Project Report

On

Ecommerce Product Recommendation Chatbot

Developed By:
Himanshu Kriplani (15012121007)

Abhi Patel (15012121013)

Jhanvi Patel (15012121016)

Pratik Patel (15012121019)

Guided By:
Prof. Prachi Pancholi

Submitted to
Department of Computer Science & Engineering
U V Patel College of Engineering

GANPAT UNIVERSITY
॥ विद्यया समाजोत्कर्षः ॥
U V PATEL COLLEGE OF ENGINEERING

Year: 2019

CERTIFICATE

This is to certify that the report entitled **"E-commerce Product Recommendation Chatbot"** by Himanshu Kriplani (15012121007), Abhi Patel (15012121013), Jhanvi Patel (15012121016) and Pratik Patel (15012121019) of Ganpat University, towards the fulfillment of requirements of the degree of Bachelor of Technology – Computer Science and Engineering, is record of bonafide final year Project work, carried out by them in the Computer Science and Engineering Department. The results contained in this Project have not been submitted in part or full to any other University/Institute for award of any other Degree.

Name & Signature of Internal Guide

Name & Signature of Head of Department

Name & Signature of Principal

Place:
Date:

ACKNOWLEDGEMENT

A major project is a golden opportunity for learning and self-development. I consider myself very lucky and honored to have so many wonderful people lead me through in completion of this project. First and foremost, I would like to thank Dr. Hemal Shah, Head of Department, Computer Science and Engineering, who gave us an opportunity to undertake this project. My grateful thanks to Prof. Prachi Pancholi for his/her guidance in project work Ecommerce Product Recommendation Chatbot, who despite being extraordinarily busy with academics, took time out to hear, guide and keep us on the correct path. We do not know where would have been without his/her help. CSE department monitored our progress and arranged all facilities to make life easier. We choose this moment to acknowledge their contribution gratefully. We are also grateful to Angshuman Chatterjee for their kind help.

Signature of Students

Himanshu Kriplani (15012121007)

Abhi Patel (15012121013)

Jhanvi Patel (15012121016)

Pratik Patel (15012121019)

Contents

List of Figures

CHAPTER:1
INTRODUCTION

1. INTRODUCTION

E-commerce nowadays has been booming as the world has moved towards digitization. It has also opened doors to globalization to a literal huge extent. E-commerce has grown with the growth in technologies. In the era of big data and analytics, the power which a data driven business has been exponentially increasing. E-commerce is the answer to the solution to easy and accessible to everyone business system. Faster shipping and better delivery logistics are the key to satisfy any user which has been taken care of by the e-commerce giants. Greater integration of AI and Machine learning has played a vital role in development of systems which can benefit both the Business and Business user as well. These days, it's becoming increasingly difficult to talk about mobile without mentioning voice search. In 2019, voice will be one of the leading drivers of innovation in the ecommerce space – and not just on mobile. Loyalty bonus and discounts are one of best marketing tactics used today. Our project is one of a kind which facilitates the user by creating more personalized content. Content here means product, user get personalized recommendation of the products. With an attempt to increase user interaction and ease of usage, we built a chatbot interface for recommendation engine which is trained on data from Retail Rocket an e-commerce management tool.

Recommendation system [1] for Explicit dataset can be implemented using basic coding and if-else rules. But, building a recommendation system is arduous when the data is implicit. Explicit data means a data in which we have ratings, scales, reviews, and lot of other detail on which we can decide the best product. Implicit data [2] as the name suggest doesn't contain all the details. Implicit dataset contains user and item interactions on which we need to build a system for recommendation. Our dataset is an implicit data-set which contain events showing a user's interaction with a product(item) in 3 particular actions viz. View, Transacted and added to cart. The following section describes about project scope, data exploration and methodology. Moreover, the later section shows the project conclusion and future scope. We studied various methods which have been developed before. One of the methods used web mining [4] data but our data is not as similar as that.

CHAPTER:2
PROJECT SCOPE

2. PROJECT SCOPE

We plan to create a recommendation system for retail rocket's ecommerce dataset which could be used as to ease the search of products. Chatbot will act as an interface to our system helping out users and giving easy access to product list with recommendations. Further, we plan to use genetic algorithm to improve the recommendations of the products.

Figure 2.1: Importance of Recommendation

We, can see in the figure 1, Importance of product recommendation [5]. The research shown by Accenture, suggests that a person will 58% a product if the product is recommended the product on past purchases. Also, the he/she will buy the product for a probability of 65% if he/she knows the purchase history. Thus, Our recommendation engine will recommend the users using the chatbot and increase the overall probability for a person to buy an item.

CHAPTER: 3 SOFTWARE & HARDWARE REQUIRMENT

3. SOFTWARE AND HARDWARE REQUIREMENTS

We have used python as our base tool to do data exploration and create recommendation engine. As a development environment, we have used Anaconda product set such as Spyder and Jupyter notebook. Dialog-flow is an online platform where we have built our chatbot features and used its services to connect it to multiple platform. Flask is a web framework which has been used as an interface between dialog flow and our recommendation system in python.

