter the patterns found. Alternatively, the pattern test module are often integrated with the mining module, counting on the implementation of the info mining method used[4].

2. APRIORI ALGORITHM:

Apriori is part of the organization's governance system, which works to obtain a combination of items based on customer-driven items. Types of assembly rules include a priori, standard rule input, and hash-based algorithms. Mining Association Rule is a way to create data for obtaining rules for merging between compounds, for example, store tracking analysis. With the availability of data and visuals, it can be known about a customer's potential to buy bread and milk[2]. By using this scenario, service providers can use these scenarios by controlling the placement of goods or designing marketing strategies. All non-empty subset of frequent itemset must be frequent. The key concept of Apriori algorithm is its anti-monotonicity of support measure. Apriori assumes that. All subsets of a frequent itemset must be frequent(Apriori propertry). If an itemset is infrequent, all its supersets will be infrequent[3].

TID	items
T1	I1, I2 , I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

Minimum support count = 2

minimum confidence = 60%

Step-1: K=1

(I) Create a table containing support count of every item present in dataset – Called C1(candidate set)

Itemset	sup_count
I1	6
I2	7
I3	6
I4	2
I5	2

(II) Compare candidate set item's support count with min support count(here min_support=2 if support_count of candidate set items is less than min_support then remove those items)[2]. This gives us itemset L1.

Itemset	sup_count
I1	6
I2	7
I3	6
I4	2
I5	2

Step-2: K=2

- Generate candidate set C2 using L1 (this is called join step). Condition of joining Lk-1 and Lk-1 is that it should have (K-2) elements in common.
- View all itemset subsets that occur frequently or not and otherwise always delete that set item. (For example a set of {I1, I2} is {I1}, {I2} that happens frequently. Check each item)
- Now find the calculation of support for these itemsets by searching the database [3].

Itemset	sup_count
I1,I2	4
I1,I3	4
I1,I4	1
I1,I5	2
I2,I3	4
I2,I4	2
I2,I5	2
I3,I4	0
I3,I5	1
I4,I5	0

(III) Compare candidate (C2) support count with minimum

support number (here min_support = 2 if the support_count item set is less than min_support and subtract those items) this gives us itemset L2.

Itemset	sup_count
I1,I2	4
I1,I3	4
I1,I5	2
I2,I3	4
I2,I4	2
I2,I5	2
I2,I5	2

Step-3:

- Generate C3-set candidate using L2 (joining step). The condition for joining Lk-1 and Lk-1 is that it must have the same (K-2) properties. So here, with L2, the first thing should be the same. Therefore the element produced by joining L2 is {I1, I2, I3} {I1, I2, I5} {I1, I3, i5} {I2, I3, I4} {I2, I4, I5} {I2 , I3, I5}
- See if all subsets of these item sets are standard or not and if not, then delete that set item. (Here the sub-set of {I1, I2, I3} is {I1, I2}, {I2, I3}, {I1, I3} which are common. In {I2, I3, I4}, the sub-set of { I3, I4} rarely remove it. Similarly check the entire item)
- find support calculations for these remaining assets by searching in the database [3].

Itemset	sup_count
I1,I2,I3	2
I1,I2,I5	2

(II) Compare the number of candidate candidates (C3) with the minimum number of support (here min_support = 2 if the support_count of the candidate set item is less than min_support and subtract those items) this gives us an item set..

Itemset	sup_count
I1,I2,I3	2
I1,I2,I5	2

Step-4:

- Generate a candidate set C4 using L3 (join step). The condition for joining Lk-1 and Lk-1 (K = 4) is, they must have the same (K-2) properties. So here, for L3, the first 2 items (items) must match.
- See all subsets of these common metals or not (Here the itemset formed by joining L3 is {I1, I2, I3, I5} so its subset contains {I1, I3, I5}, unusual).
- So there is nothing in C4 Stop here because there are no regular purchases available.

Thus, we have discovered all the frequent item-sets. Now generation of strong association rule comes into picture. For that we need to calculate confidence of each rule.

