iRECOMMENDER FOR E-COMMERCE

17-035

Project Proposal

M.S.D Dharmawardhana

W.W.G.B.P Bandara, K.M.S Bandarage, U.G.D Ugayanga

Bachelor of Science Special (honors) In Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

March 2017

iRECOMMENDER FOR ECOMMERCE

17-035

Project Proposal Report

(Proposal documentation submitted in partial fulfillment of the requirement for the Degree of Bachelor of Science Special (honors) In Information Technology)

M.S.D Dharmawardhana – IT14048906

W.W.G.B.P Bandara – IT14015472

K.M.S Bandarage – IT12010554

A.G.D Udayanga - IT14034350

Ms. Dinuka Wijendra

B.Sc (Special Honours) in Information Technology

Sri Lanka Institute of Information Technology

March - 2017

Declaration

We declare that this is our own work and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
M.S.D Dharmawardhana	IT14048906	
W.W.G.B.P Bandara	IT14015472	
K.M.S Bandarage	IT12010554	
A.G.D Udayanga	IT14034350	

The supervisor/s should certify the proposal report with the following declaration.

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor: Date

Abstract

This proposal document is prepared for the Sri Lanka Institute of Information Technology as a fulfillment of fourth year research project.

This document will discuss various existing issues on recommendation engines of ecommerce websites and the solutions those are researchers proposed to overcome the each discussed issues and the our solution to overcome the each discussed issues with the development of “iRecommender” as an optimized technology for the future trend of e-commerce.

iRecommender will overcome the issues of poor product suggestions on e-commerce websites, increase the efficiency of product suggestions and speed up the finding exact product that is required by specific users.

Table of Contents

Contents

List of Figures

List of Tables

1. Introduction

The following document is composed with the intension of describing a project iRecommender, which is to create an effective and efficient product suggestion system for ecommerce websites. This document will briefly describe the process of proposed continuation of project yet before going in to technical detail. The document will also come up with the facts of existing researches on related fields and the achievements made so far.

1.1. Background & Literature survey

With the evolution of Internet technology people have the trend of using internet services to get their ornaments. When it comes to the marketing, most of the market places are converted as ecommerce platforms where goods and services are offered through the internet. Even though there are number of ecommerce recommendation platforms available, the owners are struggling on the personalized product suggestions for their specific customers.

Since emergence of the ecommerce concept most of the university and non-university researchers have followed number of research methodologies with integration of newest technologies available to find out the best recommendation engine for product suggestions. Their goal was to make the product recommendation platform more efficient and more accurate. Even though the most of researches have been involved, they couldn’t come up with the best solution for an efficient and accurate solution.

1.2. Existing system Types

1.2.1. Content based recommendation Engines.

Content-based filtering methods are based on a description of the item and a profile of the user's preference.

1.2.2. Collaborative Recommendation Engines

collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating)

1.3. Drawbacks of the existing systems

1.3.1. Cold Start Problem

Once a new user is registered, this issue emerges with the recommendation engine. As the Engine, doesn't have sufficient data to predict a product recommendation the new users are not getting personalized recommendations.

1.3.2. Database Delay

One of the most important factor is the quick responsiveness of the database with the executed queries. In most of the recommendation engines, they use relational databases which are very inefficient when considering a growing data set.

1.3.3. No scalability of the system

Scalability is known as

- High availability.
- Redundancy in the face of server failures, both for the data and for the operational service.
- Managing increasing read load.
- Managing increasing data set size.
- Managing increasing write load.

1.3.4. Inefficient data cluster management

To achieve best performance of the system, users should be clustered into different categories to predict the best matching product for the customer.

1.3.5. Usage of only one type of recommendation technology

Almost every existing recommendation engine only uses a single technology out of Content based recommendations and Collaborative Recommendation. In order to achieve the best personalized prediction, the both technologies should be used.

1.4. Evaluation of existing Recommendation Engines.

	Netflix	**eBay**	**Amazon**	**Ali Express**	**Walmart**
Cold Start Problem	✓	✓	✓	✓	✓
Database delay	✓	X	✓	✓	✓
No scalability	X	X	X	X	X
Inefficient cluster management	✓	X	X	✓	✓
Usage of only one type of recommendation technology	✓	✓	✓	✓	✓

Table 1: Evaluation of existing recommendation Engines

1.5. Research Gap

Most of the recommendation engine researches have done their research projects, in order to make the best performing recommendation engines. when considering about previously done researches we could clarify that most of the researchers couldn't be able to achieve their best performing recommendation engine in terms of high recommending accuracy, scalability and availability. there are two recommendation methodology approaches available. Collaborative based recommendation [1] and content based recommendation [2]

Most of the researchers used one recommendation approach to predict the customer tastes [1] [2]. Then we found that the most of the time that test predictions are not feasible and accurately personalized for each specific user, Because of lack of using the recommendation methodologies and technologies. And in here we have found out many of the researchers are trying form past to now, to make a solution for overcome the "Cold Start" [3] problem. researchers suggest deep learning [3], content based recommendation [2], Comparing product specifications [4], Hybrid Recommendation (Collaborative based+ Content based) [4].and the other serious problem we found the response time is very high because most of the recommendation engines use relational databases.

Even though those methods and technologies used to overcome the "Cold Start" problem, the "Cold Start" problem is still there in the ecommerce websites. And we found that the personalized recommending accuracy is not best at most of the e commerce web sites.

So to overcome "Cold Start", "scalability", "availability", "high response time" and "Low personalized recommending accuracy rate "problems We are going to find out a "social media data used, hybrid graph based solution"

1.6. Research Problem

There are several procedures on suggesting products to the users of ecommerce websites. But almost all of them are making their suggestions based on search results of the user. Even if the user has purchased a product, the current suggestion systems are suggesting the same category of the product based on the search results of the user. As this is an inefficient method, users are not getting the exact product suggestions. The failure of product suggestions on an ecommerce website leads to more issues. Users tend to use ecommerce to save their time. As suggestions are failed, users will have to search again and again for products. Sometimes the users have no idea on which keywords to search. By taking in to consideration above issues, the impression that users have on the website are getting reduced. This will lead to decrement in sales on the site.

2. Objectives

2.1. Main Objectives

- Automate the search for each customer according to their tastes.
- Save customers' time by automating the process of search.

2.2. Specific Objectives

Customer Perspective

• Without any trouble, user could buy the products which he needed most.

• Time saving

• Customer doesn’t need to have any special technical knowledge to search.

Business Perspective

• Increase sales of the e-commerce web sites

• Recognize the customer taste when browsing.

• Increase loyalty of Customers.

3. Methodology

3.1. What is iRecommender

Social Networks of specific users are analyzed when they are using online retail stores to buy a specific product. On the point of registering and login to store, “iRecommender” tracks the current users public shared contents (Eg: - “twitter: - text”) and analyze the user tweets to predict the taste of the customer. Our main target is Studying Customer Behavior through the Social Networks.