3.1 Advantage of Dialogflow:

- Delivers natural and rich conversational experiences
- Understands what users are saying about machine learning
- Works with an array of platforms
- Offers cross-device support
- Helps chatbots to speak 14+ languages

3.2 Advantage of flask:

Flask is light-weighted web-framework of python. It is one of the fastest web-framework of python. With python, it gives a huge flexibility in creating web applications.

3.3 Other Requirements:

3.3.1 Software Requirements

- Spyder
- Flask library
- NGROK
- Jupyter notebook
- Seaborn
- Dialog-Flow platform access
- Web Browser/Android/iOS Phone

3.3.2 Hardware Requirements (minimum):

- 4GB RAM
- i3 Processor CPU
- OS (Windows, Linux, Mac)

CHAPTER: 4
PROJECT PLAN

4. PROJECT PLAN

Down below is the graphical timeline of our project Ecommerce Product Recommendation System (EPRC). We now have our chatbot available on Facebook, telegram and web. We can integrate the chatbot using just like adding a microservices to any website.

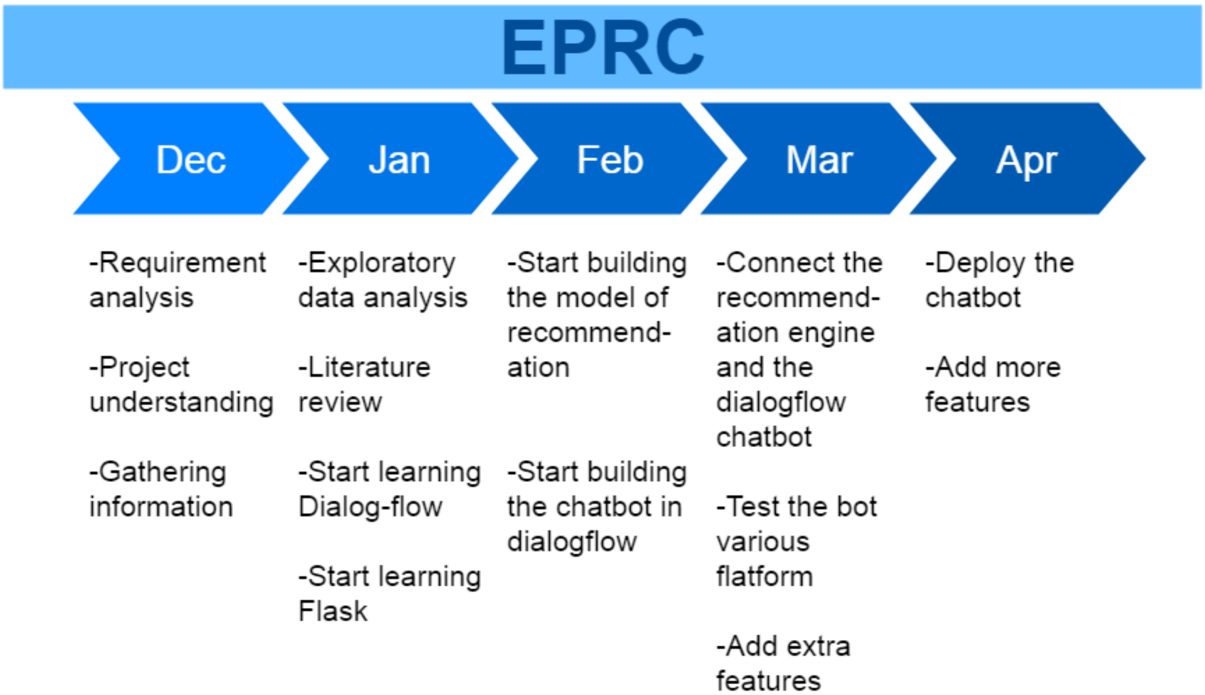

Figure 4.1: Project Plan

Our chatbot recommends products based on related users and its history. Also, a user can find out similar products of a particular product. Moreover he can lookup similar profiles for product recommendations.

CHAPTER: 5
DATA EXPLORATION

5. DATA EXPLORATION

The dataset we have is fully hashed and consists of only numbers. We have an event related data which shows the interaction between users and items based on 3 events (view, add to cart and transacted). We have a category tree, which shows the categories of different items. Items properties files shows the properties of all the items.

We also have used 2 other datasets not given to us. As the we only have user id and item ids, we used a list of products and list of names to replace ids with names. This way we the product more user friendly.

