Prof.Mr.Atul Pawar, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 11, Issue 12, (Series-I) December 2021, pp. 19-24

RESEARCH ARTICLE OPEN ACCESS

A Review on Grocery Management System Using Machine Learning Algorithms

Prof.Mr.Atul Pawar[1], Sushank Dehankar[2], Akshay Dandawate[3], Omkar Bansode[4] , Avishkar Gawade[5]

[1]*Associate Professor, Department of Computer Engineering ,Pimpri Chinchwad College of Engineering,Savitribai Phule Pune University, INDIA,*
[2]*Student,Department of Computer Engineering ,Pimpri Chinchwad College of Engineering,Savitribai Phule Pune University, INDIA,* [3]*Student, Department of Computer Engineering, Pimpri Chinchwad College of Engineering,Savitribai Phule Pune University, INDIA,*
[4]*Student, Department of Computer Engineering, Pimpri Chinchwad College of Engineering,Savitribai Phule Pune University, INDIA,*
[5]*Student, Department of Computer Engineering, Pimpri Chinchwad College of Engineering,Savitribai Phule Pune University, INDIA,*

ABSTRACT
Smartphones have become acrucialpart of human life. As we know ,most people faced problem of forgetting some of the groceries that they need to buy after visiting the grocery store. It might be a small bag of condiments, a tube of toothpaste, or even a packet of biscuits. Also, sometimes there is buying of unwanted items that we may never use that leads to wastage. So Digital Grocery lists are the best replacement for paper-based grocery System. This paper discusses about the Machine Learning Algorithms in the development of a mobile application for creating and managing a grocery list collaboratively. This Systemwill helps us to generate, manage and maintain the required monthly grocery list that can be customized based on grocery requirements for any period of time. The evaluation suggested that this application is sufficiently usable in terms of -ease of use, satisfaction, time-saving and the future development of the household needs. The main purpose of this review paper is to trackthe monthly consumption of grocery products, calculate the consumption pattern of customer and help us by giving a recommendation of essential groceries. Algorithms studied in this paper are Apriori Algorithm,Maxminer Algorithm, Support Vector Machine and Random Forest Algorithm.

Keywords—Algorithms, Machine Learning ,Patterns Statistical

Date of Submission: 01-12-2021 Date of Acceptance: 15-12-2021

I. INTRODUCTION

Nowadays, everyone tends to shop online. as it is time and energy saving. Online grocery systems are helping people, provides access to many items, and helps to purchase them from home. So retailers are entering the user environment. Most traditional Shopping Methods are time-consuming and don't provide the variety of choices in the product selection which has led to a lack of consumer satisfaction.Whenever I go with the family to the star bazaar we land up into buying extra grocery than the required Quantity .Due to Which Some part doesn't get consumed before the expiry date and which Leads to Wastage of the Product. So these things can be eliminated with help of the online grocery prediction system .

On an online platform, grocery items are in hundreds , and making a choice may not be easy for customers. This is where the grocery recommendation system helps the user. Most e-commerce websites use recommendation systems that are based on specific algorithms. Some personalization algorithms are the basis of these recommendations systems. These algorithms suggest items that are likely of interest to the customer. Collaborative filtering and content-based filtering algorithms are mostly used to develop such recommendations system. But these systems may have some disadvantages and limitations. To overcome such problems, a hybrid recommendation system can be used.

We aim to study different Machine learning algorithms and create a user-friendly system that

will provide customer satisfaction.Machine learning is a method of data analysis that automates analytical model building. It is a branch of Artificial Intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

In this paper, we have compared some machine learning algorithms which are Random Forest Algorithm, Support Vector Machine, and Apriori Algorithm on various different parameters Like Area Under ROC Curve, Cross-Entropy Loss, Root Mean Square Error and Kernal Functions.

The proposed system will recommend items based on the user's previous purchase history, and by considering factors like previous buying quantity and period(item is bought once a week or once a month).

