IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 3, November 2022

Predicting the Frequent Item Sets for Supermarket Data

Mr. D. Srinivasa Rao[1], B. Sai Mahitha[2], G. Apoorva[3], G. Yamuna[4], M. Pallavi[5]
Assistant Professor, Department of Information Technology[1]
IV B.Tech Students, Department of Information Technology[2,3,4,5]
Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India

Abstract: *This Python notebook uses the Apriori algorithm to analyze datasets from various supermarkets, retail organizations, and minimarkets, resulting in a more accurate analysis of customer behaviour and better product prediction and forecasting. The dataset that is used in this model typically involves customer purchases in supermarkets or any other organization. The datasets contain item details as well as the number of transactions purchased by customers. This model can be used by retailers and supermarkets of all sizes in both urban and rural areas. This algorithm implementation enables accurate forecasting and allows products to be sold efficiently and profitably in stores. Supermarkets, for example, can use the resulting data to forecast future sales volume using a variety of machine-learning techniques. It displays the most frequently purchased items or associated items by the user. This prediction is primarily focused on figuring out the rules of the association. It identifies the set of items or attributes that occur together or frequently in the dataset using association rules. If this apriori model meets a minimum threshold value for support and confidence, it produces a set of items known as a frequent itemset. This Python notebook implements a prediction model based on the apriori algorithm, which improves the efficiency of level-wise generation of frequent item sets by utilizing an important property known as the Apriori property, which aids in reducing the search space.*

Keywords: Dataset, Apriori, Machine learning, prediction, forecasting, confidence, support

I. INTRODUCTION

Due to today's transition from visiting physical stores to online shopping, predicting customer behaviour and forecasting future sales is an essential component of any organization. The organization's information system was unable to provide the customer with shopping habits. As the system cannot provide information on the relationship between items purchased by customers, there is frequently a vacancy in one stock of goods that are frequently purchased concurrently. Many businesses analysts state that service problems are common in retail businesses, specifically minimarket owners or retail stores that still place shelf positions without regard to the relationship between goods, making it difficult for consumers to find the goods. Accurate forecasting of future sales assists businesses in developing and improving business strategies, as well as gaining proper market knowledge. It has the potential to boost customer satisfaction and sales, resulting in higher conversion rates and a more personalized shopping experience. Models for predicting customer behaviour can be built using machine learning and artificial intelligence algorithms and supplementary customer data for supermarket cart prediction.

II. LITERATURE REVIEW

The conventional Apriori method and its use in supermarkets are reviewed by Pragya Agarwal et al. in their study [4]. One of the core algorithms that can be used to mine the common things is apriori. Algorithms' current drawbacks include the necessity for data repetitions, the difficulty of identifying uncommon data sets, etc. By utilising Association Rule Mining (ARM), which more accurately anticipates customer-related sales trends, the proposed initiative will address these shortcomings.

The Association rule mining utilizing modified Apriori method by Minal G. Ingle et al. [5] is highlighted. By minimizing the number of scans while parsing the database, which was a weakness in conventional databases, the

Apriori algorithm is improved in this case. Additionally, frequent item sets are generated more precisely. Their suggested technique is intended to be further improved using multilevel association rules.

Shopping cart analysis was performed using uninorms by Raymond Moodley et al. [7].Uninorms are frequently employed in multi-criteria decision-making as aggregating variables. In this case, the authors substituted the uninormsmonotonicity principle for the conventional Apriori approach. The ideal item set combinations for marketing the products were found using uninorms.

III. PROPOSED SYSTEM

The proposed system creates a predictive model using a Machine Learning algorithm to assist retailers in forecasting product sales. Associative learning is the foundation of this model which contains the Apriori algorithm in Machine Learning, the Apriori algorithm includes three factors: support, lift, and confidence.Apriori is an algorithm related to mining Association Rules. It searches the datasets for a series of frequent sets of items. The Apriori Algorithm is a machine-learning technique used to delve deeper into structured interactions between the various items involved. It is centered on the itemsets associations and correlations. The algorithm's most predominant field of application is to give recommendations based on the products already in the customer's cart. This machine learning algorithm with its ability to collect and analyze customer data in real-time is assisting in gaining a better understanding of customer behavior and needs, and ultimately in developing a personalized customer experience strategy.