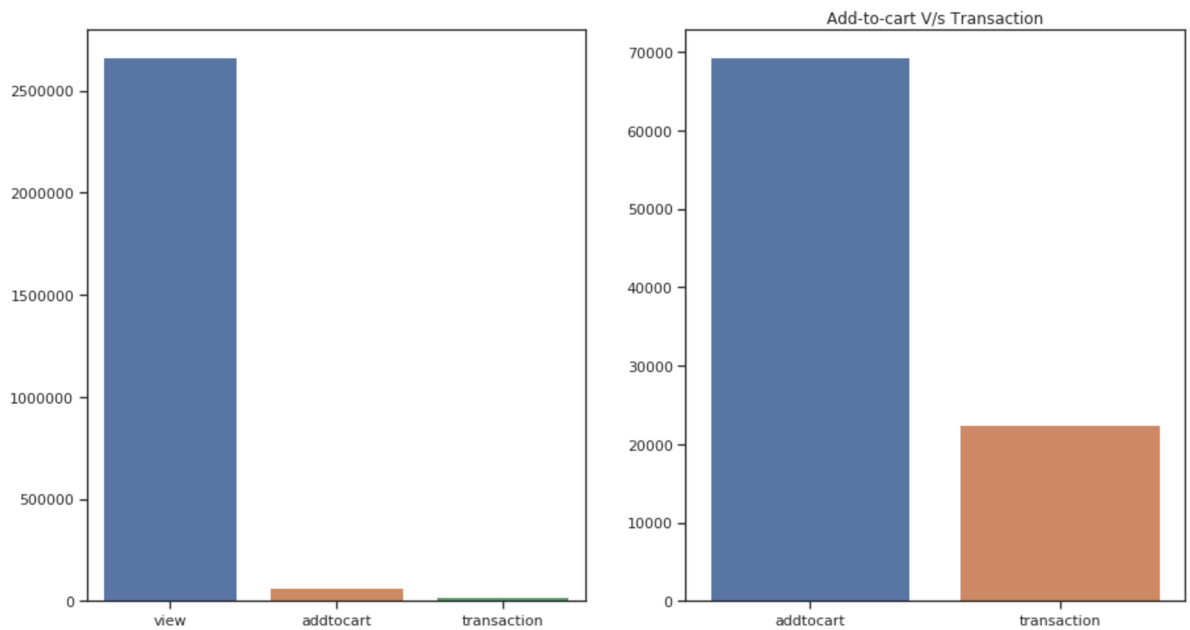

Figure 5.1: Count of Different Events

We can see in Figure 3 the count of different events in the events file. To get a better understanding of add to cart and transaction count, We plotted that graph to compare it.