The users are prompted to link their social networks on the point of registering with the online retail store. “iRecommender” analyses their social network shared contents (Texts). And The determined data are stored and exposed to machine learning techniques and data mining techniques to predict the needs of customer.

“iRecommender” automates the searching process based on the predicted outcome. With the solution provided the time wastage of the user is reduced and the searching process is made more accurate to the customer. And this solution causes to increase the sales of the online store. “iRecommender” solves the struggling issues that were faced by the store owners in suggesting best suitable products for their customers. the specific online stores which are using “iRecommender” will get a higher reputation and higher income among the competitors.

3.1.1. Retrieval phase

This System takes texts(tweets) from twitter using twitter public API. As well as, we hope to extract user’s Geographical location and the age. In this Phase, system filter the key words (product names) and the related feeling/emotion from the twitter tweets. And finally, these set of key words are passed to the next level. (These set of key words are stored inside .csv file.)

3.1.2. Analyzing phase

Analyzing phase, which is described mainly in this document, takes set of key words (Product and Emotions) and determine whether these emotions are Very Negative, Negative, Neutral, Positive or Very Positive. As well as, in this phase, the system stored these output negativity or positivity according to relevant users in a .csv file.

3.1.3. Predicting Phase

In this phase, the System evaluate the trend of the current users to the help of the suggesting several products. Output of this phase is taken by next phase to suggest the accurate products accordingly.

3.1.4. Suggestion Phase

In this phase, the System takes the output of the Predicting Phase and according to their taste suggest the product. As well as, in this phase this System do web mining for validating the key words.

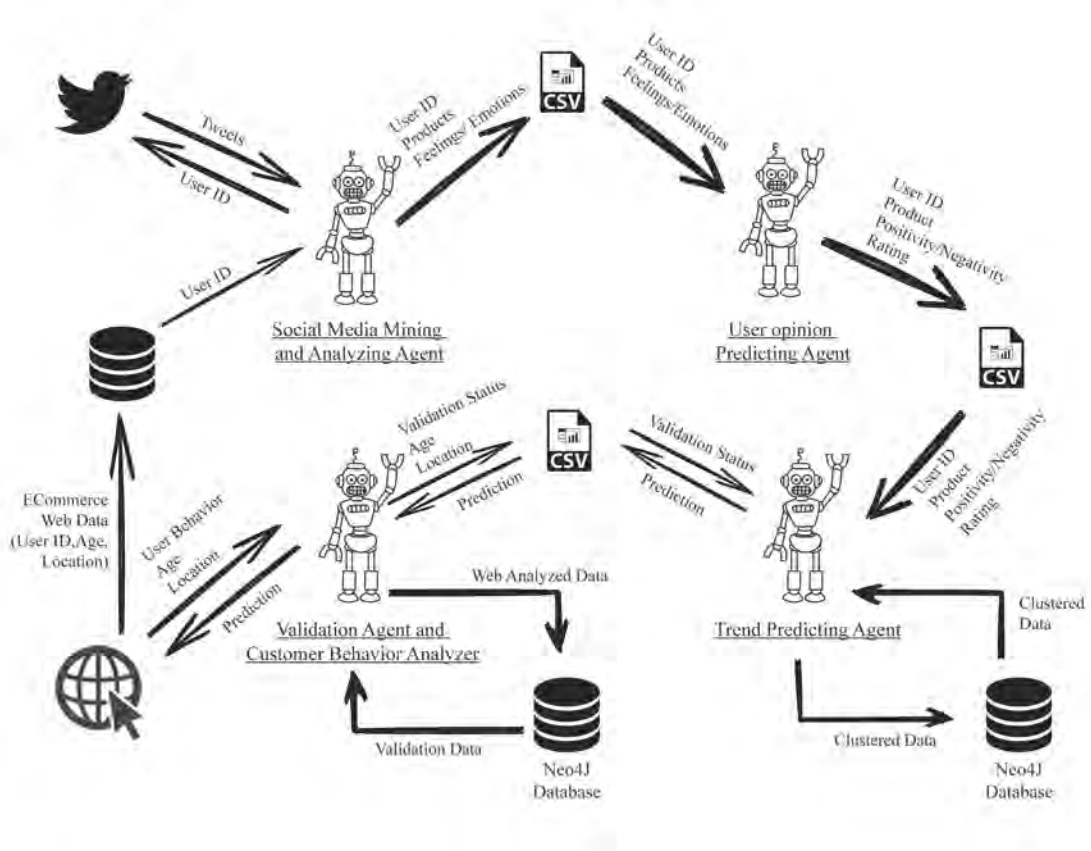


Figure 1.0: High-End Diagram

3.2. Sub part 1 :- Social media mining and analyzing agent development.

In this research our main target is studying customer behavior to predict customer tastes and recommend products to satisfy their needs. For that we need some data to study customer behavior. We mainly focus customer's social media (twitter) to study their behavior by looking text (tweets).

3.2.1. Social media mining and analyzing agent's responsibilities.

i. Analysis each customer's social media
ii. Remove noise words of tweets
iii. Identify the products name if customer have mentioned web site's selling products.
iv. Identify the emotional words which can get customer's opinion about that products.
v. Finally Pass the Data to user opinion predictor agent.

3.2.2. Social Media Analyzer

In the current research, we used Twitter and a system for collecting user generated data (i.e. tweets) for a period of time. Twitter has attracted much interest the last few years from researchers, because Twitter has provided a public API with many resources to developer perspectives to do some research.

Twitter provides the Application Program Interface (API) which allows programmatically accessing and retrieving tweets by a query term. The Twitter API takes a set of parameters related to features such as the language, the format of the results, the published date, the number of tweets and a query and returns the tweets that contain the query and meet the parameters. In our system, only tweets written in English were collected. An application was developed to collect and store the data. The application queried the Twitter. The data were stored in a ". json" format to facilitate our analyze.

3.2.3. Noise word removal

Sentiment classification over Twitter is usually affected by the noisy nature (abbreviations, irregular forms) of tweets data. A popular procedure to reduce the noise of textual data is to remove noise words by using pre-compiled noise words lists or more sophisticated methods for dynamic noise words identification.

By referring research papers, we have found few noise words identification method to twitter data from different dataset and observed how removing noise words affected. using pre-compiled lists of noise words negatively impacts the performance of Twitter sentiment classification approaches. On the other hand, the dynamic generation of noise words lists, by removing those infrequent terms appearing only once in the corpus, appears to be the optimal method to maintaining a high classification performance.

As mentioned above, hope to develop an algorithm to identify and remove noise words from the tweets by considering following aspects.

- By using pre-define noise words list.
- Removing single letters / punctuation marks and other symbols.
- Removing words with low inverse document frequency.

Then after removing noise words the data will store as a word list for the lexical analyze.

3.2.4. Identifying the products name and emotional words

In this function have to identify the products names by looking tweets words which are selling e-commerce site. So, that we have to develop products dictionary by using e-commerce web site's database data. Then After we should identify which are the products customer have mentioned by using lexical analyze.

