2018 4th International Conference for Convergence in Technology (I2CT)

# i-Shopping with Sensor Fusion for finding Customer Behavior using Deep Learning Algorithm

Santhi Venkatraman
*Deparment of Computer Science and Engineering*
*PSG College of Technology*
Coimbatore, India
vsr.cse@psgtech.ac.in

Gokul Nithin Kumar R
*Deparment of Computer Science and Engineering*
*PSG College of Technology*
Coimbatore, India
r.gokulrk9@gmail.com

Heartlin Karan Machado
*Deparment of Computer Science and Engineering*
*PSG College of Technology*
Coimbatore, India
karan_machado@yahoo.com

***Abstract*— Nowadays many people are opting for online Retail shopping because there is no need to hassle at the checkout counters and the process of shopping is fast. What if people have the liberty to purchase the required products and leave without having to spend unnecessary time at the checkout counters? This could attract more customers to the retail stores and hence increase profits. 'i-Shopping' which is an amalgamation of deep learning techniques with analytical processing aims at removing checkout counters at Retail stores. This is possible through sensor fusion with RFID tags to determine the presence or absence of an item. Deep learning algorithms can be made applicable to determine the customer behavior and assess which item the customer has picked or returned to the aisles. The products are automatically billed to the customer account on the app and the soft invoice is provided to the customers.**

***Keywords*— *i-Shopping, Customer Behavior, Deep Learning, Sensor Fusion, RFID, Android.***

## I. Introduction

Retail Stores are indispensable for today's shopping needs. The growing demand of the population in the country is met by the numerous Retail stores set up in the country. But the people who go to these stores for purchases has seen a remarkable decline during the past ten years because of the growing e-commerce and also people consider it as a waste of time to spend around an hour or so for just shopping. People spent a large amount of time worthlessly at the checkout counters where they could be doing several other useful duties. What if a person experiences the comfort of e-shopping while being able to physically see and pick up the products? Our project aims to conceal the distance between these two by removing all the checkout counters from the stores. And also provide in-store and out-of-the store recommendations to the customers. For achieving these objectives, the following things should be happened in the proposed system. Sensor fusion makes it possible for connecting all the numerous sensors to automate the process. A Radio Frequency Identification (RFID) [1] tag has its impact on the various products by assigning a unique tag to each one of it. Near Field Communication (NFC) [2] is applicable for establishing a link between the central system and the mobile phone. Deep Learning plays an important in recommending suitable products (in-store and out-of-store recommendations) to each customer.

Descriptive analysis [3] is carried out to gain an insight about the overall working nature of present day retail sales system. This analysis is done with respective to customer classes to find a working solution for each category of customers and suggest an admissible provocation to increase the sales. Once descriptive analysis arrives at a workable solution the next process includes making convenient recommendations designed as in-store and out- of-store.

## II. Data Acquisition

The dataset for the proposed system is true and authentic and is acquired from various super markets across different continents. The dataset is updated based on customer's purchase records. The Dataset has a total of 9 attributes (Table 1) and 5,41,910 rows. Further Datasets includes the Description, Quantity purchased, Unit Price, and Country of purchase of each customer. These datasets are analyzed in order to form the training sets for the prediction model.

### *A. Dataset*

The following are the datasets are being used for training and testing

- Whole foods (USA & UK) – 546MB [4]
- Publix markets(Europe and Asia) - 671MB [5]
- Spar supreme(Africa) – 325MB [6]

TABLE I. Supermarket Purchase Dataset Attributes

| Attributes | Data type | Description |
|---|---|---|
| Invoice No | Number | Bill ID |
| Description | Plain Text | Product Description |
| Quantity | Number | Quantity of the purchased product |
| Unit Price | Float | Individual price of the product |
| Customer ID | Number | Unique ID of the customer |
| Country | Number | Country of purchased product. |
| Total | Float | Total price of the product |
| Date | Date & Time | Date & Time when the product was billed |

### *B. Dataset Preprocessing*

The Datasets must be pre-processed in order to fit the data for further classification. We must first clean the data, integrate them and then transform it to the desired processing format. The Not Applicable values are removed from the dataset in order to make the data suitable for all types of processing. The Dataset is split up based on the countries and the processing is done for country based product purchases.