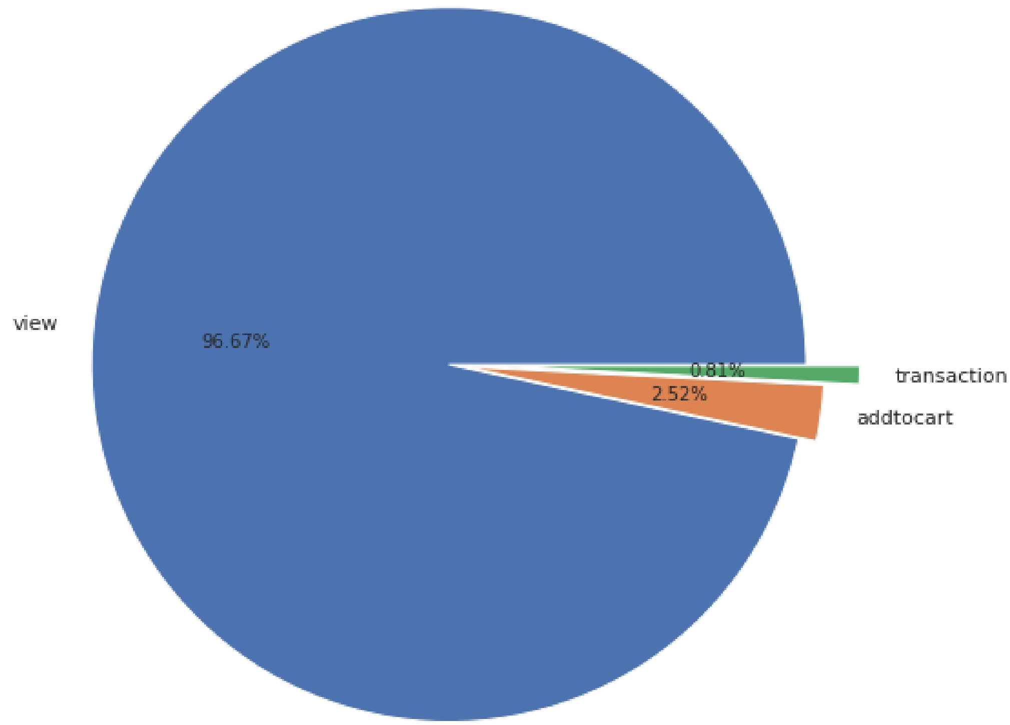

Figure 5.2: Distribution of Events

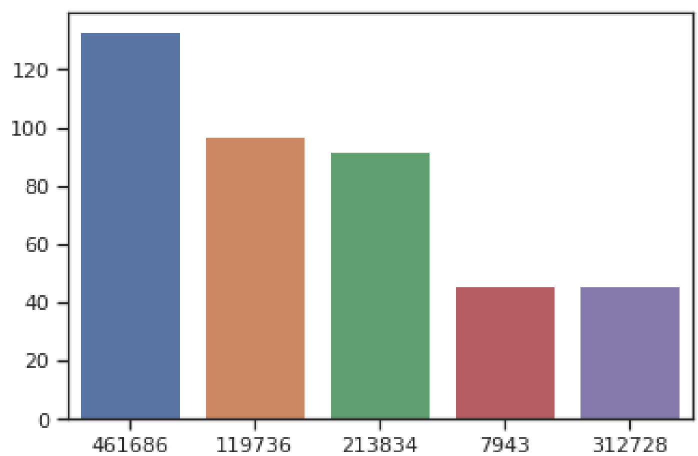

Figure 1.3: Top5 Transaction Item ID

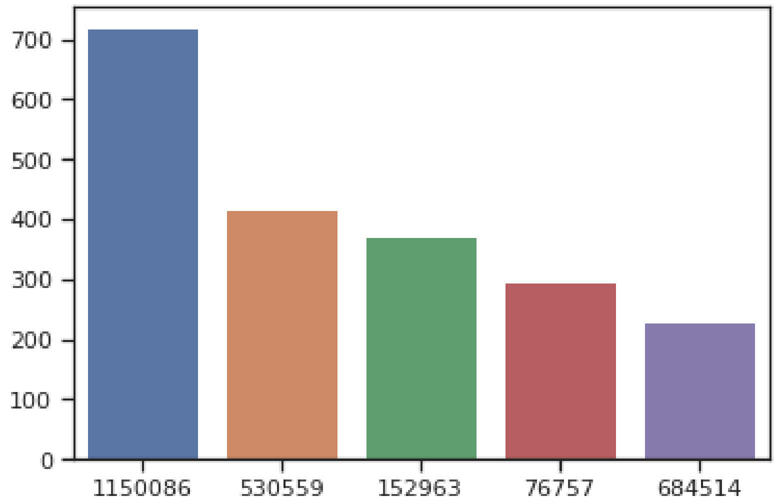

Figure 5.4: Top 5 Visitors Who Transact Any Item

```
events[events.event == 'transaction'].itemid.value_counts().describe()
```

```
count    12025.000000
mean         1.867526
std          2.710560
min          1.000000
25%          1.000000
50%          1.000000
75%          2.000000
max        133.000000
Name: itemid, dtype: float64
```

Figure 5.5: Description of Transactional Data

CHAPTER:6

METHODOLGY

6. METHODOLOGY

Systems based on Implicit data has been catching a lot of eyes in the market due to data collection techniques. Each organization has understood the value of data and is collecting a lot of data. The unstructured data can have some game changing insights. Thus, an implicit data-based recommendation system would make things a lot of easier for the organization. Because, they wouldn't rely on user's survey forms or other methods involving user which may be manipulated/tampered/duplicated.

Interactions of users and product can be easily tracked and stored. On such type of data, we have created a recommendation system with as a chatbot interface.

6.1 Content based Recommendation engine

With the advancements in data storage and wrangling techniques all over the world, Companies are trying to save each and every element or property related to applications. By storing the previously viewed/added to cart/ transaction made by a particular used can be really and helpful to make predictions on various needs of users. This is a basic algorithm about how this engine will work:

1. Find out similar products based on actions and category values
2. Look for the previous product viewed or the product for which the used needs suggestion
3. Sort and filter top 5 products based on the product information given by the users
4. Recommend those products to the user

6.2 Collaborative Based Recommendation engine

This method involves working with the User profiles and the product interest by the user. Firstly, I will find out users based on various actions performed. Unfortunately, we don't have any other parameter to work on as the dataset is limited. Based on actions we create list of profiles matching to a particular user. After studying the products used by the other users in the top matching list of users, we can recommend those products to the particular user.

6.3 Alternative Least Squares Method

Alternative least squares [2] is matrix decomposition technique which help in factorization of complex matrix.

$$\begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} * \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

Implicit data is a complex matrix which we factorize it to 2 matrices. One of which is User interaction data and the other is item interaction. We factorize a user v/s test dimension into 2 matrices which represent user and its features and the other represent relation between item and its features.

ALS is a repetitive process of getting a perfectly factorized data. By randomly assigning the values in ***H*** and ***K*** and using *least squares* iteratively we can arrive at what weights yield

the best approximation of ***T***. The *least squares* approach is to find the minimum value in a in a sum of squared distances loss method
With the ALS approach we iteratively alternate between optimizing **H** and fixing **K** and vice versa. We do this repetitively to get closer to ***T = H x K.***
The paper highlights following information:
First of all we define the Preference score and the confidence using these 2 following equations.

$$P_{ui} = \begin{cases} 1 & R > 0 \\ 0 & R = 0 \end{cases} \text{ where } R = r_{ui}$$

$$C_{ui} = 1 + \alpha r_{ui}$$

$$\min_{y*,y*} \sum_{u,i} C_{ui} P_{ui} - X_u^T Y_{i)}^2 + \lambda \left(\sum_{u} || \ x_u \ ||^2 + \sum_{u,i} || \ y_i \ ||^2\right)$$

We find the user(x) vector and the item(y) vector by differentiating the above cost function by x and y respectively.

$$X_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$Y_i = (X^T C^i X + \lambda I)^{-1} X^T C^i p(i)$$

We transform the above equation into the following form.

$$X_u = (Y^T Y + YT(C^u - I)Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$Y_i = (X^T X + XT(C^i - I)X + \lambda I)^{-1} X^T C^i p(i)$$

Finding the similar items can be done using below formula and similarly the next formula helps in making recommendations

$$score = V.V_i^T$$

$$score = U^i.V^T$$

We can find the similar users and similar items using above formula and get recommendations.