IV. TECHNOLOGIES USED

4.1 Python Colab-Notebook

Colab, in particular, is a cloud-based Jupyter notebook environment that is completely free. Most importantly, no configuration or setup is required, and the notebooks we create can be edited concurrently by all team members, just like editing documents in Google Docs. Colab supports a large number of popular machine-learning libraries, which can be easily loaded into the notebook. Write and execute code in Python. Document your code that supports mathematical equations moreover Create/Upload/Share notebooks and Importing/Save notebooks from/to Google Drive. Import external datasets example from Kaggle. All the above-mentioned functionalities can be achieved using google collabs. The developed notebook uses python and machine learning algorithms to perform analysis and prediction.

4.2 Python Libraries

The Python collection module is defined as a container that is used to store collections of data,The collection Module in Python provides different types of containers. The collection module is used in this model. It has different characteristics based on the declaration and the usage. A Container is an object that is used to store different objects and provide a way to access the contained objects and iterate over them.Examples include the Tuple, List, Dictionary, and other built-in containers. It was included to enhance the functionality of the built-in collection containers.The developed model used the counter() function.

counter()

Python Counter is a container that keeps track of the count of each element in the container. The counter class is a subclass of the dictionary class.The Counter, like hashtable objects, stores the data in an unordered collection. As values, the elements in this resemble the keys and the count.This is an easy way to count the items in an iterable list. A Counter can easily perform arithmetic operations such as addition, subtraction, intersection, and union. Function counter() is used to perform the above-stated operations.

frozenset()

The function frozenset() is used during the development of the model. A frozenset is basically the set object's unalterable or immutable version. Because a frozenset is an immutable set, it is appropriate for use cases where there is a need to create a group of keys or identifiers that, don't want the users to customize.

issubset()

When all elements of a set existent in another set (passed as an argument), the issubset() method returns a boolean True; anything other than that returns a boolean False. This function can be found in the Standard Library.

append()

In this model, the append() function accepts the input parameter as a single item and appends it to the end of the specified list. In Python, append() does not return a new list of items; rather, it returns null. It simply alters the original list by appending the item at the end. Append functionality as available at Python Standard Library.

sorted()

To sort the items this model uses Python's sorted() method. It may sort elements in descending order as well as ascending order by default. It requires four arguments and returns a sorted collection. When it comes to dictionaries, it sorts only the keys—not the values.

4.3 Dataset

The data set was downloaded from an Excel or CSV sheet on the Kaggle website. A random supermarket is included in the dataset. It consists of the grocery purchases made by customers. It includes various customers' purchase.

	A	B	C	D
2	DIAPERS	BREAD	MILK	MAGGI
3	FACEWASH	SUNSCREEN	PERFUME	SOAPS
4	TUBELIGHT	BULBS	WIRE	PLASTER
5	SHAMPOO	SOAPS	WATCH	NAILPOLISH
6	SHOES	SOAKS	MOISTURISER	PERFUME
7	BREAD	JAM	VEGETABLES	EGGS
8	SAUCE	OILS	VEGETABLES	EGGS
9	TUBELIGHT	BULBS	ICECREAM	CHOCOLATES
0	PASTA	SAUCE	OILS	ICECREAM
1	BREAD	BUTTER	CHEESE	OILS
2	FACEWASH	SUNSCREEN	PERFUME	SOAPS
3	DIAPERS	WETWIPES	ALCOHOL	BREAD
4	PERFUME	SHAMPOO	NAILPOLISH	HAIRBAND
5	JAGGERY	SAUCE	OILS	MILLETS
6	ALCOHOL	DIAPHERS	LIPSTICK	PERFUME
7	NAILPOLISH	HAIRBAND	VEGETABLES	SAUCE
8	CHEESE	SPICES	TUBELIGHT	BULBS
9	SPICES	GHEE	OILS	RICE
0	ALCOHOL	PLATES	GLASSES	DIAPER
1	FACEWASH	SUNSCREEN	NAILPOLISH	HAIRBAND
2	PANNER	GRATER	ALCOHOL	DIAPER
3	BOOKS	SHOES	PENS	PENCILS
4	VEGETABLES	GRATER	RICE	GHEE
5	SAUCE	OILS	FRUITS	EGGS
6	BREAD	JAM	MILK	BISCUITS
7	JAM	SAUCE	SPICES	GHEE
8	MILKSHAKES	CHIPS	COFFEE	OILS
9	TOWELS	SOAPS	TOOTHPASTE	FACEWASH
0	CHIPS	COKE	FACEWASH	SUNSCREEN
1	TUBELIGHT	BULBS	FAN	WIRE
2	VEGETABLES	FRUITS	MILK	EGGS
3	MILLETS	SAUCE	OILS	CORNFLAKES

