European Journal of Computer Science and Information Technology
Vol.7, No.5, pp.1-15, October 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

AN INTELLIGENT PRODUCT SUGGESTION ALGORITHM USING PREDICTIVE ANALYSIS FOR PERSONALIZED USER INTERFACE BUILDING

R. H. R. Perera, K. L. Jayaratne
University of Colombo, School of Computing

ABSTRACT: *The main objective of this research was to propose a technological solution to the long queues that are often seen in many retail outlets. As the solution this research proposes a self-checkout application. The application populates a list predicted next purchasing item set making the user interface intelligent and user friendly. The research introduces a model named RFR-U model to generate the next purchasing item list of the customer. It uses the parameters; relevance, recency and frequency to determine the next purchasing item set. The algorithm uses a rule based approach with weighted ratings. Although collaborative method is a popular method in finding such results, in the studied scenario, it is not applicable as the store does not maintain a comprehensive user profiles or facilitates the users to rate products. The proposed algorithm and the solution was evaluated both quantitatively and qualitatively and results show an accuracy above 80%.*

KEYWORDS: Recommendation systems, purchasing patterns, Recency, Frequency, Relevance, self-checkout

INTRODUCTION

Today, where everything is expected to be instant, long queues that are often seen in many retail outlets may lead to dissatisfied customers. The main objective of this research is to propose a technological solution to the long queue problem, as it may negatively impact the businesses. As the case study for the research, a food outlet in a renowned super market chain was selected.

During the problem analysis, many causes for long queues were identified. High arrival rate, less experienced cashiers, number of items in customers tray, language barriers are to name a few. It was observed that most of the customers who visit the outlet are regular customers and there are repetitive patterns in their purchases. Furthermore, the selected store maintains a large set of historical data of the customers and their transactions. The research problem is formally defined as;
“The traditional method of item checkout has failed to serve the rising demands, leading to long queues in outlets and dissatisfied customers.”

Therefore as the technological solutions for the stated problem, this research proposes a self-checkout application that can be used by the customers to check out the items bought. Since customers can be of many types, the application should be simple and user friendly. The application can be made user-friendly and personalized by filtering out the items that the user is likely to purchase.

To predict the next purchasing item set, this research introduces a simple rule based weighted rating model, named as RFR-U model. RFR-U stands for Recency, Frequency, Relevance and User. This uses historical data of the user to predict the next purchasing item set.

LITERATURE REVIEW

Predictive analysis is a branch of the advanced analytics which is used to make predictions about the future. Many areas have been benefited from predictive analysis, such as weather forecasting, healthcare, supply chain and automobile.[10][11][12][13] This research focuses on the retail sector and predictive analysis has been widely used in the retail sector to solve many problems. One such scenario is identifying consumer buying patterns. Many researches are done to predict and identify consumer buying patterns. The results of these are widely used in decision making, planning, sales and marketing etc. But the main focus of this research is to use the identified buying patterns to recommend or suggest products for a self-checkout application.

Numerous product recommending algorithms are currently available to predict customers next purchasing item set. Product recommending algorithms are mainly classified into, content based methods, collaborative filtering and hybrid approaches [7][6]. Collaborative filtering methods compare users and make suggestions depending on the user similarities. This needs a large set of data and most of the algorithms are time consuming. Content-based methods uses historical data of the user. There are many hybrid approaches already available such as, RFM model, rule based models and matrix models [4][5][6][9].

RFM model has been proposed by Hughes in 1994 [2]. It proposes three behavioural variables, namely Recency (R) Frequency (F) and Monetary (M). Recency denotes the latest purchases, frequency denotes the total number of purchases and monetary denotes the monetary value spent during one period [3]. There are drawbacks of the model as well and therefore revised models has been introduced. One such model is the revised model proposed by Macus. He used the number of purchases (F) and the average purchase amount (A) to construct two-dimensional matrix model based on CLV (customer lifetime value) to correct the RFM method [4].