Other thing is, we should have to identify the emotional words. To do that, we will use algorithm is the emotional dictionary of the "Linguistic Inquiry and Word Count" (LIWC) software (Pennebaker J. and R., 2001). LIWC contains a broad dictionary list combined with emotional categories for each lemma that were assigned by human.

Example of final outcome,

Tweet: I like lather wristwatch → Outcome: Like Wristwatch

3.3. Sub Part 2 :- User Opinion predicting agent development

3.3.1. Who is User Opinion predicting agent?

This is the one of major parts of the iRecommender System which is analyzing the user's feelings/emotions and extract the opinion as Positive, Negative or Neutral. These user feelings/emotions are extracted and stored in a csv file by the Social media mining and analyzing agent. These data will be input to the User Opinion predicting agent. Using Natural Language Processing (NLP) techniques, develop a User Opinion predicting Algorithm. This Algorithm will be output the Negativity, Positivity or Neutrality. As well as, rating of that opinion will also extract. These set of data store inside a csv file according to each user. Each users have different csv file.

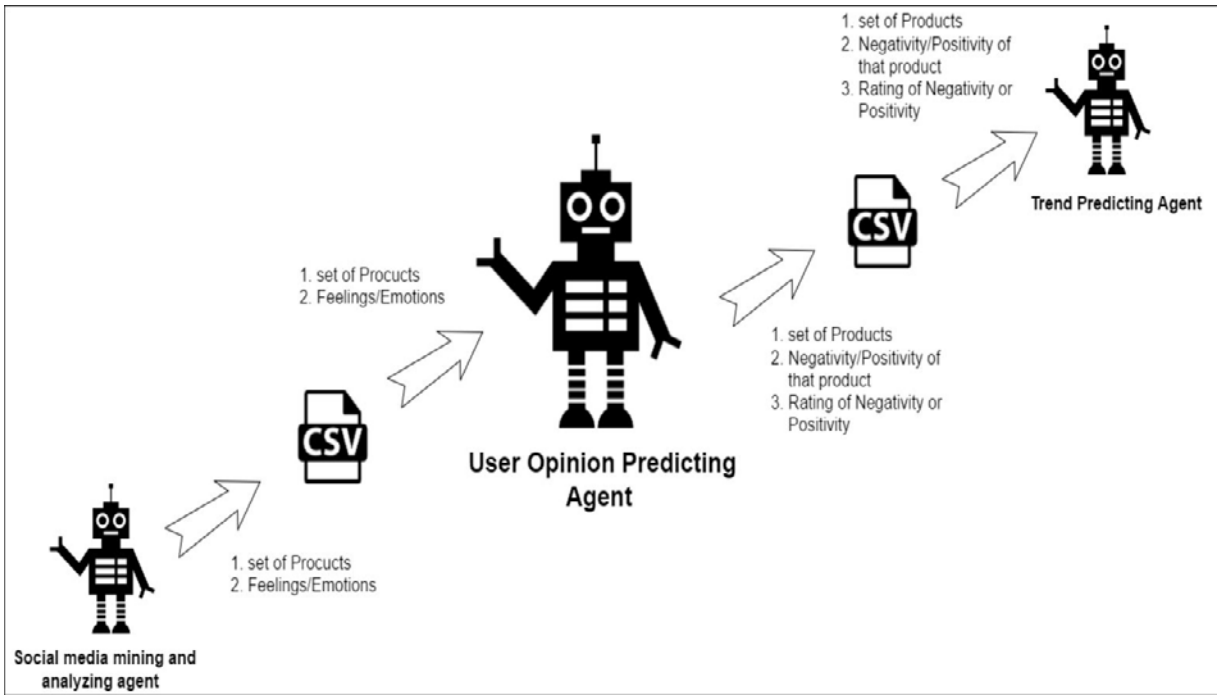


Figure 1: User Opinion predicting agent

3.3.2. What is an Opinion

The simple meaning of the Opinion is "a personal belief or judgment that is not founded on proof or certainty".

But, “the fact that an opinion has been widely held is no evidence whatever that it is not utterly absurd”.

Word of mouth is powerful though.

3.3.3. What is an Opinion to a Machine

It is a "quintuple", an object made up of 5 different things

Oj	-	The thing in question (i.e. product) / Target entity
Fjk	-	A feature/aspect of oj
Soijkl	-	The sentiment value of the opinion of the opinion holder hi on feature fjk of object oj at time tl. soijkl is Positive, Negative or Neutral
Hi	-	Opinion holder
Ti	-	The time when Opinion is expressed

These 5 elements have to be identified by the machine
(defined by Bing Liu in the NLP handbook)

3.3.4. The Main Process

The main Objective of this section goes to extract the Negativity or Positivity of the feelings (Key words) which are taken from the .csv file given by the early 3.2 section. This Main Approach is divided in to 3 major sub-parts.

- Get set of data.
- Execute User Opinion Predict Algorithm for Opinion Mining.
- Store the output result.

3.3.4.1. Get set of data

In the first step of this section get the set of the words from the .csv file which are given from the early stage. These data should read and input to the Algorithm which is create from this section.

3.3.4.2. Execute User Opinion Predictor Algorithm for Opinion Mining.

In this step, System will analyze the extracted Key words (Feelings) and determine the Positivity or Negativity of these Key words. To do that, develop User Opinion Predictor Algorithm.

3.3.4.3. User Opinion Predictor Algorithm

This can be done by using Sentiment Analysis. Basic Sentiment Analysis algorithms use Natural Language Processing (NLP) to classify words as positive, neutral, or negative. Keyword spotting is the simplest technique leveraged by sentiment analysis algorithms.

Keyword spotting is the simplest technique leveraged by sentiment analysis algorithms. Input data is scanned for obviously positive and negative words like 'happy', 'sad', 'terrible', and 'great'. Algorithms vary in the way they score the words to decide whether they indicate overall positive or negative sentiment. Different algorithms have different libraries of words and phrases which they score as positive, negative, and neutral.

After getting extracted data set which are stored in .csv file, we are ready to execute the sentiment analysis algorithm on each Key words (feelings/emotions). Then, we will calculate an average score for all the Key words (feelings/emotions) separately.

Using this User Opinion Predictor Algorithm, we give the Key words as input to this Algorithm. Output of this will be the Negative, Positive or Neutral. Also, these negativity or positivity divide in to 5 levels. Such as Very Negative, Negative, Neutral, Positive and Very Positive. This Algorithm also give the level as 0 – 4. That means,

0 - Very Negative

1 - Negative

2 - Neutral

3 - Positive

4 - Very Positive

3.3.4.4. Store the output result

The output of the User Opinion Predictor Algorithm, should be stored inside a .csv file. This is stored as combination of product and related Opinion on this Product which is analyzed and computed by Algorithm.

3.4. Sub part 3 :- Trend Predicting Agent Development

3.4.1. Who is the Trend predicting agent?

This is the Trend Predicting Component (System) of the iRecommendar system. This component is going to predict the taste of the customer by analyzing Initial, Self-Learned, Web mining and Validation data set.