	A	B	C	D
69	ALCOHOL	DIAPERS	FISH	MEAT
70	BOOKS	SAUCE	OILS	MILK
71	SHOES	CHOCALATES	WATCH	HAIRBAND
72	NOTEBOOK	PENS	NAILPOLISH	TOWEL
73	PERFUME	CHIPS	MILKSHAKES	OILS
74	SHOES	BOOKS	PENCILS	JUICER
75	SAUCE	OILS	TRAY	SPEAKERS
76	BAT	FOOTBALL	VALLYBALL	CRICKET BALL
77	BREAD	JAM	NOTEBOOK	PENS
78	BOOKS	SHOES	SOCKS	SPEAKERS
79	FACEWASH	SUNSCREEN	LOTIONS	LIPSTICK
80	SPICES	CHILIPOWDER	PEANUTS	SPEAKERS
81	VEGETABLES	FRUITS	SAUCE	SOAPS
82	CORNFLAKES	SUGAR	JAGGERY	MILLETS
83	NOTEBOOK	PENS	SHOES	CHOCOLATES
84	ICECREAMS	TISSUES	FRUITS	VEGETABLES
85	JEWELLERY	MAKEUP	TRAY	VEGETABLES
86	FRUIT	SAUCE	OILS	FISH
87	BREAD	CHOCOLATES	MIKSHAKES	CHIPS
88	NOTEBOOK	PENS	PENCILS	SHOE
89	ALCOHOL	DIAPERS	BALL	BAT
90	NAILPOLISH	HAIRBAND	SHOES	SOCKS
91	JEWELLERY	MAKEUP	VEGETABLES	FRUITS
92	JEWELLERY	MAKEUP	NAILPOLISH	FRUITS
93	NOTEBOOK	PENS	SHOES	FRUITS
94	CHILI POWDER	SAUCE	FISH	NUTS
95	MAGGI	PASTA	ALCOHOL	DIAPHERS
96	SAUCE	OILS	BREAD	BUTTER
97	NAILPOLISH	HAIRBAND	FACEWASH	SOAPS
98	NOTEBOOK	PENS	PENCILS	PAPERS
99	BREAD	MEAT	VEGETABLES	SAUCE
100	SAUCE	OILS	NAILPOLISH	HAIRBAND

V. EXPERIMENTAL RESULT

Here, nine transactions were used in this model. The dataset yields the results as stated above after the functions have been applied. It provides all of the dataset's singular items. Repeated products won't be shown. The items were also arranged alphabetically. Lists representing the outcomes are shown.

```
data = [
 ['T100',['Alcohol','Diapers','Chocolates']],
 ['T200',['Diapers','Soaps']],
 ['T300',['Diapers','Chocolates']],
 ['T400',['Alcohol','Diapers','Soaps']],
 ['T500',['Alcohol','Chocolates']],
 ['T600',['Diapers','Chocolates']],
```

['T700',['Alcohol','Chocolates']],
['T800',['Alcohol','Diapers','Chocolates','Vegetables']],
['T900',['Alcohol','Diapers', 'Chocolates']]
]

Users determine the minimal support in the Apriori algorithm for generating association rules. The support value entered by the user is returned by this section of the code. It is determined using the dataset's length and the provided support value. The support value, in this case, is 1.

```
[ ] #choosing a value for the support. For this example we will choose support to be 20%.
    sp = 0.2
    s = int(sp*len(init))
    s

    1
```