The case study for the reported release being bakery items purchases, it is highly likely that the results depends on the recent purchases than purchases made months or years back. Therefore the two behavioural variables, recency and frequency in the RFM model can be considered as two important variables in this research as well.

Customer segmentation methods can also be used for product recommendation and those methods can be classified into two categories, i.e. market research or data mining methodologies. This research has used neural networks and fuzzy logic to come up with a solution. [3]

The paper titled, “An Intelligent Product Recommendation Model to Reflect the Recent Purchasing Patterns of Customers” [1] uses the recent purchasing patterns to recommend product to customers. It has considered both customer information and product information for the solution. The research has used various data mining classifiers such as the decision tree, neural network, support vector machine, random forest, rotation forest, sliding-window scheme for the recommendation model.

Another research uses a combination of attribute based approach and a sequential based approach. It makes a personal preference matrix for each user to generate results. In the sequential based approach it uses weighted association rules to identify latest patterns. It suggests that the purchasing processes usually have some time-dependency relationship and is repeatable and also periodical. [8] This is similar to what was observed during initial problem analysis of this research. Rule based approach is also a simple approach of recommendation. Rules are usually derived from database of previous transactions. This uses data mining techniques such as web usage mining, decision tree induction to find association rules. [7]

However, for the research stated, an efficient method had to be used, since the results should be generated real time. Although collaborative filtering is the most commonly used approach this method is not suitable for this research as the store does not maintain a comprehensive user profile and the store does not have a rating mechanism on the products. Therefore, for this research rule based, weighted rating system was used.

PROPOSED SOLUTION

As digitalization is on the rise in the global market and all the organizations are turning toward digitalization, it is reasonable to look at the trending digitalization concepts to derive a solution. Data analysis, artificial intelligence based solutions, machine learning, and mobility are some of those popular topics in the digitalization stack. Further, many organizations encourage users to use self-operated applications.

Therefore this research proposes a self-checkout application as the solution for the long queue problem. To create a more personalized application, the self-checkout application is designed to suggest the next purchasing item set. This will make it easier for the user to select the items the user has purchased from the long list of items available.

The self-checkout application could address the following problems identified;

- Delay in entering the items to the system: Cashier need not to enter each individual item to the system. By scanning the code, the list of items will be automatically populated.
- Less experienced cashiers: It is not necessary for the cashiers to remember the codes as the items will auto populate..
- Language barriers and misunderstandings: Since the customer has the facility to add the item to the bill, the necessity of communication decreases

The solution consists of three main parts. The mobile application, the algorithm and the dashboard. Below image illustrates the high-level system design.

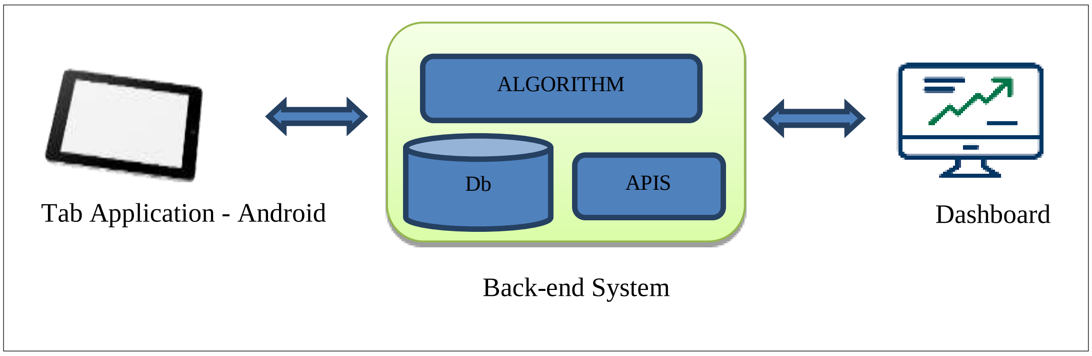

Figure 1. High level system diagram.