II. METHODOLOGY

2.1. APRIORI ALGORITHM

- Apriori Algorithm is given by R. Agrawaland R. Srikant in 1994. It is used for finding frequent itemsets in a dataset for the boolean association rule.
- Definition ;

" All subsets of a frequent itemset must be frequent(apriori property).

- If an itemset is not frequent, all its supersets will be not frequent."
- Limitation :

a) The algorithm scans the database too many times, which reduces the overall performance.
b) Due to this, the algorithm assumes that the database is permanently in the memory.
Time complexity: O(2D) where D is dataset size

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

- Pseudocode: The pseudo-code for the algorithm is given below for a transaction database {\displaystyle T}T, and a support threshold of {\displaystyle \varepsilon }\varepsilon. Usual set theoretic notation is employed, though note that {\displaystyle T}T is a multiset. {\displaystyle C_{k}}C_{k} is the candidate set for level {\displaystyle k}k. at each step, the algorithm is assumed to generate the candidate sets from the large itemsets of the preceding level, heeding the downward closure lemma. {\displaystyle \mathrm {count} [c]}{\displaystyle \mathrm {count} [c]} accesses a field of the data structure that represents candidate set {\displaystyle c}c, which is initially assumed to be zero. Many details are omitted below, usually, the most importantpart of the implementation is the data structure used for storing the
candidate sets, and counting their frequencies.

Code :

```
Apriori(T, ε)
   L1 ← {large 1 - itemsets}
   k ← 2
   while Lk−1 is not empty
      Ck ← apriori_gen(Lk−1, k)
      for transactions t in T
         Dt ← {c in Ck : c ⊆ t}
         for candidates c in Dt
            count[c] ← count[c] + 1

      Lk ← {c in Ck : count[c] ≥ ε}
      k ← k + 1

   return Union(Lk)

apriori_gen(L, k)
    result ← list()
    for all p ⊆ L, q ⊆ L where p1 = q1, p2 = q2, ...,
pk-2 = qk-2 and pk-1 < qk-1
       c = p ∪ {qk-1}
       if u ⊆ c for all u in L
          result.add(c)
    return result
```

- **Working :**

Given :
Min support = 2
(It means min frequency of the item to get selected is 2times)
TID = Transaction ID
Items = Products Purchaesd

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

Step 1: It will count each item is repeated
and will eliminate the item that's. The count is less than 2 (min frequency)

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

Step 2: a) It will form pair of items and will count each pair that is repeated.
b)and will eliminate the pair whose count is less than 2 (min frequency)
c) Here (I1,I4) ,
(I3,I4) , (I3,I5) , (I4,I5)are eliminated as their count is less than 2

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

Step 3: a) It will form a group of 3 items and will count each group
that is repeated.
b) and will eliminate the pair whose count is less than 2 (min frequency)

Step 4: a) It will form a group of 4 items
and will count each group that is repeated.
b) and will eliminate the pair whose count is less than 2 (min frequency)
c) Here only T8 has 4 items
Thus count is 1 ie < 2
So , it is excluded
Result :

- So (I1,I2,I3) & (I1,I2,I5) are the most frequent item sets.
- Thus we have successfully found the most frequent item bought by you.
- Limitation :
- Apriori Algorithm is slow when dealing with a large no. of sets.

To overcome the problem **"Max-Miner "** Algorithm is used as it deals directly with existence of products in respective Transaction IDs rather than counting individually.

2.2.MAX-MINER ALGORITHM

- Max-Miner has a similar implementation to the Apriori with the difference being that Max Miner Algorithm but only concerns itself with finding the frequent maximal itemsets and not all frequent itemsets.
- Max-Miner provides reliable results even with a low Support by attempting to "look ahead" in order to quickly identify long frequent patterns.
- By doing this in the early stages, max-miner is able to prone all subsets of the long frequent pattern from further consideration.