Look up each item's count in the dataset. This number is known as the support Count. This indicates that it calculates the count of each item in transactions. This set is represented as C1 and is also called a one-frequent set.

```
C1:
['Alcohol']: 6
['Chocolates']: 7
['Diapers']: 7
['Soaps']: 2
['Vegetables']: 1
```

The itemsets in C1 will be compared with minimum support value. The items which satisfy the given constraint will be formed as L2.

```
L1:
['Alcohol']: 6
['Chocolates']: 7
['Diapers']: 7
['Soaps']: 2
['Vegetables']: 1
```

Two frequent item sets should be produced in the case of C2. Similar to how L2 will emerge from C2, so too will L3. similar C4,L4. They will be contrasted with support values to derive these. By using this apriori property we can reduce the size of the subsets.

```
C2:
['Vegetables', 'Alcohol']: 1
['Diapers', 'Chocolates']: 5
['Chocolates', 'Alcohol']: 5
['Vegetables', 'Diapers']: 1
['Alcohol', 'Soaps']: 1
['Diapers', 'Soaps']: 2
['Vegetables', 'Soaps']: 0
['Vegetables', 'Chocolates']: 1
['Diapers', 'Alcohol']: 4
['Chocolates', 'Soaps']: 0

L2:
['Vegetables', 'Alcohol']: 1
['Diapers', 'Chocolates']: 5
['Chocolates', 'Alcohol']: 5
['Vegetables', 'Diapers']: 1
['Alcohol', 'Soaps']: 1
['Diapers', 'Soaps']: 2
['Vegetables', 'Chocolates']: 1
['Diapers', 'Alcohol']: 4
```

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 3, November 2022

Impact Factor: 6.252

```
C3:
['Vegetables', 'Diapers', 'Alcohol']: 1
['Diapers', 'Chocolates', 'Alcohol']: 3
['Chocolates', 'Alcohol', 'Soaps']: 0
['Vegetables', 'Diapers', 'Soaps']: 0
['Vegetables', 'Alcohol', 'Soaps']: 0
['Diapers', 'Chocolates', 'Soaps']: 0
['Vegetables', 'Chocolates', 'Alcohol']: 1
['Vegetables', 'Diapers', 'Chocolates']: 1
['Diapers', 'Alcohol', 'Soaps']: 1

L3:
['Vegetables', 'Diapers', 'Alcohol']: 1
['Diapers', 'Chocolates', 'Alcohol']: 3
['Vegetables', 'Chocolates', 'Alcohol']: 1
['Vegetables', 'Diapers', 'Chocolates']: 1
['Diapers', 'Alcohol', 'Soaps']: 1

C4:
['Chocolates', 'Alcohol', 'Diapers', 'Soaps']: 0
['Chocolates', 'Alcohol', 'Vegetables', 'Diapers']: 1
['Alcohol', 'Vegetables', 'Diapers', 'Soaps']: 0
```

```
[10]
L3:
['Vegetables', 'Diapers', 'Alcohol']: 1
['Diapers', 'Chocolates', 'Alcohol']: 3
['Vegetables', 'Chocolates', 'Alcohol']: 1
['Vegetables', 'Diapers', 'Chocolates']: 1
['Diapers', 'Alcohol', 'Soaps']: 1

C4:
['Chocolates', 'Alcohol', 'Diapers', 'Soaps']: 0
['Chocolates', 'Alcohol', 'Vegetables', 'Diapers']: 1
['Alcohol', 'Vegetables', 'Diapers', 'Soaps']: 0

L4:
['Chocolates', 'Alcohol', 'Vegetables', 'Diapers']: 1

C5:

L5:

Result:
L4:
['Chocolates', 'Alcohol', 'Vegetables', 'Diapers']: 1
```