Mobile Application

The mobile application is an Android application designed for tablets. The application can be used by customers to check-out their items. It was developed as a proof of concept. The application is named as "I-Checkout" to resemble its self-checkout functionality.

The user could enter his/her loyalty id and login to the system. Upon login, the user is directed to the home screen where a list of predicted items are displayed. The application uses the RFR-U algorithm to generate the said list. The result set of the algorithm is obtained using the services hosted in the back-end.

The user could tap on the icons to add preferred items to the food basket and using the checkout button a unique code can be generated. This code can be used for further processing such as payments.

Below given are few user interfaces of the mobile application.

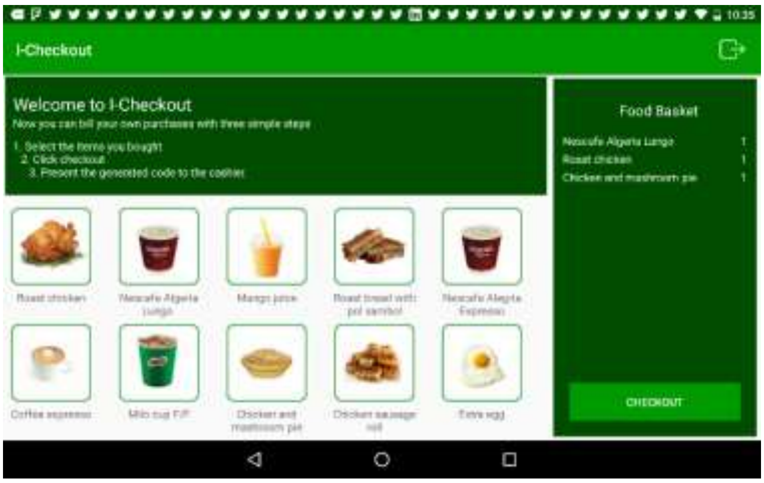

Figure 2. Home Screen.

Vol.7, No.5, pp.1-15, October 2019

Published by ECRTD- UK

Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

Figure 3. Checkout Screen.

DASHBOARD

The dashboard displays the results of the algorithm. It was implemented for evaluation and presentation purposes. It contains three input fields; user id, date and time. A search button is also present in the dashboard. Once the user id, date and time is entered, the search button can be used to generate results for the entered date and time of the entered user id. The dashboard displays results of several variations of the RFR-U model as well.

The dashboard could be further improved to be used by the organization to make decision on the meal plans, procurements and item quantities.

DATA ANALYSIS

The data was collected from a set of sample users over a period of six months. Further their purchasing behaviours were closely monitored. During problem analysis, the following variables were identified as potential deciding factors of a purchase.

- Time of purchase
- Day of purchase

- Ongoing promotions
- Customer preference of taste
- Weather

Out of these, the first three factors were analysed to check if the relationship with the purchase is significant. The factors analysed were time, date and frequency.

Below figures illustrate how the dataset of user 1 changes when it is filtered by different variables.

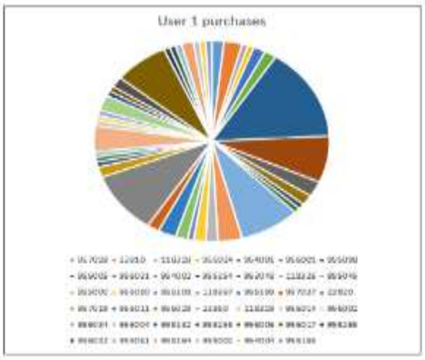

Figure 4. Items purchased by user 1 during a given period of time

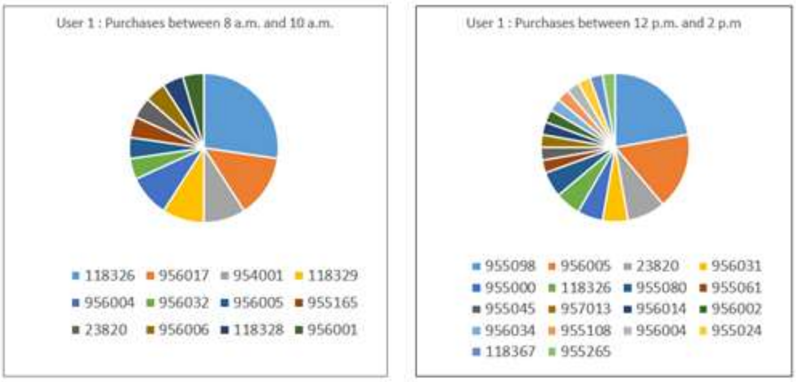

Figure 5. Items purchased by user 1 when filtered by time.