- Definition :

GROUP OF 3	COMMON	COUNT
I1,I2,I3	T8,T9	2
I1,I2,I5	T1,T8	2
I1,I3,I5	T8	1 (Eliminated <2)
I2,I3,14	0	0 (Eliminated <2)
I2,I3,I5	T8	1 (Eliminated <2)
I2,I4,I5	0	0 (Eliminated <2)

"It is a highly customizable miningalgorithm that implicitly represents allfrequent itemsets by extracting only the maximal frequent itemsets. "

- It applies techniques such as Superset frequency based pruning for reducing the number of itemsets considered.
- The result is orders of magnitude in performance improvements over apriori-like algorithms especially when frequent itemsets are long.

- **Pseudocode:**

```
MAX-MINER(Data-set T)
    ;; Returns the set of maximal frequent itemsets present in T
    Set of Candidate Groups C ← { }
    Set of Itemsets F ← {GEN-INITIAL-GROUPS(T, C)}
    while C is non-empty do
        scan T to count the support of all candidate groups in C
        for each g ∈ C such that h(g) ∪ t(g) is frequent do
            F ← F ∪ {h(g) ∪ t(g)}
        Set of Candidate Groups C_new ← { }
        for each g ∈ C such that h(g) ∪ t(g) is infrequent do
            F ← F ∪ {GEN-SUB-NODES(g, C_new)}
        C ← C_new
        remove from F any itemset with a proper superset in F
        remove from C any group g such that h(g) ∪ t(g)
                has a superset in F
    return F
```

I1	I2	I3	I4	I5
T1,T4, T5,T7, T8,T9	T1,T2,T 3,T4,T6 ,T8,T9	T3,T5, T6,T7, T8,T9	T2, T4	T1,T8

Step 1 :

PAIRS	COMMON	COUNT
I1,I2	T1,T4,T8,T9	4
I1,I3	T5,T7,T8,T9	4
I1,I4	T4	1 (Eliminated <2)
I1,I5	T1,T8	2
I2,I3	T3,T6,T8,T9	4
I2,I4	T2,T4	2
I2,I5	T1,T8	2
I3,I4	0	0 (Eliminated <2)
I3,I5	T8	1 (Eliminated <2)
I4,I5	0	0 (Eliminated <2)

Step 2 :

On Grouping 4 items (I1,I2,I3,I5) .Here T8 is only commonelement Thus count is 1 i.e.< 2 . So, it is excludedSo (I1,I2,I3)&(I1, I2, I5)are the mostfrequent item sets.
The result is orders of magnitude inperformance improvements overApriori-like Algorithms, especially whenfrequent item setsare long.

2.3. RANDOM FOREST ALGORITHM

- Random forests is an Ensemble Learning Method for Classification, Regression, thatcreates multiple Decision trees during the trainingand results in the output in the form of Meanof the individual decision tree.
- Random forest uses Bagging Technique. It is Simple and Powerful Ensemble Learning.
- Technique Which reduces Over fittingAndVariance.
- So, it creates multiple decision trees based on the previous purchased history of the customer and then calculatesmen of individual tree andBased on the Final Result obtain it recommends the customer the particular product he/ she should purchase. The importantpart of Random Forest is thatInstead of sticking to particularsimple it selects multiple Randomsamples due to which it provides most accurate result than other MachineLearningalgorithms.
- Result: By comparingRandom Forest with different MachineLearningalgorithms using different Datasetsandparameters Like are Under ROC Curve, Cross-Entropy Loss, Root MenSquare Error Random Forest algorithmwas found to bemost accurate with an accuracy of 93.53%.