6.4 LightFM

LightFm [3] is a Python implementation of a number of popular recommendation algorithms. LightFM includes implementations of BPR and WARP ranking losses (A loss function is a average of cost of each iteration the model takes to increase the prediction accuracty.).

6.4.1 Types of Loss function in LightFM

- **Bayesian Personalized Ranking pairwise loss:** This loss function when we have positive interactions. It increases the difference between the predicted examples in which one is a positive one and the other is negative example selected arbitrary
- **k-OS WARP**: It is similar to WARP but uses the k-th positive datapoint in the context of a visitor for pairwise updates.
- **Weighted Approximate-Rank Pairwise loss:** Iteratively increase the rank of positive data points by sampling negative points.

6.4.2 Algorithm

Follow these steps for each item and user:

1. We find the latent representation of the features of Users

$$q_u = \sum_{j \in f_u} e_j^u$$

2. Than we find the latent representation of the features of the Items.

$$P_i = \sum_{j \in f_i} e_j^I$$

3. We than find the bias using following formulas.

$$b_u = \sum_{j \in f_u} b_j^U \qquad\qquad b_i = \sum_{j \in f_i} b_j^I$$

4. Then we find the relativity score using the following function. Here the function can be any identity function. The author has used sigmoid function.

$$\hat{r}_{ui} = f(q_u * p_i + b_u + b_i)$$

We 250 epochs to train the lightfm model, we got an accuracy of 99.3% on training data and 81% in testing dataset. We used WARP loss function.

6.5 DialogFlow

Dialogflow is a platform to built chatbots based on user needs. It has huge variety of features. It uses machine learning to enhance the user experience by understanding natural languages.

6.5.1 Advantages

- Delivers natural and rich conversational experiences.
- Understands what users are saying about machine learning.
- Works with an array of platforms.
- Offers cross-device support.
- Helps chatbots to speak 14+ languages.
- Helps to track chatbot's performance with built-in analytics tool.