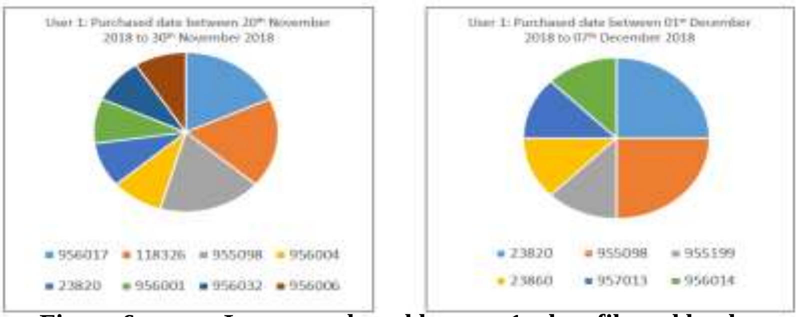

Figure 6. Items purchased by user 1 when filtered by date.

European Journal of Computer Science and Information Technology

Vol.7, No.5, pp.1-15, October 2019

Published by ECRTD- UK

Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

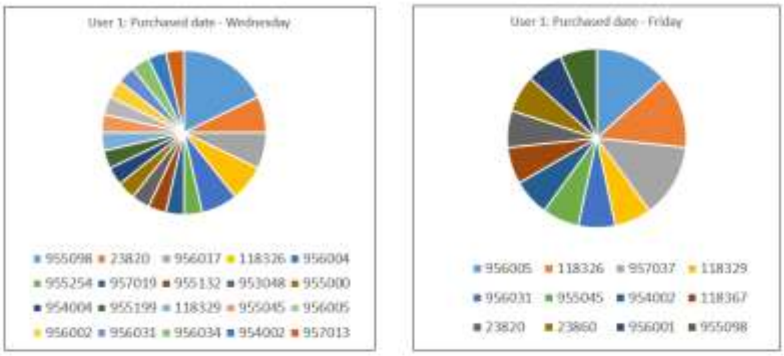

Figure 7. Items purchased by user 1 when filtered by day.

Comparing the graphs, it could be observed that when the data set is filtered by the mentioned variables, the variation of the data set is drastically reduced. Hence it is clear that, date of purchase, time of purchase, frequency of purchase and day of purchase significantly impact the purchasing patterns of the users. Therefore, the hypothesis for the research can be defined as,

"Factors such as Time, date affects the users purchase and these variables could be used to develop an algorithm to predict the next purchasing item list of the user and it could be used for intelligent user interface building for a self-checkout application"

RFR-U MODEL

As described in the data analysis section variables such as date of purchase, time of purchase and day of purchase significantly affect the buying patterns of users. Therefore these variables can be used to derive an algorithm to predict the next purchasing item set. In addition, the algorithm should generate results within seconds as it will be used in a real time application. Therefore, to predict the results only the recent purchases have been considered. It is also important that the algorithm capture the patterns with a limited amount of data.

As the first approach a matrix method was used. However, the large number of different items made it more complex to determine the next purchasing item set.

The second approach used is a combination of rule based and content-based models. Different rules related to identified variables; time, date, day and frequency was experimented to determine a suitable filtering approach. As found during the literature survey, some of the researchers have used rating methods to prioritize items. Therefore different rating methods were also examined during the experiments. After evaluating different approaches the RFR-U model was developed to generate the next purchasing item set. RFR-U stands for Recency, Frequency, Relevance and User. The design of the algorithm is as follows.