The association rules with the greatest degree of confidence will be picked. The items with the highest value will be taken into consideration in this model after the first confidence for each rule has been determined. The retailer will take into account and pay attention to the aforementioned rules 1,4,5, and 7 .These rules will help in business strategy.

```
['Vegetables', 'Chocolates', 'Alcohol'] -> ['Diapers'] = 100.0%
['Diapers'] -> ['Vegetables', 'Chocolates', 'Alcohol'] = 14.285714285714285%
['Diapers', 'Chocolates', 'Alcohol'] -> ['Vegetables'] = 33.333333333333333%
['Vegetables'] -> ['Diapers', 'Chocolates', 'Alcohol'] = 100.0%
['Vegetables', 'Diapers', 'Chocolates'] -> ['Alcohol'] = 100.0%
['Alcohol'] -> ['Vegetables', 'Diapers', 'Chocolates'] = 16.666666666666664%
['Vegetables', 'Diapers', 'Alcohol'] -> ['Chocolates'] = 100.0%
['Chocolates'] -> ['Vegetables', 'Diapers', 'Alcohol'] = 14.285714285714285%
choosing: 1 4 5 7
```

VI. FUTURE SCOPE

Direct development of this prediction model is possible in supermarkets. Additionally, if new data analysis tools or programs were to be produced, the data might be analyzed in a better format. Due to this model's extremely accurate sales forecasting, it may be heavily utilized. Additionally, there won't be any possibility of them misplacing or losing the items they wanted to buy. If further development is required, a front-end GUI can be created, with the code serving as the back-end, and the user's provided data flowing from front to back through the database.

VII. CONCLUSION

The frequent items, associated items, or most combinational items are the outcomes of this prediction model. Additionally, using association rules, the data can be examined in a more efficient manner. The association rule, which simply expresses correlation, is crucial to this endeavor and asserts that the presence of one item or collection of things suggests the presence of another item with some probability and predicts a target. With the use of this model, merchandise may be efficiently and economically sold in stores while also assisting retailers with forecasting.

REFERENCES

[1]. Vidhya. G, Marimuthu. M, Vinothkumar. R. B, Vidyabharathi. D, Theetchenya. S, Basker.N, Mohanraj.G- “An enhancement of Apriori Algorithm for Shopping Cart Analysis of Customers” in International Journal of Advanced Science and Technology Vol. 29, No. 08, (2020), pp. 2124- 2131.

[2]. Loraine Charlet Annie M.C.1, Ashok Kumar D2-“Market Basket Analysis for a Supermarket based on Frequent Itemset Mining” in IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012.

[3]. Putri Agung Permatasari1, Linawati2, Lie Jasa3-“Analysis of Shopping Cart in Retail Companies Using Apriori Algorithm Method and Model Profset” in International Journal of Engineering and Emerging Technology, Vol.5, No.2, July — December 2022.

[4]. Agarwal, Pragya1, Madan Lal Yadav2,Nupur Anand3-"Study on Apriori Algorithm and its Application in Grocery Store." in International Journal of Computer Applications 74.14.2017.

[5]. Minal I, N. Suryavanshi, “Association Rule Mining Using Improved Apriori Algorithm” in International Journal Of Computer Applications, Volume 112,Issue 4, February 2015.

[6]. Ila Padhi, Jibitesh Mishra,Sanjit Kumar Dash-“Predicting Missing Items in Shopping Cart using Associative Classification Mining” in International Journal of Computer Applications (0975 – 8887) Volume 50 – No.14, July 2012.

[7]. MoodleyR, Chiclana.F, Caraffini. F, Carter.J- “Application of uninorms to shopping cart analysis” in International Journal of Intelligent Systems, 34(1), 39-49.2019

[8]. Kasun Wickramaratna,Miroslav Kubat,Kamal Premaratne-“ Predicting Missing Items in Shopping Carts” in IEEE Transactions on knowledge and Data Engineering, vol. 21, no. 7, july 2009.

[9]. Kasun Wickramaratna Miroslav Kubat Kamal Premaratne –“ Predicting Missing Items in Shopping Carts” in IEEE Transactions on Knowledge and Data Engineering (Volume: 21, Issue: 7, July 2009).

[10]. M. Nirmala and V. Palanisamy –“ An Efficient Prediction of Missing Itemset in Shopping Cart” in Journal of Computer Science Volume 9 No. 1, 2013, 55-62.

[11]. Putri Agung Permatasari, Linawati, Lie Jasa–“ Analysis of Shopping Cart in Retail Companies Using Apriori Algorithm Method and Model Profset” in International Journal of Engineering and Emerging Technology, Vol.5, No.2, July — December 2020.