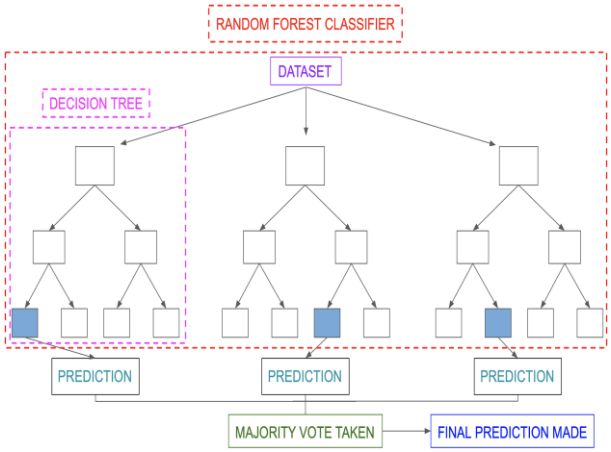

Fig : Working of Random Forest Algorithm

2.4.SUPPORT VECTOR MACHINE

- SVM is Supervised Machine Learning Algorithm mainly used for Classification of problem.
- As in SVM, we shall be using Past 6-months dataset to predict the items which are purchased more or not purchased product.There will be three columns such as Name of grocery, months, purchased data of item.These purchased columns will be in binary format denoting whether the inidividual has bought the product(1) or not(0).
- This Algorithm Classified the data items into n-dimensional space and finding the best decision boundary which is called hyperplane that differentiate the classes.Different Kernal Functions can easily generate a set of decision functions(called as support vectors) which uses subsets of training datasets to give output.So it is memory efficient Algorithm.In SVM ,the line that is used to separate the classes is referred as hyperplane.Tha data points on either side of the hyperplane that are closest to that hyperlane are called support vector which is used to plot the boundary line.
- This Kernal Functions converts low dimensional space i.e. not separable problem to high Dimensional spaces i.e. Separable problem.It separates the data based on the label or outputs that we will defined.
- We can used different kernals functions for the predicting the datasets .For training the model "rbf" kernal will used,rbf means Radial Basis Function.

Based on the confusion matrix,it will show the number of correct and incorrect predictions on purchased or not purchased products.

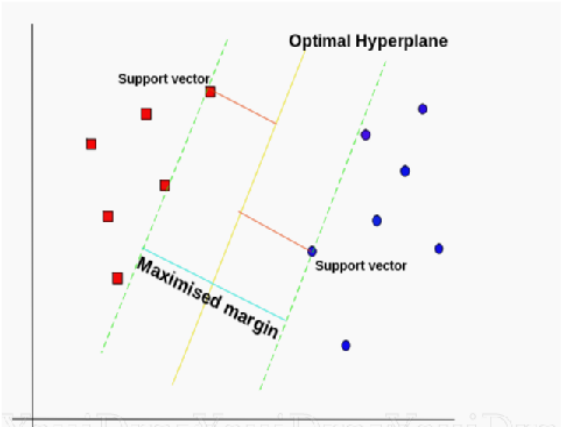

Fig: Working of Support Vector Machine Algorithm

III. CONCLUSION

Analysing the proper requirement of the consumer and suggesting him products accordingly is the main goal of the paper.

So to achieve it different machine learning algorithms have been considered .Apriori and Max-Miner algorithms that analyses the previous data and depicts the frequency of product bought by the consumer .By comparing Apriori and Max-Miner Algorithm , Max-Miner Algorithm showed more accuracy than Apriori Algorithm. Random forest and support vector machine Algorithm can be used to classified the products based on purchased product and not purchased product in the grocery list. Comparing both the algorithm Support vector machine showed more accuracy than random forest algorithm. This algorithm used to build a system where grocery purchasing and predicting and recommending the most visitied grocery in the list.So it will beneficial for consumer to save their valuable time .

ACKNOWLEDGEMENT

The authorwe wish to acknowledge Mr.Atul Pawar ,Professor at Pimpri Chinchwad College of Engineering, Pune for his guidance and encouragement towards this work.

REFERENCES

[1]. Harsha Jayawilal W.a & Saminda Premeratne, "*The Smart Shopping List an effective mobile solution for grocery list-creation process",2020.*

[2]. Neha Kumari Department of CSE VIT UniversityChennaiCampus, India ,"*Grocery Prediction using Machine Learning*" ,ISSN 2321 3361 © 2020 IJESC.

[3]. Purvika Bajaj1, Renesa Ray2, Shivani Shedge3, Shravani Vidhate, Prof. Dr. Nikhilkumar Shardoor,Student, Dept. of Computer Science and Engineering, MIT School of Engineering, MIT ADT University, Pune Maharashtra, India,"S*ales Prediction using Machine Learning Algorithms*",InternationalResearchJournal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 06 ,2020.