Vol.7, No.5, pp.1-15, October 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

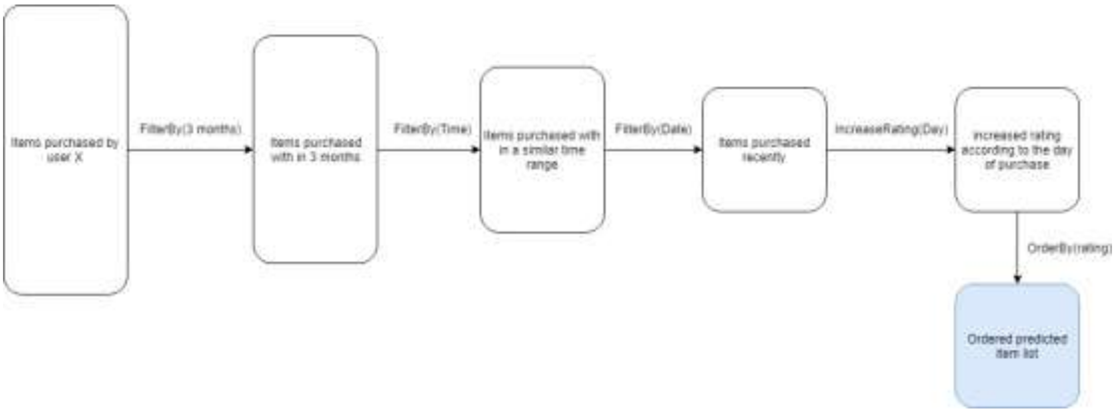

Figure 8. High level design of the algorithm.

The algorithm uses historical data of three months from the date of purchase to generate the results. A simple rule based method is used to produce results. Thus makes the algorithm simple and fast.

The algorithm will calculate a rating for each items according to the time, date, frequency and the day factors. The comparisons are made between the current purchase date and time and the historical item purchased date and time.

The algorithm first filters the records using the relevance variable (time). A rating based on the frequency and the weight is assigned to each filtered item. Thereafter, the rating of the items are further increased based on the recency and the relevance (day) variable. The result set is then sorted in a descending order. Based on the rating, the probability of purchase for each item is calculated. Thereafter the result is further filtered by eliminating items which has the probability percentage less than 0.5.

Relevance

It was identified during the research that the users purchase different products during different times of the day and day of the week. Therefore this parameter checks how relevant is an item to the purchasing date and time.

Time variable checks how relevant is an item in the list, to the time of the day. The weight table for the difference is given below. The algorithm uses addition to increase the rating of the item.

Table 1. Weight table for time variable.

Time difference	**weight**
Purchasing time is between $\pm 30\ mins$ of item purchased time	7
Purchasing time is between $\pm 1.25\ hr$ of item purchased time	5
Purchasing time is between $\pm 2\ hr$ of item purchased time	3

Recency

This refers to recently purchased items. It is likely that the user purchase the same item he/she bought recently. The reasons for this behaviour could be an ongoing promotion, weather condition, current health situation etc.

This parameter compares the item purchase dates in-terms of the week of purchase. Since this is parameter is the most important parameter the algorithm uses multiplication to increase the rating.

The weight table for the recency variable is given below.

Table 2. Weight table for date variable.

Date difference	weight
Item purchased date is within 1 week of the purchasing date	10
Item purchased date is within 2 week of the purchasing date	8
Item purchased date is within 3 week of the purchasing date	5

Frequency

This refers to the number of times the user has purchased the item within the given time frame. This is used as initial rating and the base for weight calculations.

User

User parameter indicates that the algorithm uses past data of the particular user only. The algorithm uses historical data of three months.

EVALUATION

The evaluation of the algorithm and the mobile application was done in both quantitative and qualitative approaches. For quantitative approach different types of validation equations have been used and for quantitate approach, feedback obtained from different users were used.