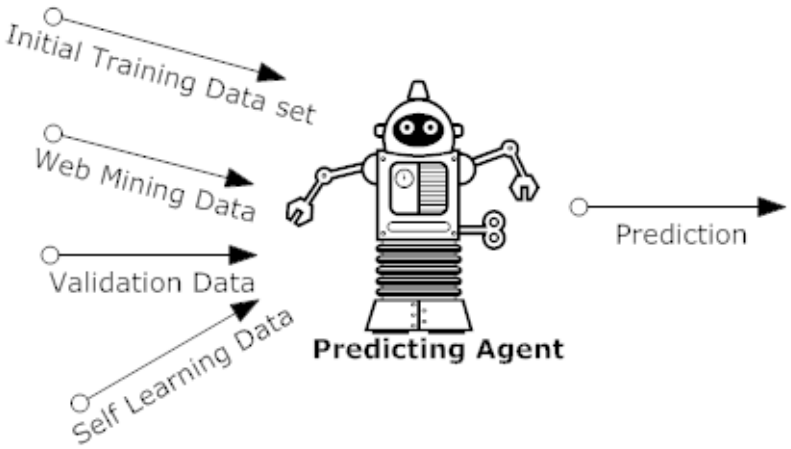


Figure 2: Trend Predicting Agent

3.4.2. How does he works

When the User's social media data analyzed into the positivity and negativity and pass to the Trend predicting agent System, it is going to cluster [5], [6] the data set accordingly similar product wise and Similar customer wise using an optimized similarity functions [7], [8]. And from Ecommerce website, predicting agent can get the Location and the age of the customer and those data also going to cluster in to different age groups and Difference Location Groups. This system is mainly target for the overcome the *Cold Start"* [9] *Problem and the database delay* [10]*" Problem.*

3.4.2.1. To Overcome the" Cold Start" Problem: -A Hybrid filtering system

There are two recommendation engine approaches are there as we mentioned in the above part of the document. The Content based filtering [11] and the collaborative based [7] filtering So in order to get more accurate prediction (Product Suggestion) I am going to develop a hybrid [12] system as stated in Figure 3 by combining above two methodologies.

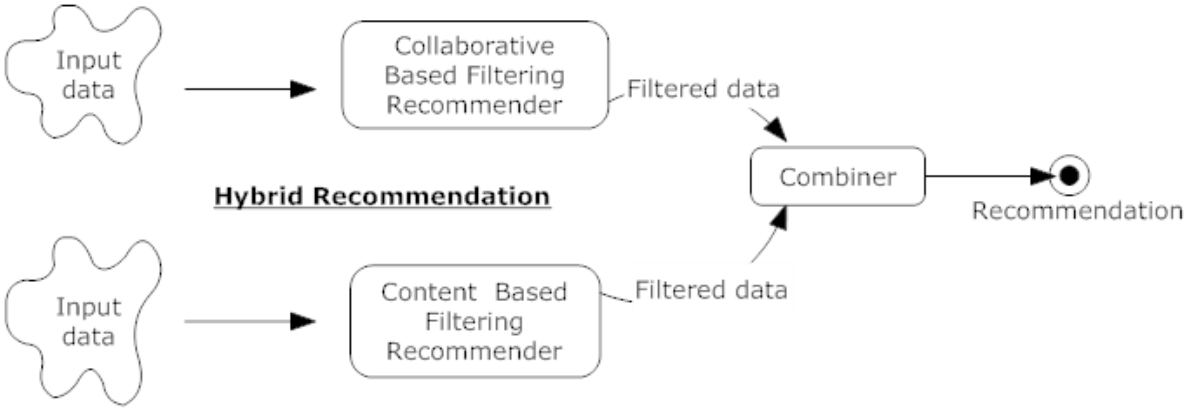


Figure 3: Hybrid Recommendation to overcome Cold Start Problem

3.4.2.1.1. Develop a content based filtering system

Content-based filtering methods are based on a description of the item and a profile of the user's preference. So in here I am going to develop an algorithm(s) for calculate product similarity for opinion positive items(products). and according to the similarity we can suggest the products to the user who interested in the similar product set as stated in Figure 4.

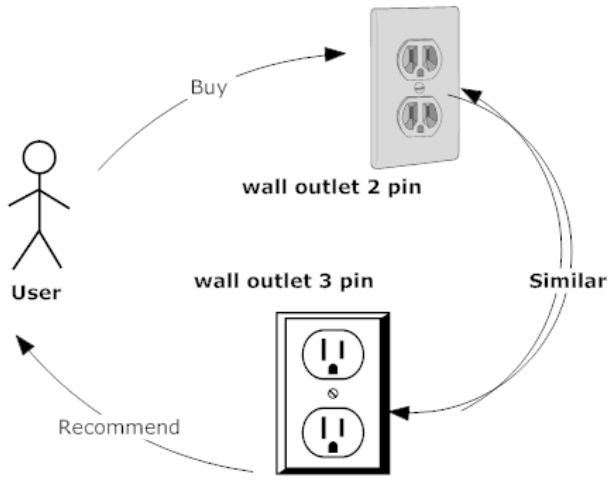


Figure 4: Content Based Filtering System

Then the initial suggestions for a newly registered user will be depend on the content based filtering method. Most of the researchers used users' initial e commerce platform registering data to develop the content based filtering System .and they tried to overcome the "Cold start Problem [*11]*" That works up to some level but that couldn't overcome the hundred percent "Cold start Problem". So in here I am going to use Pre analyzed users' social media data to predict the customer's preference and That mechanism can totally overcome the "Cold start Problem".

3.4.2.1.2. Develop a collaborative based filtering system

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users as stated in Figure 5. So to do the Users Clustering task I am going to develop an algorithm(s) that takes web mining data and Self Learning data (Learn by Learning agent). Web mining dataset means user ratings on products, purchase events, add to shopping cat events, Item View events etc.