Confidence:– A confidence of 60% means that 60% of the customers, who purchased milk and bread also bought butter.Confidence(A->B)=Support_count(A∪B)/Support_count(A)

So here, by taking the example of any common property item, we will show the ruling generation. Itemset Items {I1, I2, I3} // from L3 SO rules can be [I1 ^ I2] => [I3] // confidence = sup (I1 ^ I2 ^ I3) / sup (I1 ^ I2) = 2/4 * 100 = 50%

[I1 ^ I3] => [I2] // confidence = sup (I1 ^ I2 ^ I3) / sup (I1 ^ I3) = 2/4 * 100 = 50%

[I2 ^ I3] => [I1] // confidence = sup (I1 ^ I2 ^ I3) / sup (I2 ^ I3) = 2/4 * 100 = 50%

[I1] => [I2 ^ I3] // confidence = sup (I1 ^ I2 ^ I3) / sup (I1) = 2/6 * 100 = 33%

[I2] => [I1 ^ I3] // confidence = sup (I1 ^ I2 ^ I3) / sup (I2) = 2/7 * 100 = 28%

[I3] => [I1 ^ I2] // confidence = sup (I1 ^ I2 ^ I3) / sup (I3) = 2/6 * 100 = 33% So if the lowest confidence is 50%, first

The 3 rules can be considered as strict organizational rules.

3. LIMITATIONS OF APRIORI ALORITHM:.

[3] Apriori Algorithm can be slow. The main limitation is time required to hold a vast number of candidate sets with much frequent itemsets, low minimum support or Large itemsets i.e. not working well approach for large number of datasets. For example, if there are 10 ^ 4 from always 1- shopping items, requires to generate more than 10 ^ 7 candidates into 2-length which respectively will be checked and accumulate. Moreover, to detect always pattern by size 100 i.e. v1, v2… v100 it have to generate 2^100 candidate itemsets that yield on costly and wasting of time of candidate generation. So, it will check for many sets from candidate itemsets, also it will scan database many times repeatedly for finding candidate itemsets[2]. Apriori will be very low and inefficiency when memory capacity is limited with large number of transactions.

4. ASSOCIATION RULE MINING:

Association Mining searches for frequent items in the data-set. In frequent mining usually the interesting associations and correlations between item sets in transactional and relational databases are found. In short, Frequent Mining shows which items appear together in a transaction or relation[2].

Need of Association Mining: Frequent mining is generation of association rules from a Transactional Dataset. If there are 2 items X and Y purchased frequently then its good to put them together in stores or provide some discount offer on one item on purchase of other item. This can really increase the sales. For example it is likely to find that if a customer buys Milk and bread he/she also buys Butter. So the association rule is ['milk]^['bread']=>['butter']. So seller can suggest the customer to buy butter if he/she buys Milk and Bread[3].

IMPORTANT DEFINATIONS

1. Support : It one among the measure of interestingness. This tells about usefulness and certainty of rules. 5% Support means total 5% of transactions in database follow the rule. Support(A -> B) = Support count(A ∪ B)
2. Confidence: A confidence of 60% means 60% of the purchasers who purchased a milk and bread also bought butter.
   - Confidence(A -> B) = Support count(A ∪ B) / Support count(A)
   - If a rule satisfies both minimum support and minimum confidence, it is a strong rule[2]. Support count(X) : Number of transactions in which X appears. If X may be a union B then it's the amount of transactions during which A and B both are present.
3. Maximal Itemset: An itemset is maximal frequent if none of its supersets are frequent.
4. Closed Itemset: An itemset is closed if none of its immediate supersets have same support count same as Itemset.

5. K- Itemset: Itemset which contains K items may be a K-itemset. It can therefore be said that aitetet occurs more often when the corresponding support value is greater than the lower support value [1].
6. Example of Availability of Frequent Objects - Consider the database provided for the given functions [2].

TransactionId	Items
1	{A,C,D}
2	{B,C,D}
3	{A,B,C,D}
4	{B,D}
5	{A,B,C,D}

- Suppose the minimum support number is 3
- Relationship capture is too high => closed => always.