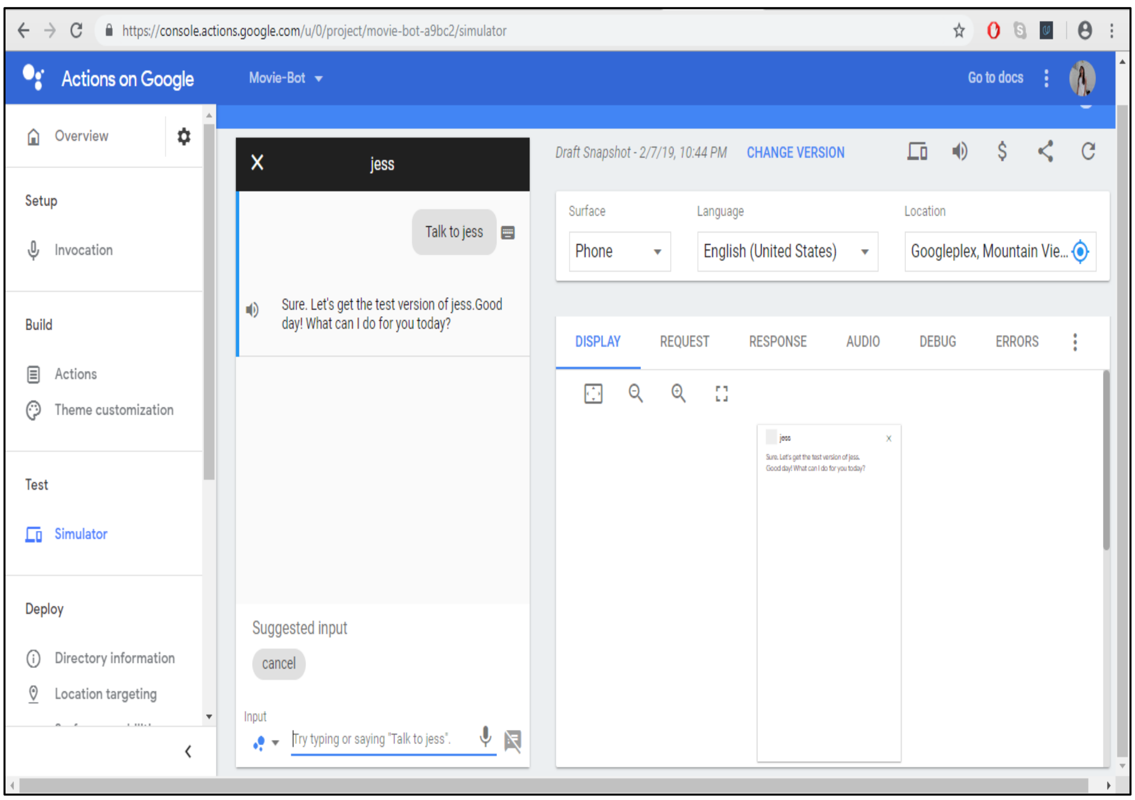

Figure 6.1: Dialogflow

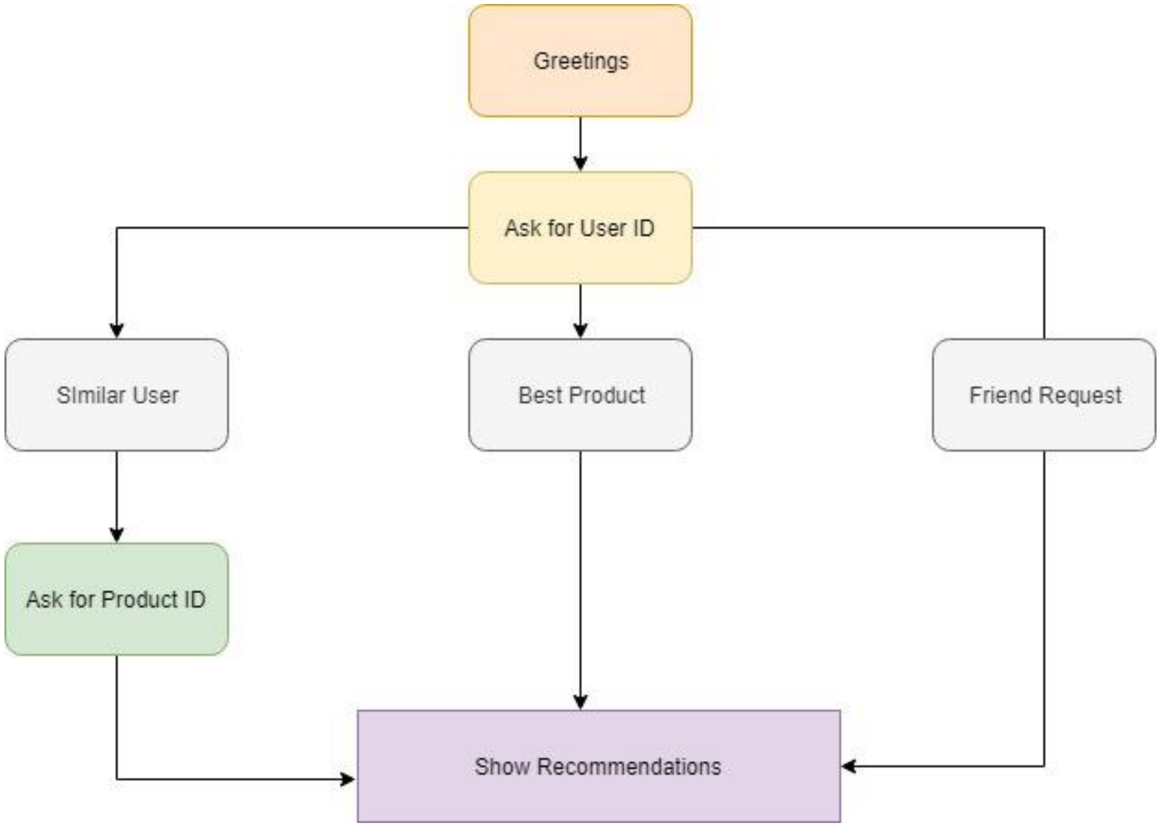

Figure 6.2: Flow of Working of the Bot

We have linked the chatbot to google assistant making the need of storefront apps evitable. Here is how our bot would look in an android phone.