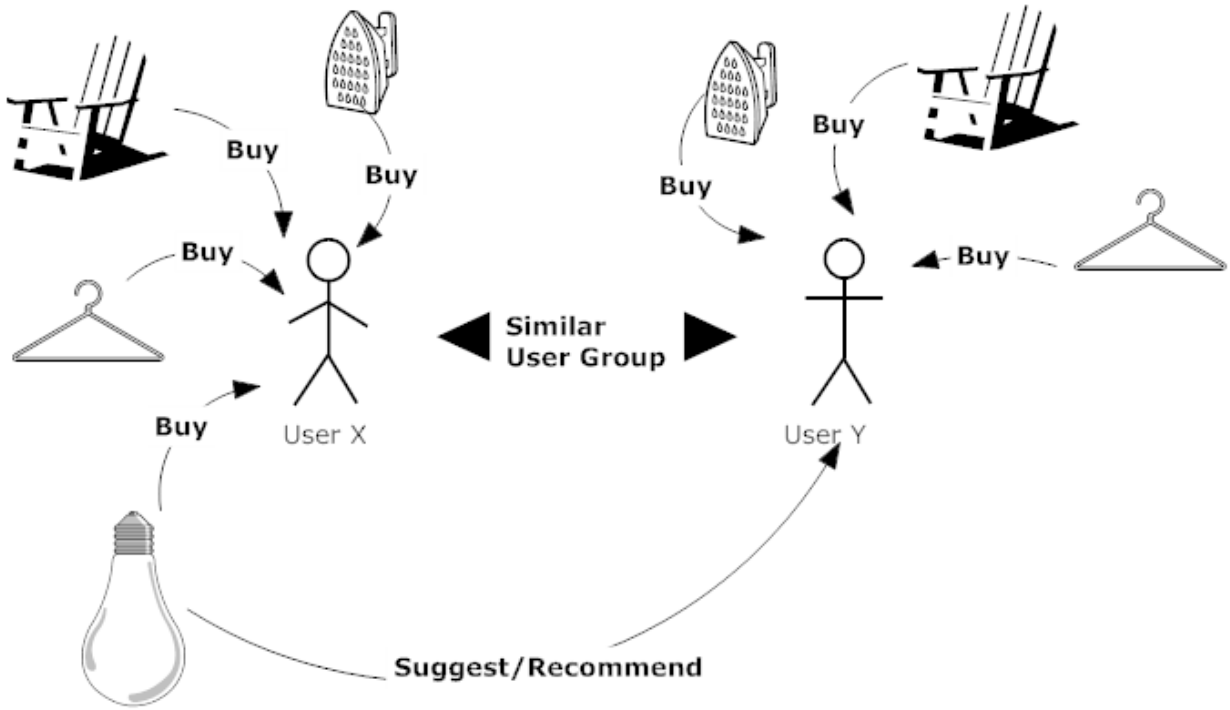


Figure 5: Collaborative based Filtering System

3.4.2.2. To Overcome the "database delay" Problem *[10]*: -Use a graph database

When the data set size is growing up day to day, if we used a relational database that causes to slow down the system. So most of the available e commerce recommendation engines now days trying to migrate to another solution. Most of the researchers have done their researches on graph based database [10] [13] [14]solutions and have found graph has good impact on the performance. So, for our developments we are going to use a graph base database to overcome the database delay problem and system scalability problems.

3.4.3. Initial Step (At the point of new user registration) use case

At the Initial Step, the predicting agent do the prediction as stated in Figure 6, according to the content base filtering algorithms

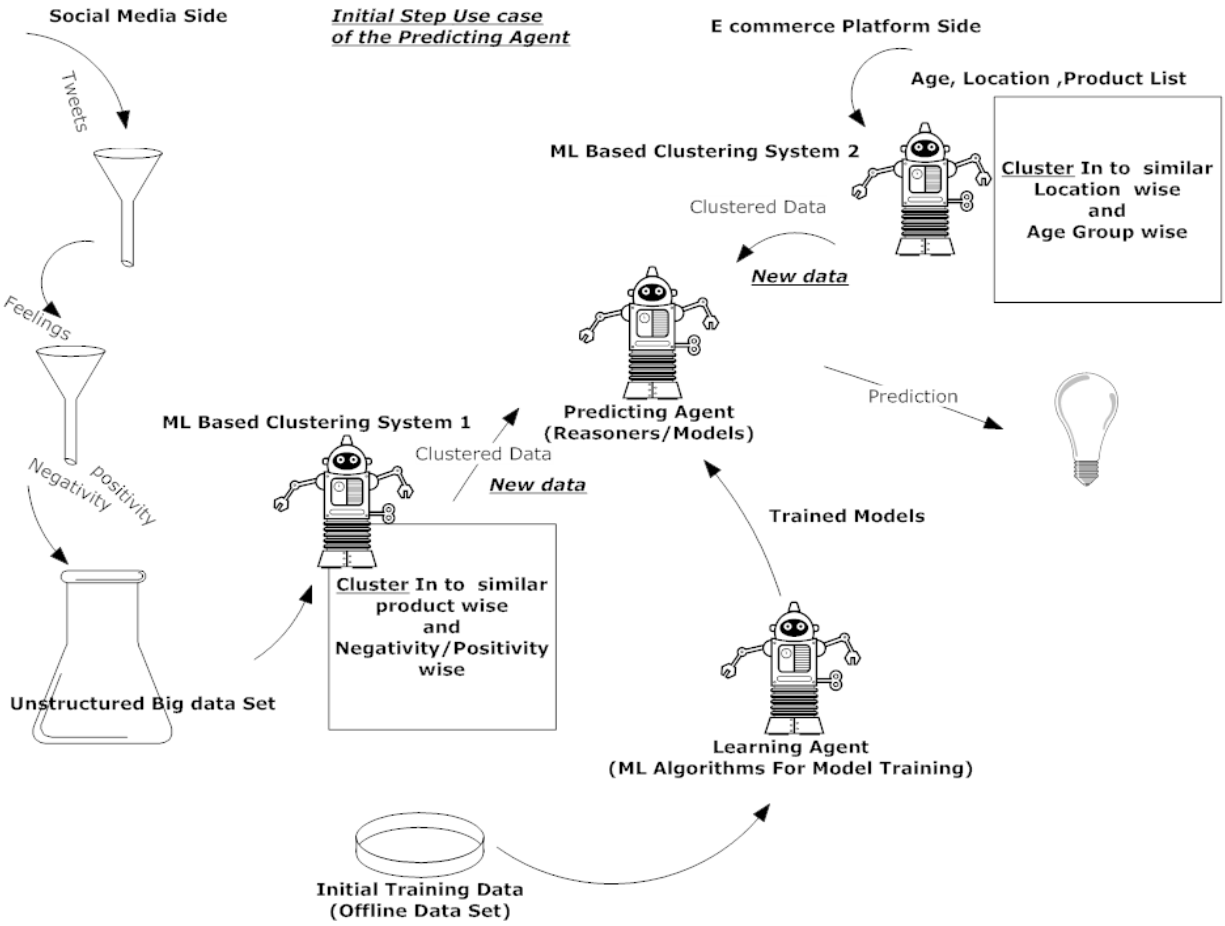


Figure 6: Initial Process of Predicting Agent

3.4.4. Full prediction system functioning step use case

After passing the learning time of the algorithms, it is predicting the suggestions based on web mining data from validation agent, opinion predictor and initial web mining data (Age, Geographical Location). predictions are based on both collaborative and content based filtering algorithms