1- always:

{A} = 3; // is not closed due to {A, C} and not size

{B} = 4; // is not closed due to {B, D} and no higher

{C} = 4; // is not closed because {C, D} is not high

{D} = 5; // object set as not set faster than the same number [2]. Not the top

2-always:

{A, B} = 2 // is rare because of the support value <minimum support number so ignore

{A, C} = 3 // is not closed due to {A, C, D}

{A, D} = 3 // is not closed due to {A, C, D}

{B, C} = 3 // is not closed due to {B, C, D}

{B, D} = 4 // is closed but not large due to {B, C, D}

{C, D} = 4 // is closed but not large due to {B, C, D}.

3-always:

{A, B, C} = 2 // ignores not always because support number <minimum support number

{A, B, D} = 2 // Ignore is rare because the number of support <minimum support number

{A, C, D} = 3 // maximal frequency

{B, C, D} = 3 // very high

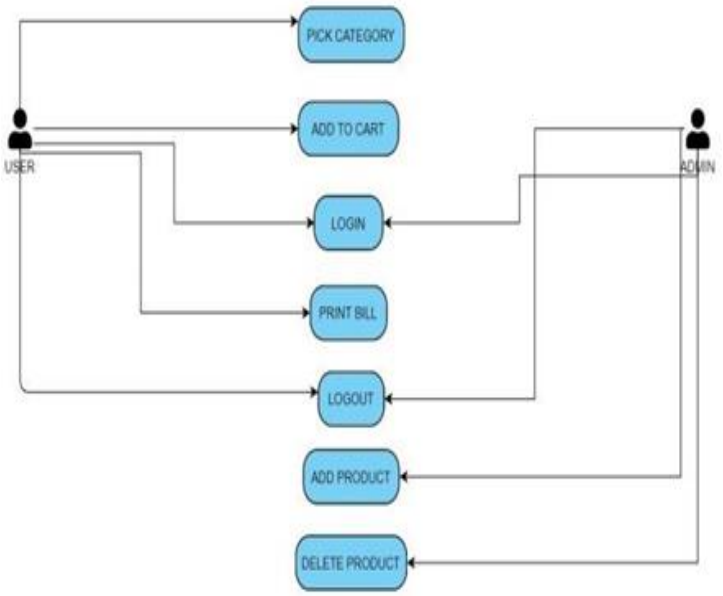
4-frequent:

{A, B, C, D} = 2 // Ignoring is not uncommon

</ [2].

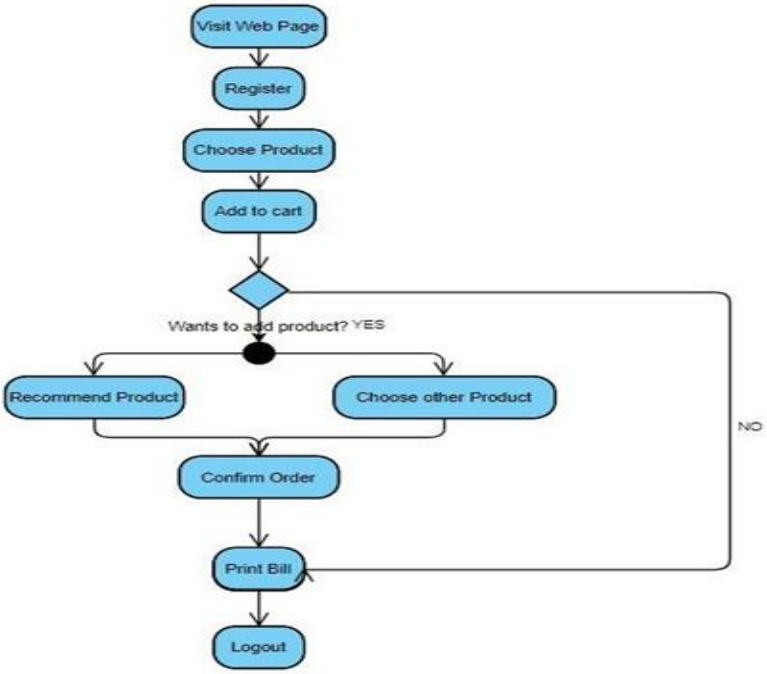
5. DISCUSSION:

1. Use case diagram:

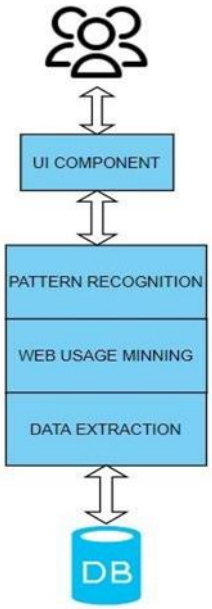


2. Activity Diagram:

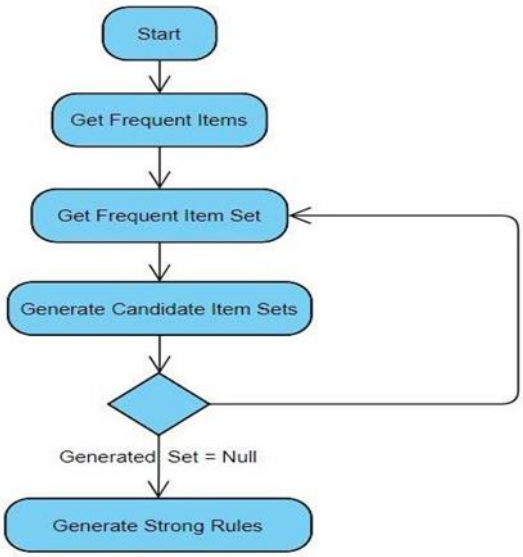
An activity diagram is a diagram that describes the flow of functionality from the system. The types of activity diagrams this time are registers, logins, change passwords, log out, add advertisements, delete ads, add categories, delete categories, and activity diagrams select categories[1].



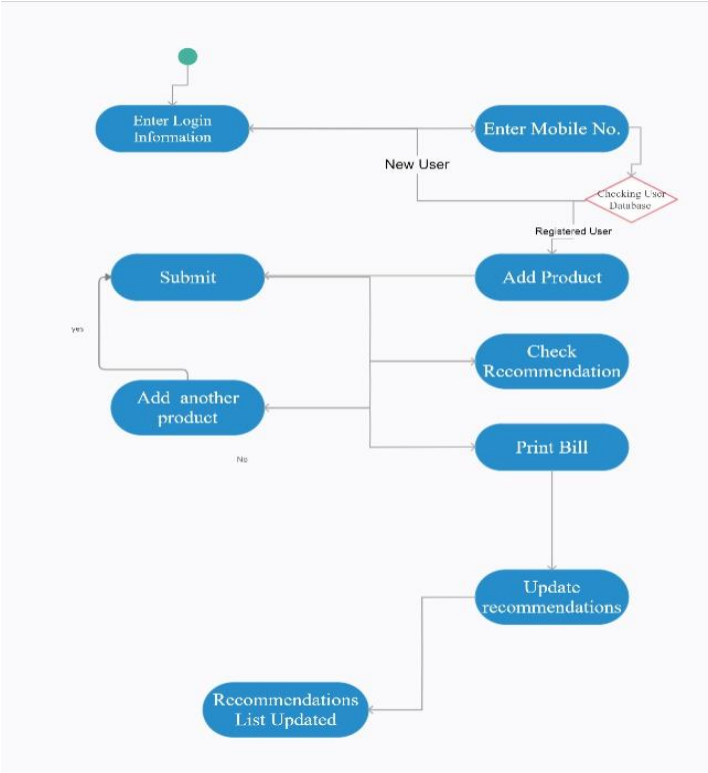
3. System Diagram:



4. System_flow Diagram:

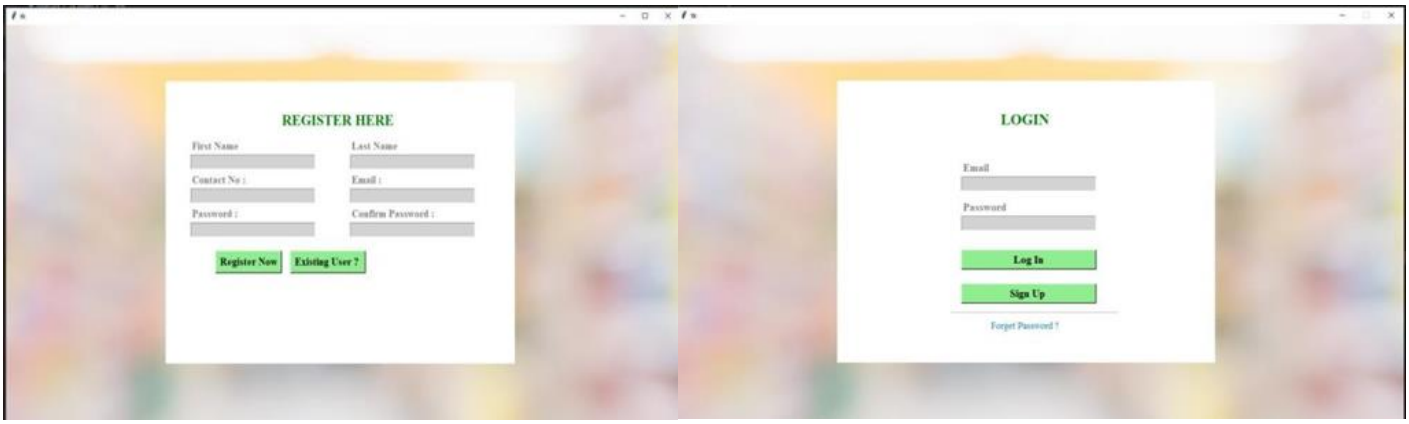


5. UML diagram:

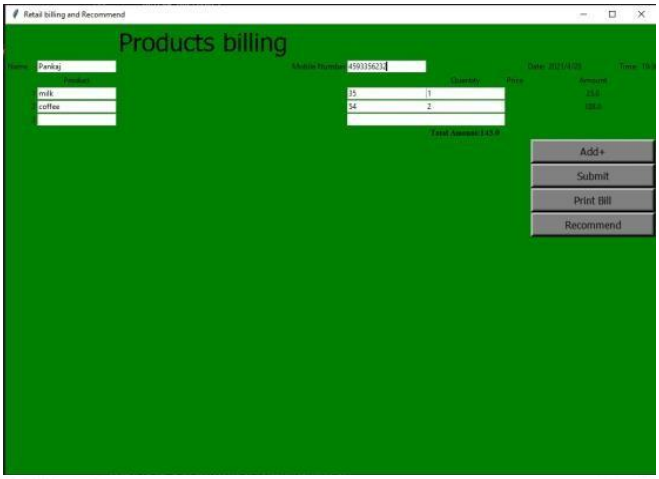


6. IMPLEMENTATION:

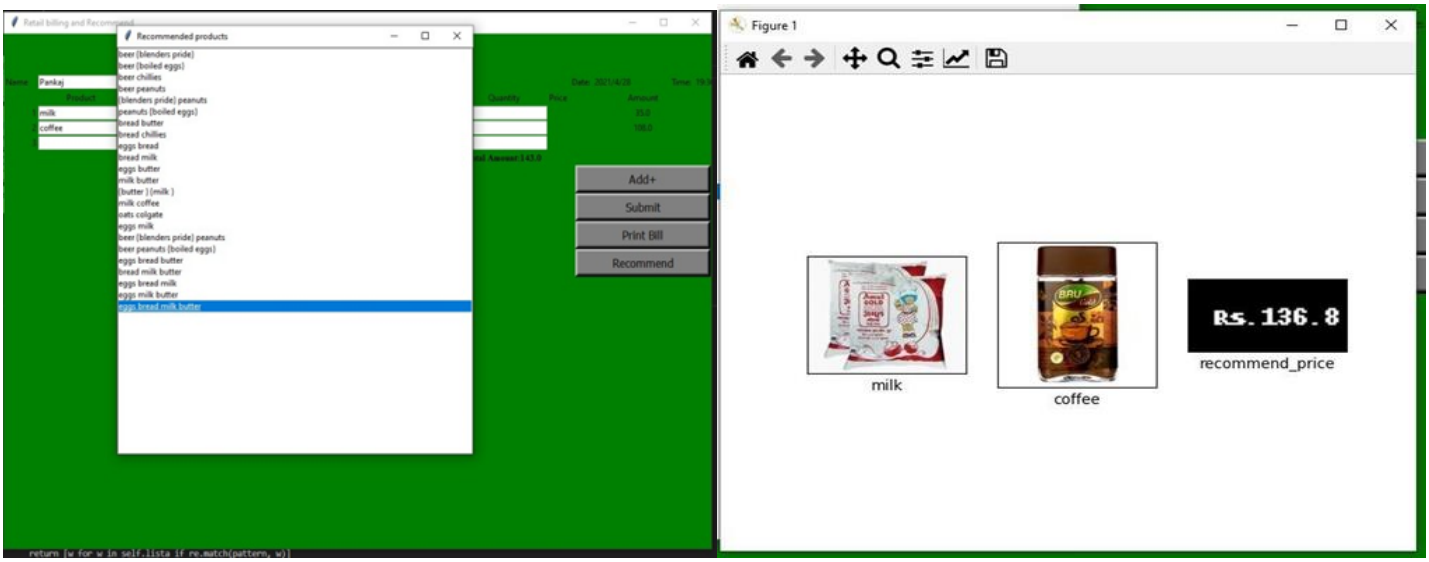
Registration and Login page.