European Journal of Computer Science and Information Technology
Vol.7, No.5, pp.1-15, October 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

Fresh bakery item purchasing bills were collected over a period of six months. As the algorithm uses historical data of three months, to predict the next purchasing item list, the bills collected during the last few months were used for evaluation.

QUANTITATIVE EVALUATION

A total of 30 bills were selected from the available bill collection which covers different user and date combinations. The user id, the purchased date and time in the bill was entered using the dashboard created and then the predicted item set was obtained using the algorithm. Below is the test case used for the quantitative evaluation.

1. Pick a bill from the bill set
2. Open the dashboard
3. Enter the user id, date and time in the dashboard
4. Click search button to generate results
5. Compare the results against the list in the bill
6. Enter the findings in the evaluation excel

The generated (Predicted) item set was compared against the actual item set in the bill. If the items predicted by the algorithm matches the items in the bill, it can be concluded that the algorithm has 100% accuracy. If few items are present it can be concluded that the algorithm is partially accurate. To calculate the accuracy percentage of the individual bills and the total accuracy, the following set of equations were used.

Let,

$$P(n) = \begin{array}{c} Correctly\ predicted\ number \\ of\ items\ (within\ first\ n\ records) \end{array}$$

$$N = Number\ of\ records\ in\ the\ bill$$

1. Rate of accuracy

The accuracy of the predicted items of a single bill

$$Rate\ of\ Accuracy(n) \quad = \quad RA(n) = \frac{P(n)}{N} \times 100\%$$

2. Strict rate of accuracy

The algorithm may sometimes predict a list with many items. Therefore, it should be evaluated if the algorithm could correctly define the probability of purchase. This equation compares the items in the bill with the first five items which was predicted by the algorithm. Therefore this equation is used to check if the algorithm could correctly order the items predicted

$$Strict\ Rate\ of\ Accuracy\ (SRA) \quad = \quad \frac{P(5)}{N} \times 100\%$$

3. Weighted rate of accuracy

   This equation is used to calculate the general accuracy of a single bill prediction

$$Weighted\ Rate\ of\ Accuracy\ (WRA) = \frac{\left(3 \times RA(5)\right) + (2 \times RA(10)) + \left(RA(15)\right)}{6}$$

4. Average rate of accuracy:

   This equation is used to calculate the average accuracy of the complete set of bills. The accuracy can be calculate separately for strict accuracy and weighted accuracy.

$$Overall\ Rate\ of\ Accuracy\ (Weighted) = \sum_{i=1}^{m} \frac{WRA_i}{m} \ ; \ m = number\ of\ bills$$

5. Accuracy of a user:

   It is important to check the algorithms behavior for different users. This equation is used to separately calculate the average accuracy of a single user.

$$Average\ rate\ of\ Accuracy\ (U) = \sum_{i=1}^{m} \frac{SRA(u)_i}{m(u)} \ ; \ m \rightarrow number\ of\ bills,\ u \rightarrow selected\ user$$

$$Weighted\ Average\ rate\ of\ Accuracy\ (U) = \sum_{i=1}^{m} \frac{WRA(u)_i}{m(u)} \ ;$$
$$m \rightarrow number\ of\ bills, u \rightarrow selected\ user$$

Using the individually calculated accuracy rates, the overall accuracy of the algorithm was calculated. Below are the calculated accuracy levels.

- Average rate of Accuracy of the bills = 83.34%
- Weighted average rate of accuracy of the bills = 85.28%

Accuracies of few individual users are as follows;

- User 1
  - Average rate of accuracy of the user 1(Strict) = 78.57%
  - Average rate of accuracy of the user 1 (weighted) = 78.57%
- User 3
  - Average rate of accuracy of the user 2(Strict) = 91.11%
  - Average rate of accuracy of the user 2 (weighted) = 96.42%
- User 3

European Journal of Computer Science and Information Technology
Vol.7, No.5, pp.1-15, October 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

- Average rate of accuracy of the user 3(Strict) = 77.78%
- Average rate of accuracy of the user 3 (weighted) = 77.78%

It was observed that the aaccuracy of the algorithm increases as the number of records increase and the aaccuracy drops when purchasing a completely new item. 19 test cases out of the 30 test cases have an accuracy of 100% and when considering the weighted average 22 test cases have an accuracy greater than 75%.