Also, we integrated with Facebook and telegram as shown in the figures below

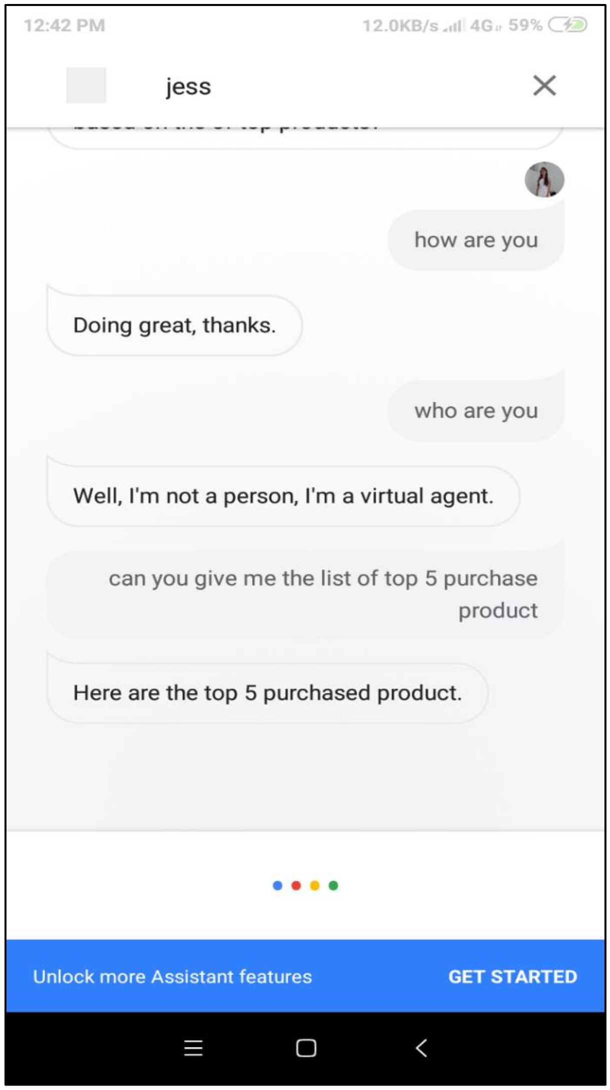

Figure 6.3: Google assistant

6.5.2 Flask and Dialogflow

To have an alternative for a web-app we used dialogflow API to create a flask based chatbot. Flask is a light-weight web framework built in python. Its one of the fastest web frameworks in python. The following screen shot shows the chatbot we built.