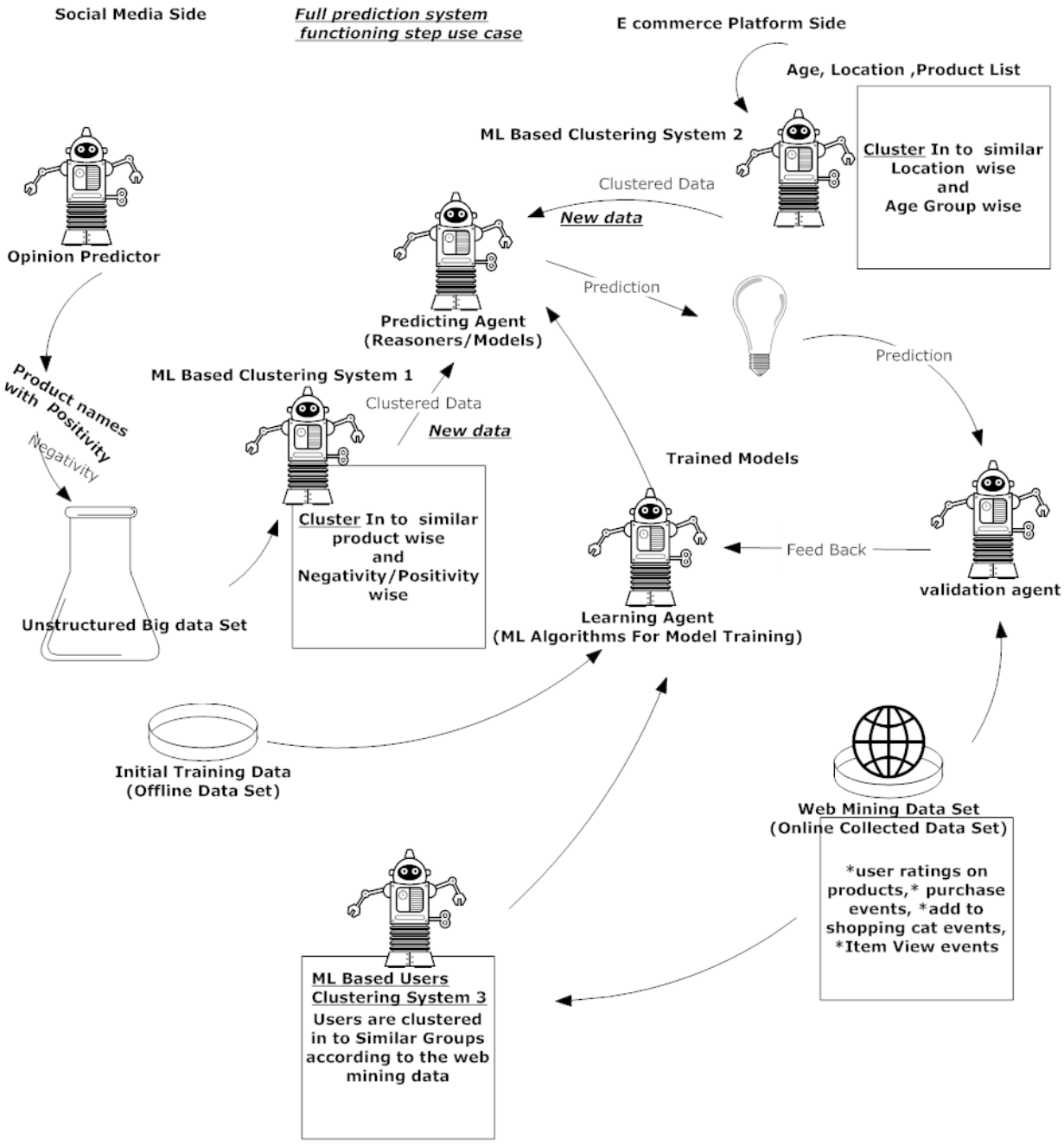


Figure 7: Full functioning of Predicting System

3.4.5. Categorization of Systems develop for specific tasks.

1.3.5.1. For Data filtering

For Collaborative filtering	**For Content Based Filtering**
ML Based Users Clustering System 3	ML Based Clustering System 1
	ML Based Clustering System 2

Table 2: Categorization of Systems for Data Filtering

1.3.5.2. For Prediction

- Predicting Agent (Reasons/Models)
- Learning Agent (ML Algorithms for Model Training)

3.5. Sub part 4:- Validation Agent and Customer behavior analyzer development.

3.5.1. What is customer behavior analyzer?

Customer behavior analyzer is the component which tracks the user activities on the e commerce website. This component is triggered at the point of login of the customer and important data which are required for the iRecommender's validation agent are collected through the actions performed by the user.

3.5.2. Who is Validation Agent

This is the component of the iRecommender system which is responsible for the validation of the prediction done by Predicting agent. The validating agent tracks and analyses the user activities that were collected by the Customer Behavior Analyzer. The analyzed data are exposed to web mining techniques to validate the product suggestion done by the predicting agent.

3.5.3. How does Customer Behavior Analyzer work?

The customer behavior analyzer is responsible on tracking the user activities on the website. Once the user is logged in to the website, the analyzer is triggered and all the activities of the customer are recorded in a log file. The Log files in different web servers maintain different types of information [15]. The basic information present in the log file are

- User name: In some web sites the user identification is made by getting the user profile and allows them to access the web site by using a user name and password. In this kind of access the user is being identified uniquely so that the revisit of the User can also be identified. International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.1, January 2011 100

- Visiting Path: The path taken by the user while visiting the web site. This may be by using the URL directly or by clicking on a link or through a search engine.

- Path Traversed: This identifies the path taken by the user with in the web site using the various links. Time stamp: The time spent by the user in each web page while surfing through the web site. This is identified as the session.

- Page last visited: The page that was visited by the user before he or she leaves the web site.

- Success rate: The success rate of the web site can be determined by the number of downloads made and the number copying activity under gone by the user. If any purchase of things or software made, this would also add up the success rate.

- User Agent: This is nothing but the browser from where the user sends the request to the web server. It's just a string describing the type and version of browser software being used.

- URL: The resource accessed by the user. It may be an HTML page, a CGI program, or a script.

- Request type: The method used for information transfer is noted. The methods like GET, POST.

In addition to the information available on the log files, the customer behavior analyzer optimizes the website to record every click that the user performs on products, purchases done by the customer and the additional items that were visited by the user [16]. These collected data are exposed to Web Mining techniques to extract important data for the validation agent [19].

3.5.4. Overview of Web Mining

Web mining is the process of extracting information from World Wide Web through the conventional practices of the data mining. There are three types of Web Mining approaches as Web Structure mining, Web Content mining and Web Usage mining. On the process of Validation Agent (Web Mining) and Customer behavior analyzer development. We are using the Web Usage Mining approach [17] [15].

3.5.4.1. What is Web Usage Mining?

In the web usage mining process as described on the Figure 8, data mining techniques are applied on the collected data from web analyzer by analyzing the trends and patterns of users performed on the e commerce website [19].