Below mentioned is the complete results of the evaluation.

Table 3. Test results.

Bill #	User	N	P (5)	P (10)	P (15)	SRA / RA(5)	RA (10)	RA (15)	WRA
1	3	2	1	2	2	50	100	100	75
2	3	1	1	1	1	100	100	100	100
3	3	2	2	2	2	100	100	100	100
4	15	2	1	1	1	50	50	50	50
5	1	1	1	1	1	100	100	100	100
6	2	6	4	4	4	66.667	66.667	66.667	66.667
7	15	3	2	2	2	66.667	66.667	66.667	66.667
8	3	1	1	1	1	100	100	100	100
9	13	2	2	2	2	100	100	100	100
10	3	2	2	2	2	100	100	100	100
11	3	1	1	1	1	100	100	100	100
12	1	1	1	1	1	100	100	100	100
13	2	3	2	2	2	66.667	66.667	66.667	66.667
14	1	1	1	1	1	100	100	100	100
15	3	2	1	2	2	50	100	100	75
16	3	2	2	2	2	100	100	100	100
17	3	2	2	2	2	100	100	100	100
18	13	3	2	2	2	66.667	66.667	66.667	66.667
19	1	3	2	2	2	66.667	66.667	66.667	66.667
20	3	1	1	1	1	100	100	100	100
21	3	1	1	1	1	100	100	100	100
22	3	2	2	2	2	100	100	100	100
23	3	2	2	2	2	100	100	100	100

European Journal of Computer Science and Information Technology
Vol.7, No.5, pp.1-15, October 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

24	3	1	1	1	1	100	100	100	100
25	3	6	4	5	5	66.667	83.333	83.333	75
26	15	3	3	3	3	100	100	100	100
27	1	2	1	1	1	50	50	50	50
28	1	1	1	1	1	100	100	100	100
29	1	1	0	0	0	0	0	0	0
30	1	1	1	1	1	100	100	100	100
Average						**83.333**	**87.222**	**87.222**	**85.278**

QUALITATIVE EVALUATION

In addition to the quantitative evaluation a qualitative evaluation was also done to check if the research has achieved its objectives. A qualitative evaluation is important to evaluate the non-quantifiable aspects of the research such as user friendliness. This is also important to understand the different opinions of users, about the concept and the solution.

As the quantitative evaluation, a feedback form was given to individuals who purchase fresh items from the selected outlet on a daily basis. The individuals were given a chance to use the developed prototype system and provide their feedback on the system and the concept through the feedback form. The feedback form contains questions related to the accuracy of the results, satisfaction, usefulness and other non-functional objectives of the solution.

A positive feedback has been received for the solution. All the users have agreed that the current checkout process can be improved and the proposed solution makes the checkout process easier. The satisfaction rate of the results produce by the algorithm stands above 80% and the satisfaction of the overall application and the concept also stands above 80%.

Most of the users liked the fact that the application filters and shows a predicted item list. Also users have mentioned that it is faster and easier to use the proposed approach than to use the existing billing mechanism.

CONCLUTION

The research elaborates the design and implement of an algorithm to predict the next purchasing item list of a given user, using his/her historical purchasing data. As mentioned in the literature review numerous methods and techniques are available to determine purchasing patterns of users. However, as the algorithm should generate results real time, a time efficient method had to be used to generate the results.

Hence a simple rule base weighted rating method was used to design the algorithm. During the research it was able to identify important variable that impact the purchases of the users. This is not only important for this research but will be important for future studies as well.