Description and Quantity (Table I) are two important attributes which must be taken into account for finding out

978-1-5386-5232-9/18/$31.00 ©2018 IEEE

the different classes of customers. Time from the date attribute is pivotal for suggesting suitable purchase times. These help in providing out store recommendations.

## III. Descriptive Analysis

The fundamental step for any project in Data Science is Descriptive Analysis [3]. It describes the basic features that are going to be implemented in the project. It provides simple summaries about what the project is all about. It also projects the various features involved in the project and the dependencies between the projects. Descriptive Analysis also includes Inferential Statistics that provides new inferences with proof based on raw data.

### A. Customer Types Chart

This chart deals with categorizing customers as Premium or Regular (Table II) based on their purchasing behavior. Customers who have purchased products for more than $ 1000 /- are considered as PREMIUM customers and others are considered as REGULAR customers, because Premium customers tend to purchase more products hence providing recommendations to such customers would lure them to buy that product thereby increasing profit to the shops.

TABLE II. Customer Types Chart

| Type | Premium | Regular | Total |
|---|---|---|---|
| **Count** | 1638 | 2734 | 4372 |

### B. Country Consideration Chart

This chart (Table III) includes the list of countries that would be considered in this project. As this project involves a pretty high investment into it, countries where the customer count is less would not be cost efficient. So in order to be cost efficient this analysis is done. Two features decide whether the country will be considered or not. Firstly, the customer count is taken into consideration. Countries where the customer count is greater than 500 are considered.

$$\text{CustomerIndividualContribution} = \frac{\text{GrossAmount}}{\text{CustomerCount}}$$

TABLE III. Country Consideration Chart

| Countries Under Consideration | |
|---|---|
| Total | 38 |
| Considered Countries | 16 |
| Not Considered | 22 |

### C. Prime Time Recommedation Chart

This is the most important chart for recommendation based systems. Recommendations cannot be provided to customers at any point of time, for example, providing recommendation to customers at 3:00 AM will have more probability for the customer to delete the recommendation. Hence in order to overcome this issue Prime Time is calculated for every country. For calculating the prime time, Customers purchase time is taken into consideration and its frequency is calculated. Then a graph is plotted between Frequency and Purchase Time. For example, Fig.1 summarizes the prime time for the country Australia.

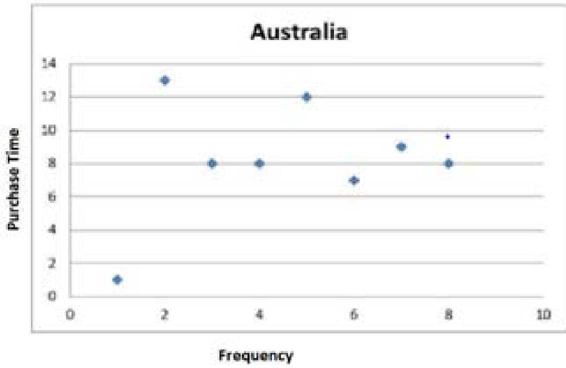


Fig. 1. Purchase Frequency Graph

### D. Product Recommendation Chart

Recommending products is based on two factors. Firstly, it depends on the count of each product brought during the period (sorted in descending order). Secondly, it based on the Cumulative Frequency of the gross total of each product. After applying these two filters, the products that are responsible for 80% of the gross total are considered for recommendation. Table IV shows the result.

TABLE IV. Product Recommendation Chart

| Product Based Recommendation | |
|---|---|
| Total products | 42 |
| Products for recommendation | 8 |
| Products not considered | 33 |

TABLE V. Product Recommendation For Premium Customers Chart

| Sl. No | Description | Unit_Price | Discount* % | Reduced_Price |
|---|---|---|---|---|
| 1 | SKULLS DESIGN COTTON TOTE BAG | 2.25 | 44.40% | 1.25 |
| 2 | SET OF 6 SOLDIER SKITTLES | 3.75 | 15.74% | 3.16 |

TABLE VI. Product Recommendation For Regular Customers Chart

| Sl. No | Description | Unit_Price | Discount* % | Reduced_Price |
|---|---|---|---|---|
| 1 | 20 DOLLY PEGS RETROSPOT | 1.25 | 45.44% | 0.68 |
| 2 | ASSORTED BOTTLE TOP MAGNETS | 0.36 | 47.99% | 0.19 |

* For testing purposes the Discount % is provided at random to the customers.