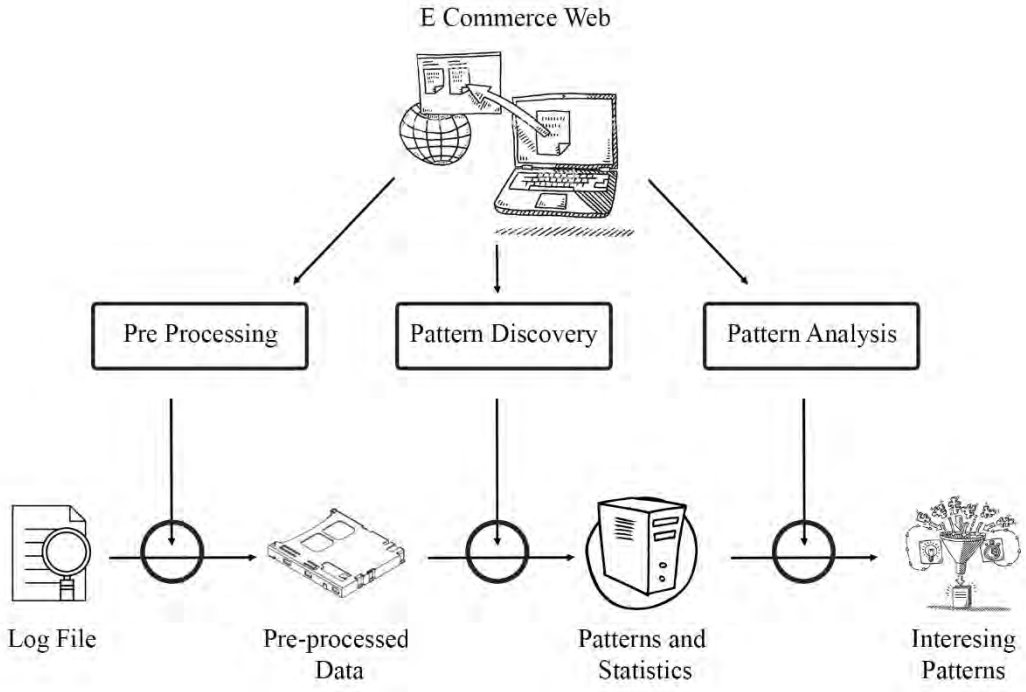


Figure 8: Web Usage Mining

The web mining process undergoes three main process as Pre-Processing, Pattern discovery and Pattern analysis [16].

3.5.4.1.1. Pre-Processing

The data on the log file are not directly analyzable. The Log file contains unwanted data that has to be cleaned. During the preprocessing step. These unwanted data are identified through algorithms and they were removed to minimize the log file obtained.

3.5.4.1.2. Pattern Discovery

The minimized log file that contains formatted data are exposed to the step of pattern discovery. Using data mining techniques these data are analyzed to identify useful information for the validation Agent. The patterns are clustered under details like Session Id and User ID identified from the log file.

3.5.4.1.3. Pattern Analysis

During Pattern Discovery process, both important information that were received from the Predicting agent and the Pattern discovery are used. By introducing suitable algorithms and exposing the collected data, the validating agent validates the suggestions made by iRecommender.

3.5.4.2. How does Validation Agent Work?

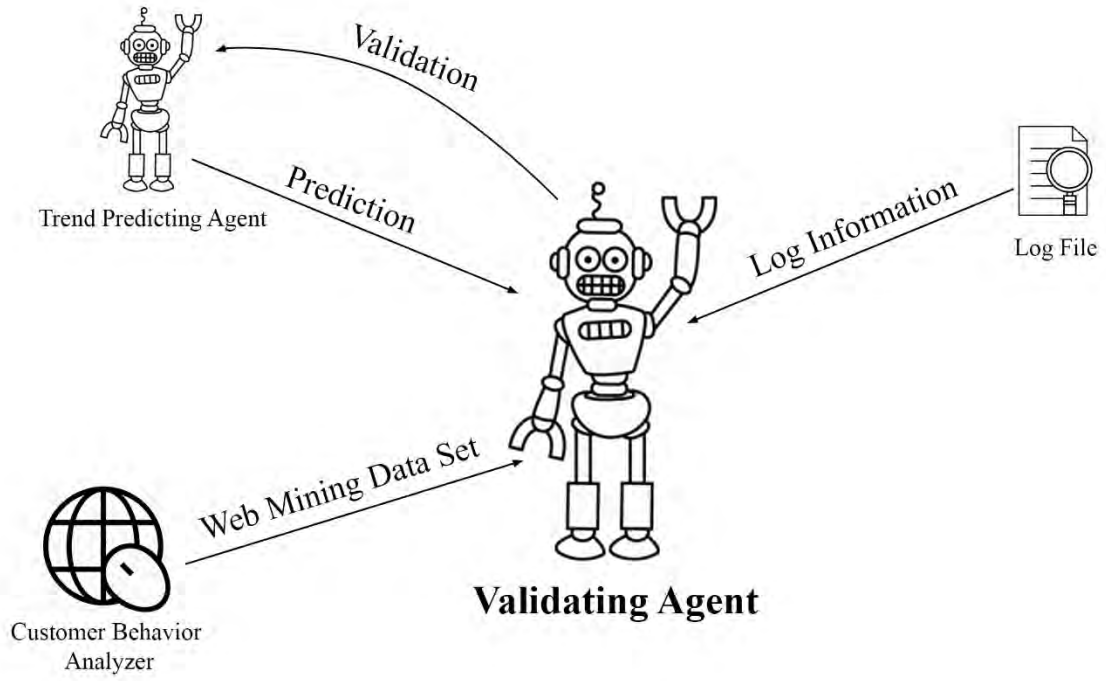


Figure 9: Validation Agent process

As Figure 9 describes, the validation agent which joins with Customer Behavior Analyzer is responsible on the existence of iRecommender. Inputs for validation agent are taken from the Predicting agent and the customer behavior analyzer. For new users, the prediction is done by analyzing the tweets. Once the prediction is done and the customer is actively using the website, the validation agent Talley's the activities of customer with the predictions done by iRecommender. The validation agent keeps its connection with the predicting agent and provide the information that are necessary to generate the future product suggestions [17,19].

4. Description of Personal and Facilities

4.1. Social media mining and analyzing agent development.

I. Tweets of each user are accessed through the Twitter API.
II. Users Geographical location and the age are extracted through the twitter API.
III. Develop an algorithm to filter the product names and emotions (key words) from the tweets using NLP techniques. (Research part)
IV. The filtered key words are passed to the predictor.

4.2. User Opinion Predictor agent development.

I. Develop an ontology to analyze the extracted data from analyzing agent. (Research Part)
II. For each user, Identifying the positivity and negativity feeling about the product using the ontology developed.
III. The analyzed data are passed to the trend predicting agent.

4.3. Trend predicting agent development.

I. Develop an algorithm to predict the product suggestions. (Research Part)
    a. Initial suggestions are based on the details collected by User Opinion Predictor agent and initial web mining data (Age, Geographical Location, Gender)
    b. After passing the learning time of the algorithm, it is predicting the suggestions based on web mining data from validation agent, opinion predictor and initial web mining data (Age, Geographical Location, Gender)
II. According to the geographical area, develop an algorithm to determine the taste of the customer using web mining techniques.

III. Systems develop for specific tasks

- For Data filtering

For Collaborative filtering	**For Content Based Filtering**
ML Based Users Clustering System 3	ML Based Clustering System 1
	ML Based Clustering System 2

- For Prediction

IV.