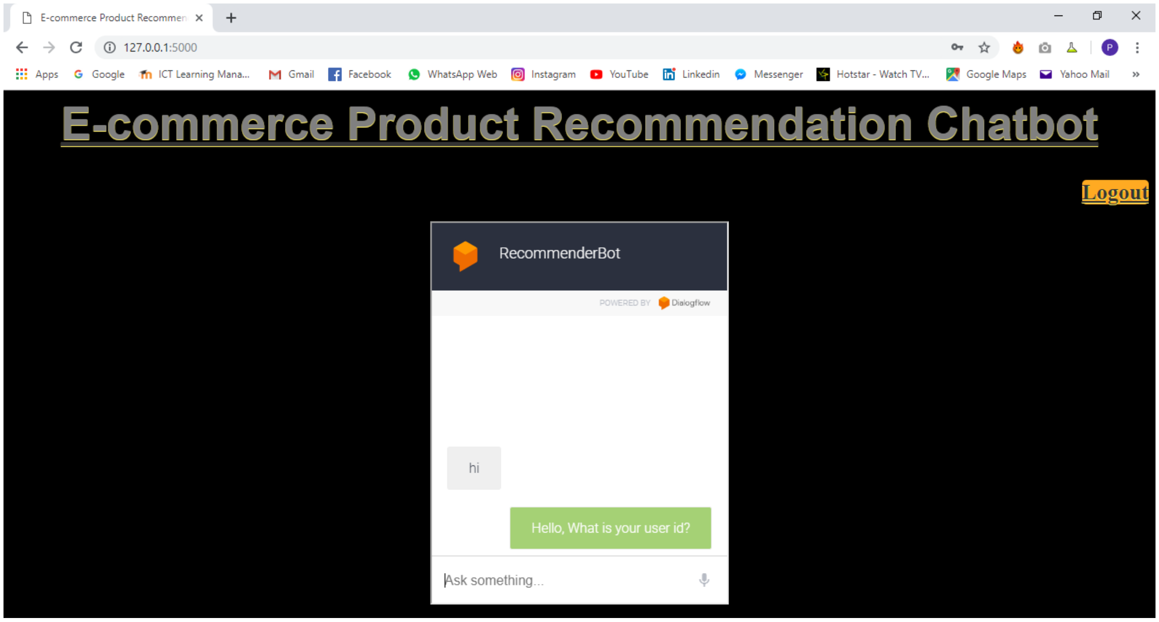

Figure 6.4: Web-Demo

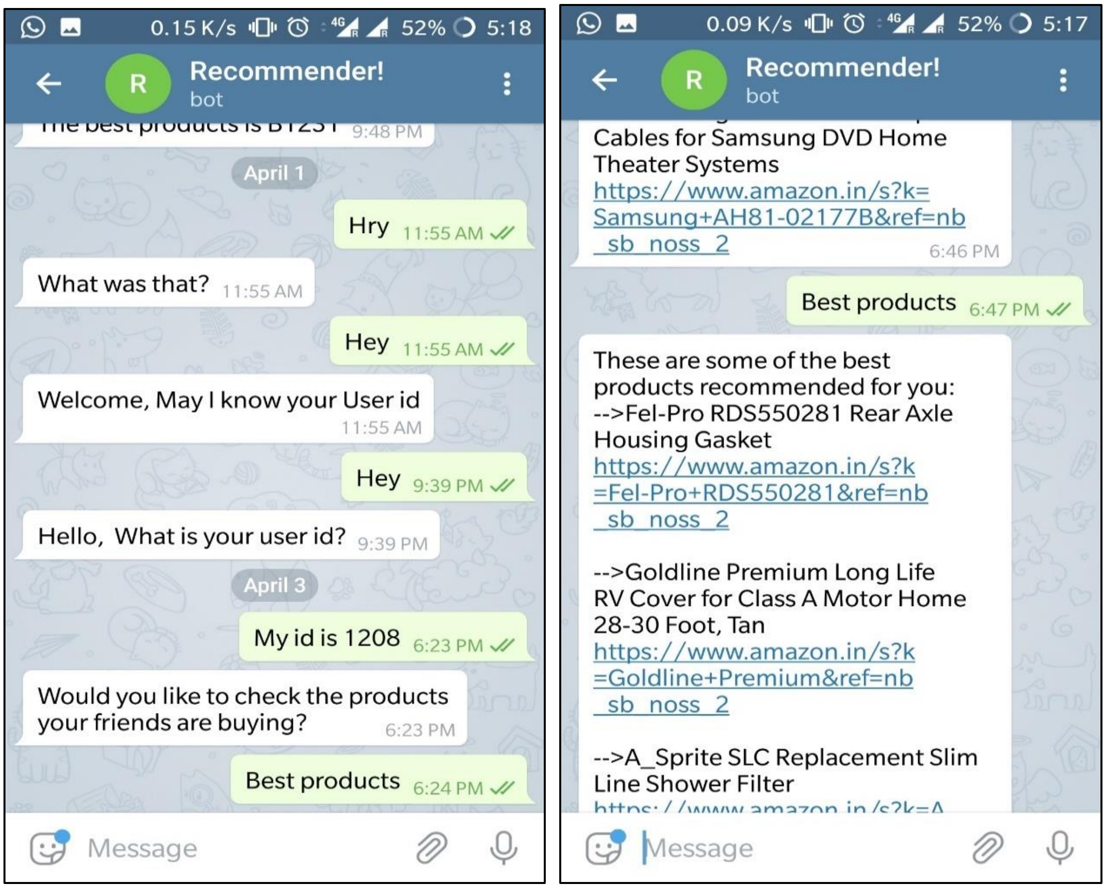

Figure 6.5: Telegram Demo

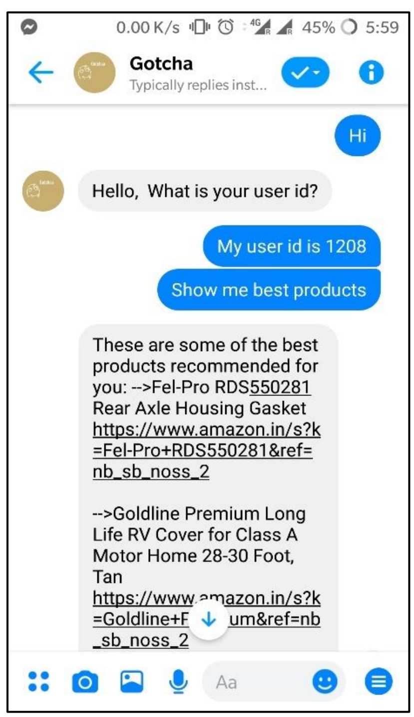

Figure 6.6: Facebook Demo

CHAPTER:7 CONCLUSION AND FUTURE WORK

7. CONCLUSION AND FUTURE WORK

We have built a chatbot which can be used as an interface to recommend products to users based on there interactions with products over the platform. We also find similar user profiles and suggest them product from those similar profiles those who have bought (or interacted in any manner with) the product. This chatbot can be an important feature to the ecommerce which suggest products which are outputted by recommendations engine at a personalized level. We can have this chatbot on a number of platforms making the need of store-front app inevitable.

The recommendation engine built using ALS and LightFM were based on implicit data, which can give a huge edge to todays growing tech presence of all the companies. With more and more interactive data rather than explicit data(likes/comments/ratings), we can discover unparalleled patterns in user-item interaction history which is much more each to store compared explicit data. With 99.3% accuracy on training dataset, we got an accuracy of 81% on test dataset on recommending the perfect product in LightFM.

7.1 Future Scope

Further, we plan to make our model to work on streaming data also. With advances in real time data, we plan to create real-time recommendations. Both, ALS and LightFM gave good results as far as our data is considered. Thus, we wish to create such a robust recommendation engine that would work with other datasets too. Moreover, our system could be integrated in any ecommerce product.

CHAPTER:8

ANNEXURE

8 ANNEXURES

8.1 References:

1. F.O. Isinkaye et al., Recommendation systems: Principles, methods and evaluation, Egyptian Informatics Journal
2. Yifan Hu, Collaborative Filtering for Implicit Feedback Datasets, arXiv
3. Maciej Kula, Metadata Embeddings for User and Item Cold-start Recommendations, arXiv
4. Yoon Ho Cho et al., A personalized recommender system based on web usage mining and decision tree induction
5. https://thegood.com/insights/ecommerce-product-recommendation/, as accessed in March, 2019

8.2 About the College

The rise in ‘real-world’ research and ‘learning by doing’ education has generated exciting opportunities with the potential to shift higher education culture and Institute of Computer Technology (ICT) under the Faculty of Engineering & Technology, Ganpat University. ICT’s provision for world class teaching and research is bolstered by an active engagement of industry experts. The Industry-Institute-Integration has resulted in ‘industry collaborated/sponsored courses’ and setting up number of sponsored labs.
In a fast-changing world, technology is changing faster than our lives. Latest technologies are being adopted as an integral part of the world; determining greater employability of young engineers. As a continued effort to offer the Programs-of-Future, Ganpat University provides an exclusive opportunity to the engineering aspirants to join Bachelor of Technology (B.Tech.) and Master of Technology (M.Tech.) – CSE (Computer Science and Engineering) in association with IBM.