### E. Discount for Premium and Regular Customers Chart

After filtering out the products that are to be recommended to the customers, they are again categorized based on which type of customers (whether premium or regular) and what products they buy. After this categorization the list is sent to the respective shops so that whenever the shop admins update the discounted value for a particular product, the recommendation will be automatically

978-1-5386-5232-9/18/$31.00 ©2018 IEEE

sent to the respective customers during their prime time (Table V and VI).

After analyzing the results of the descriptive analysis we gain insight into the situations and timings suitable for provisioning of element based recommendation supported with user specific timing specifications. This paves the way for providing customer-tailored recommendations.

## IV. RECOMMENDATIONS: INSTORE & OUT-OF-STORE

The Supervised Machine Learning Algorithm: Artificial Neural Networks (ANN) [7] is used to categorize the class of customers as premium and regular. This enhances the durability for making appropriate recommendations to the customers since only the specific class will be analyzed for making the suitable recommendations.

### *A. Artificial Neural Networks(ANN)*

Product Recommendation plays a pivotal role in finding out which products the particular customer picks up and places back. So Artificial Neural Networks makes it possible by analyzing the past shopping behavior of the customers.

**ANN Formula:**

$$C = \frac{1}{N}\sum_{i=1}^{N}(f(Xi) - Yi)^2$$

**Algorithm:**

1. Net result of hidden state 1

$$NETh1 = w1 * i1 + w2 * i2 + b1 * 1.$$

where,
w - weights of layer n
i - input layer activation function
h - hidden layers

2. Output emergence of hidden state 1 using the logistic function

$$OUTh1 = \frac{1}{1 + e^{-NETh1}}$$

3. Net Output result

$$OUTo1 = \frac{1}{1 + e^{-NETo1}}$$

4. Summation of next step average of the target with respect to the output which gives the total error.

$$Etotal = \sum \frac{1}{2}(target - output)^2$$

5. Target output for o1

$$Eo1 = \frac{1}{2}(targeto1 \; - outo1)^2$$

6. Total summation up output comprising of all the hidden states

$$Etotal = E_{01} \; + E_{y02}$$

**Hidden Layer:**

1. Total change calculation

$$\frac{\partial Etotal}{\partial w1} = \frac{\partial Etotal}{\partial OUTh1} * \frac{\partial OUTh1}{\partial NETh1} * \frac{\partial NETh1}{\partial w1}$$

$\partial$ - Change in calculations

2. Change of Total error with respect to the output

$$\frac{\partial Etotal}{\partial OUTh1} = \frac{\partial Eo1}{\partial OUTh1} + \frac{\partial Eo2}{\partial OUTh1}$$

3. Change in error calculation

$$\frac{\partial Eo1}{\partial OUTh1} = \frac{\partial Eo1}{\partial NETo1} * \frac{\partial NETo1}{\partial OUTh1}$$

4. Total change of net input of o1 change with respect to w5

$$NETo1 = w5 * OUTh1 + w6 * OUTh2 + b2 * 1$$

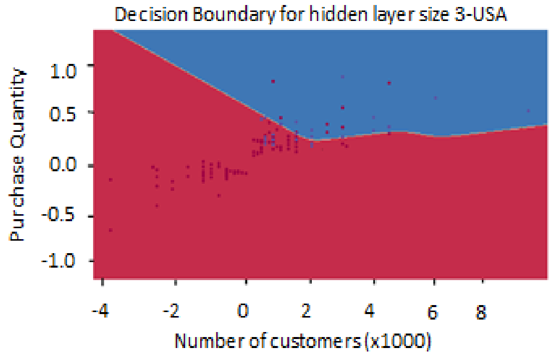


Fig. 2. ANN Decision Boundary

After recognizing the different classes of customers, the decision boundary is constructed segregating the different classes of customers. Once a customer enters a shop his shopping trends are immediately recognized and plotted accordingly. Fig.2 shows the Premium and Regular customers' categories.