[4]. Marcus Liwicki, Sandra Thieme, Gerrit Kahl, and Andreas Dengel, "*An Intelligent Shopping List - Combining Digital , Paper with Product Ontologies",*2020.

[5]. Jolly Rastogi, Assistant Professor, IET, Lucknow, Uttar Pradesh, India,Dr. Supriya Awasthi ,Assistant Professor, KhwajaMoinuddin Chishti Language University, Lucknow, Uttar Pradesh, India, "*A Supply Chain Management Based Customer Prediction Model For Grocery Store",* International Journal of Management (IJM),Volume 11, Issue 12, December 2020.

[6]. *"Decision Tree and Random Forest algorithms Including the Relevance Factor with applications in Bioinformatics",* IEEE International Conference on Bioinformatics and Biomedicine (BIBM),2019.

[7]. Nadia Tahiri§, Bogdan Mazoure, Vladimir Makarenkov," *An intelligent shopping list based on the application of partitioning and machinelearningalgorithms*", PROC. OF THE 18th PYTHON IN SCIENCE CONF. SCIPY 2019.

[8]. Noor Syafiqah ayuni Mohd-Said, Norliza Katuk , "*A Mobile app for Collaborative Grocery List Management Online Grocery ordering System*", SCRIBD 2019.

[9]. Howard, Dale S. (DesMoines, Ia) Messer, Jennifer R. (Norwalk, Ia) Delamore, Gregory J. (DesMoines, Ia) "*System and method for household grocery management ",*2019.

[10]. Harsh Valecha, A. Varma, Ishita Khare, A. Sachdeva, Mukta Goyal ,"*Prediction of Consumer Behaviour using Random Forest algorithm*", 2018.

[11]. B Pujitha ,"*Grocery Management System",2018.*.

[12]. Batta Mahesh ,"*Machine Learning algorithms - a Review*",International Journal of Science and Research (IJSR) ISSN: 2319-7064,2018.

[13]. Raewyn Bassett and Brenda Beagan ,"*Grocery lists: connecting family,household*

*and grocery store",*2018.

[14]. *Algorithms* studied from https://www.geeksforgeeks.org/, https://www.tutorialspoint.com/index.htm, https://www.wikipedia.org/ , etc.

[15]. Rohit Rathish Department of Computer Engineering, V.E.S. Institute of Technology, University of Mumbai-400074, India. ,"*Predicting online Grocery Ordering Intention",* InternationalJournal of Engineering Research & Technology (IJERT) ,Vol. 6 ,2017.

[16]. Yi Zuo∗ Institutes of Innovation for Future Society Nagoya University Furo-cho, Chikusa-ku, Nagoya, Japan ,"*Prediction of Consumer Purchasing in a Grocery Store Using Machine Learning Techniques"*3rd asia-Pacific World Congress on Computer Science and Engineering,2016.

[17]. Chad Cumby, Andrew Fano, Rayid Ghani, Marko Krema accenture Technology Labs 161 N Clark St Chicago, USA ,"*Predicting Customer Shopping Lists from Point of SalePurchaseData",*.fchad.m.cumby,andrew.e.fano,rayid.ghani,marko.kremag@accenture.com.,2016.

[18]. Raorane a.a.1 Kulkarni R.V.2 and Jitkar B.D.3 Department of computer science, Vivekanand College, Tarabai park Kolhapur, MH, INDIa ,2.Chh. Shahu Institute of business Education and Research Centre Kolhapur, MH, INDIA ,3.Department of the Computer Science, D.Y. Patil College of engineering, Kolhapur, MH, INDIA ,"*Association Rule – Extracting Knowledge Using Market Basket analysis"* , Research Journal of Recent Sciences , Vol. 1(2), 19-27, Res.J.Recent Sci.,2012.

[19]. Hetal Bhavsar, Amit Ganatra ,"*A Comparative Study of Training algorithms for Supervised Machine Learning*", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4,2012.