Main Page and Billing page:



Recommended Products:



This is an extra features, by picking up items from the recommended products, we give certain discount on the products.

6.1 Testing:

To find out the performance of the app and the algorithm created, and then test it using the Whitebox and Blackbox method. Whitebox used on the test based on manual calculation and calculation of the app as shown, as well as the Blackbox test used to measure app performance and displays.

6.2 Results:

The testing conducted by us yielded an accuracy of 91.66% where the dataset included 70+ images. One item will recommend around 5+ related items.

7. FUTURE SCOPE:

The Future Recommendation Plan has a strong future for E-commerce on the web. In the coming years this commendation program will be used in various locations to recommend various services as well. In the future the algorithm can be expanded to mine web content, web design etc. Work can also be extended to extract data from image files.

8. CONCLUSIONS:

Based on the discussion of the authors' research, it can be concluded that using an apriori algorithm, the system provides product recommendations to online store customers based on the reliability of a combination of products purchased during a given. The use of the Apriori Method in this study is to find a large combination of material based on transaction data and create an integration pattern for the object combination. This research can be further developed using a weighted product method to determine the weight of each product so that it can provide an alternative product of the highly recommended product.

ACKNOWLEDGEMENT:

This work was supported by Universal College of engineering and our guide Mr. Allan Lopes, Professor Universal college of engineering. We are thankful to our guide for helping us in the journey to reach our destination with flying colors.

REFERENCES:

[1] Aprizal, Hasriani, and Wahyu N2015Implementasi Data Mining Untuk Menentukan Posisi Barang Pada Rak Menggunakan Metode Apriori Pada PT. Midi Utama Indonesia(Techno. COM) Vol15p 335-342.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol.2. Oxford: Clarendon, 1892, pp.68–73.

[2] EzhilarasanCand RamaniS2017Performance Prediction using Modified Clustering Techniques with Fuzzy Association Rule Mining Approach for Retail2017 International Conference on Intelligent Computing and Control

[3] ItoYand KatoS2016An Apriori-Based Approach to Product Placement in Order PickingInternational Conference on Agents.

[4] Sano, N.,Yada, K.. The influence of sales areas and bargain sales on customer behavior in a grocery store.Neural Computing and Applications2015;26(2):355–361..

[5] P. Resnick and H. R. Varian, “Recommender systems,” Communica- tions of the ACM, vol. 40, no. 3, pp. 56–58, 1997

[6] Y. Jararweh, S. Alzubi, and S. Hariri, “An optimal multi-processor allocation algorithm for top performance gpu accelerators,” in 2011 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), pp. 1–6, Dec 2011.

[7] C. Castro-Herrera, C. Duan, J. Cleland-Huang, and B. Mobasher,

[8] “Using data mining and recommender systems to facilitate large-scale, open, and inclusive requirements elicitation processes,” in International Requirements Engineering, 2008. RE’08. 16th IEEE, pp. 165–168, IEEE, 2008.