The recommendations are provided as part of the automated process with differentiations between in-store and out-of-store. All these recommendations are affixed with the application which makes it easy for the users of the system to interact with the system and accept or decline a particular interest of purchase. The recommendations are originated based on the proximity of the customer to the store (viz) whether inside the premises or present above a certain distance from the store. On considering the customer's presence inside the store we have in-store recommendations.

### *B. In-store Recommendations*

After splitting the different classes of customers the recommendations are provided as in-store when a customer finds himself inside the store.

#### *i. Similarity Based Recommendations*

Similarity based recommendations is based on the principle of suggesting similar products which the customer has picked up but offered at a reduced price. Multiple recommendations should not be encouraged as it gives the wrong impression to the customers that the shop is trying to push its products just for the sake of profit and doesn't care anything about customer's satisfaction.

978-1-5386-5232-9/18/$31.00 ©2018 IEEE

*ii. Apriori Algorithm*

Apriori Algorithm [8] recognizes Hidden Patterns by considering different attributes in the dataset and provides their occurrence and confidence levels based on their associatively between the attributes. Table VII shows the Apriori functionalities for the dataset.

**Algorithm:**

$C_k$ = candidate itemset of size k.
$L_k$ = frequent itemset of size k.
$L_1$ = {frequent items }
k=1
while ($L_k \neq \emptyset$)
{
$C_{k+1}$ = candidates generated from $L_k$ .
For(each transaction)
Increment count of candidates in $C_{k+1}$
$L_{k+1}$ = candidates in $C_{k+1}$ with sufficient support
k=k+1
}

1. Support itemset frequency.
   Support - Popularity of an itemset
2. Confidence X→Y
   Confidence - Likeliness that itemset Y is purchased when X is purchased
   Conf(X→Y)=supp(XY)/supp(X) i.e P(Y/X)
3. Lift association of X and Y.
   lift - Likeliness that itemset Y is purchased when X is purchased while controlling popularity
   Lift(XY)= supp (XY) / (supp(X) * supp(Y))
4. Conviction X→Y
   Conv(X→Y) =(1-supp(Y)) / (1-conf(X→Y))

TABLE VII. APROPRO FUNCTIONALITIES

| Attribute | Function |
|---|---|
| Description | Similar product recommended |
| Time | Suitable timing recommended |
| Country | Country of recommendation |

## C. *Python Generator Function*

For making appropriate suggestions we apply the python generator function [9] instead of standard association function since each suggestion must be made in accordance with the timestamp of the central system and coordination must be maintained. The Apriori algorithm with the standard association function takes the entire set of data as input as processes it. This increases the computation time hence to overcome this situation, hence to overcome this situation python generator function is used that take each tuple from the database as input and processes it. This results in quicker computation and reduces time complexity.

## D. *Out-Store Recommendations*

Out Store recommendations refer to the process of suggesting product recommendation when the customer is out of the store. Apart from making these recommendations the stocks are frequently checked to see if the discounts are applicable or if any case the discounted products are out-of-stock.

**Discount percentages:**

The Discount percentages are given based on the suggested discount ranges published from the supermarket themselves.

TABLE VIII. SUITABLE TIMING (COUNTRY WISE)

| Country | Time |
|---|---|
| Australia | 9am |
| Belgium | 1pm-3pm |
| Norway | 12pm-2pm |
| Sweden | 11am-3pm |

Out store recommendations are provided when the primary location of the user is outside the bounded limits of the premises and suggest suitable purchase timings (Table VIII) for ease of use to the customer.

# V. IMPLEMENTATION

## A. *Android App Layout*

The customers can have their E-Cart viewed through the provision of an Android Mobile Application which enhances their hassle free shopping. This makes it possible for customers to make payments for their purchased products through various payment gateways online. The android phone will be automatically connected to the shop's WIFI when the customer logs in. The customers have a special way of logging on to their profile via the Android Application. After they have typed the username and password, a unique QR code [10] will be displayed onto their screen. The customer then needs to scan the QR code (Fig. 4) with the shop's QR code reader for successful authentication.