The algorithm introduced is named as RFR-U model as it uses, Recency, frequency, Relevance and User as parameters to determine the results. The RFR-U algorithm is able to generate a predicted next purchasing item set of a given user within seconds. The accuracy rate of the algorithm is above 80%. The accuracy level is highly satisfactory as simple rule based methods have been used to design the algorithm.

The algorithm has been used in the proposed self-checkout application to create a personalized user interface. The feedback from the users have been mostly positive on the solutions and the results. Hence it can be concluded that the solution is able to solve the problem mentioned in the research. However, the algorithm is able to achieve the said accuracy level only if the customer is a regular customer. As the algorithm uses simple techniques, the algorithm cannot predict anomalies.

FUTURE WORK

The algorithm could be further improved by incorporating factors such as weather, to generate results. In addition the results could be used for personalized promotions.

The application can be made further personalized by filtering results according to the current selection. Many researchers suggest that customers are likely to purchase certain items together. Therefore using such mechanisms the application can be made more personalized and user friendly.

The variation of the RFR-U model can be used for business decision making, such as procurements and meal plans.

ACKNOWLEDGMENT

I would like to express my deep gratitude to my supervisor Dr. K. Lakshman Jayaratne for the valuable advices and guidance given to me at all times. I would also like to thank my colleagues for the continuous support extended to me in data gathering and obtaining feedback.

REFERENCES

[1] H. Kim, G. Yang, H. Jung, S. Ho Lee, J. J. Ahn (2019), “An Intelligent Product Recommendation Model to Reflect the Recent Purchasing Patterns of Customers”,[Online]Available:https://link.springer.com/article/10.1007/s11036-017-0986-7. [Accessed March. 24, 2019]

[2] Hughes, A. M. (1994) Strategic database marketing. Chicago: Probus Publishing Company. 1994.

[3] Z. Shi, Z. Wen, Jin Xia (2015)" An Intelligent Recommendation System based on Customer Segmentation", International Journal of Research in Business Studies and Management, Volume 2, Issue 11

[4] Marcus C (1998) A practical yet meaningful approach to customer segmentation[J]. Journal of consumer marketing, 1998, 15(5): 494-504.

[5] S. Lee, D. Lee (2018) "A personalized channel recommendation and scheduling system considering both section video clips and full video clips", [Online] Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0199748 [Accessed September. 14, 2018]

[6] M. Papagelis, D. Plexousakis (2005) "Qualitative Analysis of User-based and Item-based Prediction Algorithms for Recommendation Systems", Engineering Applications of Artificial Intelligence 18

[7] H. Lu (2014) "Recommendations Based on Purchase Patterns", International Journal of Machine Learning and Computing, Vol. 4, No. 6

[8] M. Salehi (2013) "An effective recommendation based on user behaviour: a hybrid of sequential pattern of user and attributes of product", Int. J. Business Information Systems, Vol. 14, No. 4

[9] Z. Zeng (2019) "An Intelligent E-commerce Recommender System Based on Web Mining", International journal of business and management, Volume 4, Issue 7

[10] S. Tiwari, H.M. Wee, Y. Daryanto (2017) "Big data analytics in supply chain management between 2010 and 2016: Insights to industries", [Online] Available: https://www.sciencedirect.com/science/article/pii/S0360835217305508 [Accessed July. 30, 2018].

[11] Y. Wang, L. Kung, T. A. Byrd (2018) "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations". [Online] Available: https://www.sciencedirect.com/science/article/pii/S0040162516000500 [Accessed July. 30, 2018].

[12] R. Iqbal, F. Doctor, B. More, S. Mahmu, U.Yousuf (2018) "Big data analytics: Computational intelligence techniques and application areas". [Online] Available: https://www.sciencedirect.com/science/article/pii/S0040162517318498 [Accessed July. 30, 2018].

[13] G. Manogaran, D. Lopez (2018) "Spatial cumulative sum algorithm with big data analytics for climate change detection ", Computers & Electrical Engineering, Volume65,online]Available:.https://www.sciencedirect.com/science/article/pii /S004579061730811X. [Accessed July. 30, 2018]