Fig. 4. QR-Code scan Page

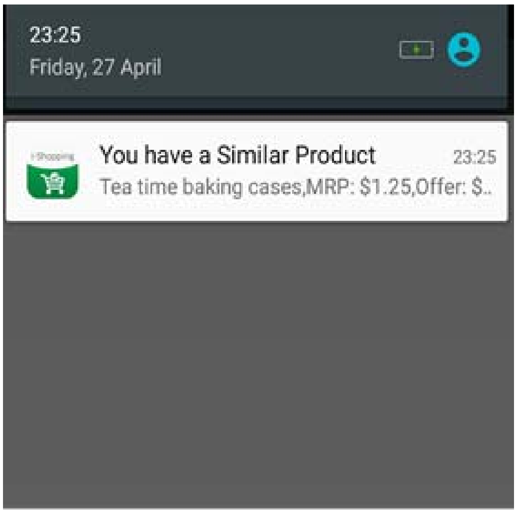


Fig. 5.In-store Recommendation

In-store recommendations (Fig. 5) are suggested based on a product which the customer picks up and similar products are provided on purchase list. When a customer, Jack picks up a product from the store says a Fiber Baking Case, products which are similar to the purchased product say Tea-Time Baking Cases are suggested to the customer. If the customer purchases a recommended product then the weight of that node is increased along with all of its successors through backward propagation. If the product is not purchased then that node is blocked and nearby neighboring nodes is explored. Deep learning is used in this case to find the association between products that belong to a different category. So when Jack purchases a recommended product he not only fulfills his need but also saves some money since the discounts come into play. These recommendations fade away when Jack exits the store.

Let's consider a customer Harry visited the store recently and the procedure through which our system interacts with Harry to lure him back to the store is achieved through out-of-store recommendations (Fig. 6). The classes of customers are initially segmented through descriptive analysis. After separating out the different possible classes of premium and regular customers the next iterative process is in making the suitable recommendations. Our customer Harry falls in one of the mentioned classes. This process now invokes the ANN algorithm which finds its application in placing Harry in one of the available classes. Through this we can narrow down our possible recommendations tailored to each of the customer in our case, Harry. Apriori algorithm is considered for making recommendations at the appropriate time respective to each country. Harry belongs to Premium class and is hailing from Belgium so the time is suggested based on these traits. Hence the products which premium customers frequently purchase are suggested to Harry at the purposeful hours with discounted prices.

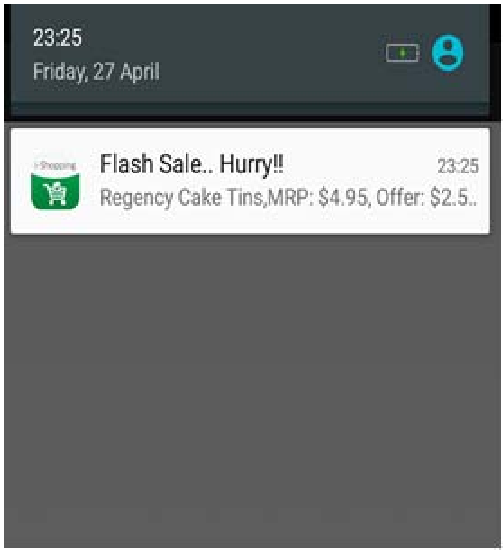


Fig. 6.Out-of-store Recommendation

## B. NFC System Layout

The customer logs in through the QR code present (Fig. 7) at the entrance of the shop and then interaction between customers mobile and store's cart takes place through NFC medium (Fig. 8). Later the customer's logs are stored into the store's database (Fig. 9).

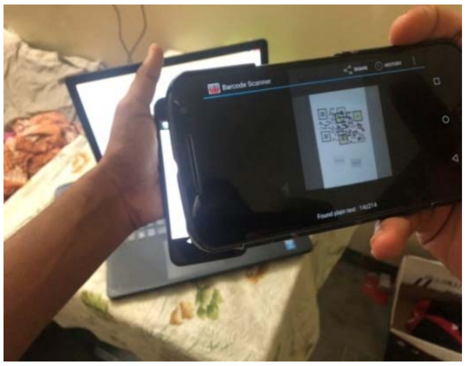

Fig. 7.Customer logs in through QR code


Fig. 8. Mobile- Cart Interaction

978-1-5386-5232-9/18/$31.00 ©2018 IEEE

| entry | out12 | sno | id1 |
|---|---|---|---|
| 2018-04-27 02PM:56:40 | NULL | 0 | 14z228 |
| 2018-04-27 02PM:56:40 | 2018-04-27 02PM:56:40 | 1 | 14z214 |
| 2018-04-27 02PM:56:40 | NULL | 2 | 14z214 |
| 2018-04-27 02PM:56:40 | NULL | 3 | 14z207 |
| 2018-04-27 02PM:56:40 | 2018-04-27 02PM:56:40 | 4 | 14z205 |
| NULL | NULL | NULL | NULL |

Fig 9 Log updation into database

### *C. RFID System Layout*

The customer picks the products from the store's isle and places it inside the cart. Once the product passes through the RFID reader (Fig. 10) embedded inside the cart it automatically gets billed into the customer's account. If the customer wishes to replace the product he may do so and the appropriate changes are reflected in the account. The selected products for checkout are displayed (Fig.11) in the Android Interface after which the customer can make their way towards online payment.

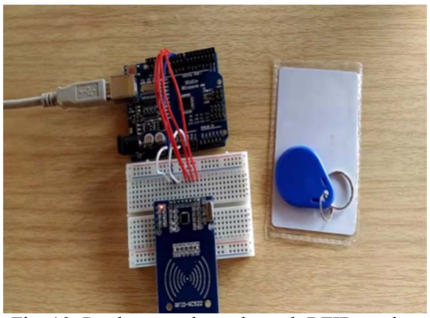

Fig. 10. Product purchase through RFID reader

```
File  Edit  Shell  Debug  Options  Window  Help
Python 3.6.1 (v3.6.1:69c0db5, Mar 21 2017, 17:54:52) [MSC v.1900 32 bit (Intel)
Type "copyright", "credits" or "license()" for more information.
>>>
 RESTART: C:/Users/DELL/AppData/Local/Programs/Python/Python36-32/purchase.py
Silver Capstone wrapped bag
>>>
 RESTART: C:/Users/DELL/AppData/Local/Programs/Python/Python36-32/purchase.py
Doritos 150gm Hot Sauce
>>>
 RESTART: C:/Users/DELL/AppData/Local/Programs/Python/Python36-32/purchase.py
Pack of 12 Hangers
>>>
```

Fig 11. Product purchase display

## VI. CONCLUSION

Shopping can be made user friendly if all the unnecessary time lags at various spots can be avoided and the products recommended according to the user's ideology. Also, if a customer can purchase something without having to wait for it then it urges them to buy that product. Hence i-Shopping would drive massive changes in the entire shopping industry. Analyzing customer’s behavior would be a key for better marketing strategies and also helps to develop the emerging e-marketing. Future extension of the system would be to implement the system with AI algorithms which would automatically recognize your future shopping trends and make product suggestions accordingly.

## REFERENCES


[1] Richard Herschel, Patricia D. Rafferty: Understanding RFID Technology within a Business Intelligence Framework. Intelligent Inf. Mgt. 4 (2012) 407-414.

[2] Hongwei Du: NFC Technology: Today and Tomorrow. Int. J. Fut. Comp. and Communi., 4 (2013) 351-354

[3] https://towardsdatascience.com/intro-to-descriptive-statistics-252e9c464ac9

[4] https://data.hawaii.gov/api/views/vcaz-dwuy/rows.csv?accessType =DOWNLOAD

[5] https://www.kaggle.com/shazadudwadia/%20superm arket

[6] http://archive.ics.uci.edu/ml/machine-learning-databases/00352/ Online%20Retail.xlsx

[7] Robert Hecht-Nielsen: Theory of the back propagation neural network. Int. 1989 Joint Conference on Neural Networks 1 (1989) 593-605.

[8] Karandeep Singh Talwar, Abhishek Oraganti, Ninad Mahajan: Recommendation System using Apriori Algorithm. Int. J. Sci. Res. and Dev. 3 (2015) 183-186.

[9] https://www.programiz.com/python-programming/ generator

[10] https://pypi.org/project/qrcode/