- Learning Agent (ML Algorithms for Model Training)
- Predicting Agent (Reasons/Models)

4.4. Validation Agent and Customer behavior analyzer development.

I. Track and record the user behavior and the activities on the website from the point of login.

II. Develop an algorithm to validate the predicted suggestions by the trend predicting agent using the analyzed data to give the most accurate product suggestions for each customer.

III. Develop Tracking algorithms and methods for the e-commerce websites in order to collect web mining data.

IV. Passing web mining data (Validity of the suggestions and the Status of the suggested product) to the Trend predicting agent to adopt the product suggestions dynamically for future logins of the user.

5. Budget and Budget Justification (if any)

6. References

[1] R. Royi, E. Yom-Tov and G. Lavee, "Recommendations Meet Web Browsing: Enhancing," *ICDE 2016 Conference,* p. 9, 2016.

[2] P. Lops, M. d. Gemmis and G. Semeraro, "Content-based Recommender Systems: State of," *Springer Science+Business Media, LLC 2011* , p. 33, 2011.

[3] J. Yuan, W. Shalaby, M. Korayem, D. Lin, K. AlJadda and J. Luo, "Solving Cold-Start Problem in Large-scale Recommendation Engines:," *2016 IEEE,* p. 10, 2016.

[4] J. Wei, J. He, K. Chen, Y. Zhou and Z. Tang, "Collaborative Filtering and Deep Learning Based," *DASC-PICom-DataCom-CyberSciTec.2016,* p. 4, 2016.

[5] X. Zang, T. Liu, S. Qiao, W. Gao, J. Wang, X. Sun and B. Zhang, "A New Weighted Similarity Method Based on Neighborhood User Contributions for," *2016 IEEE,* p. 6, 2016.

[6] "Cluster_analysis," en.wikipedia.org, 15 02 2017. [Online]. Available: https://en.wikipedia.org/wiki/Cluster_analysis. [Accessed 18 03 2017].

[7] J. Zhang and Z. Yan, "Item-based Collaborative Filtering with Fuzzy Vector Cosine and Item," *2010 IEEE,* p. 6, 2010.

[8] F. Shen and R. Jiamthapthsin, "Dimension Independent Cosine Similarity for Collaborative Filtering using MapReduce," *2016 IEEE,* p. 5, 2016.

[9] S. S. Singarani, K. Indira and M. K. Devi, "Systematic Approach for Cold Start Issues in Recommendations System," *2016 FIFTH INTERNATIONAL CONFERENCE ON RECENT TRENDS IN INFORMATION TECHNOLOGY,* p. 7, 2016.

[10] "five-signs-to-give-up-relational-database," www.neo4j.com, 27 07 2015. [Online]. Available: https://neo4j.com/blog/five-signs-to-give-up-relational-database/. [Accessed 18 03 2017].

[11] H. Li, F. Cai and Z. Zhifang, "Content-Based Filtering Recommendation Algorithm Using HMM," *2012 Fourth International Conference on Computational and Information Sciences,* p. 3, 2012.

[12] J. Wei, J. He, K. Chen, Y. Zhou and Z. Tang, "Collaborative Filtering and Deep Learning Based," *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence,* p. 4, 2016.

[13] A. Sharma and S. Shalini, "Enhancing the Accuracy of Movie Recommendation System Based on," *2015 Fifth International Conference on Advances in Computing and Communications,* p. 5, 2015.

[14] A. Sharma and S. Batra, "Enhancing the Accuracy of Movie Recommendation System Based on," *2015 Fifth International Conference on Advances in Computing and Communications*, vol. 2, p. 5, 2015.

[15] L. J. Grace, V. Maheswari and D. Nagamalai, "ANALYSIS OF WEB LOGS AND WEB USER IN WEB MINING," *International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.1, January 2011*, vol. 3, p. 12, 2011.

[16] S.-U. Guan, C. S. Ngoo and F. Zhu, "HandyBroker - An Intelligent Product-Brokering Agent for M-Commerce Applications with User Preference Tracking," *Electronic Commerce and Research Applications*, vol. 1, pp. 314-330, 2002.

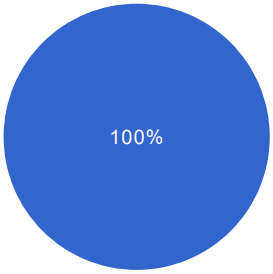
[17] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Analysis of Recommendation Algorithms for E-Commerce," *ACM (Minneapolis, Minnesota.)*, p. 10, 2000.

7. Appendices

7.1. Survey on linking Social Media with e-commerce

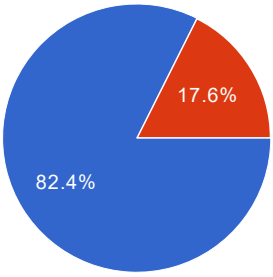
Summary

Are you using social media networks?



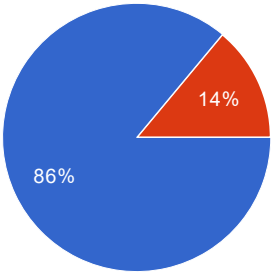
Yes	129	100%
No	0	0%

How often you use social media?



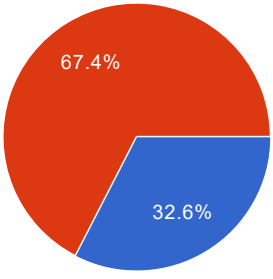
Always	103	80.5%
Sometimes	22	17.2%
Never	0	0%

Are you using e-commerce services from online retail sites?



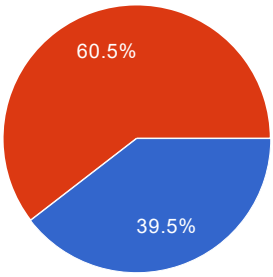
Yes	111	86%
No	18	14%

Does the e-commerce sites suggest products according to your taste at the first login?



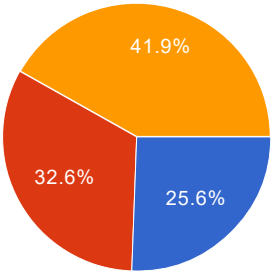
Yes	42	32.6%
No	87	67.4%

Are you getting product suggestions from e-commerce sites accurately to your taste?



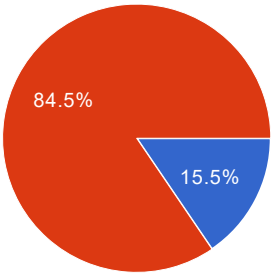
Yes	51	39.5%
No	78	60.5%

Do you always have an idea on what to search and what keywords to use?



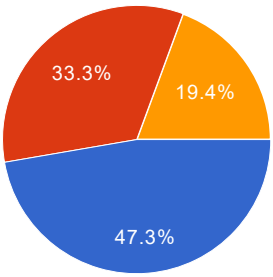
Yes	33	25.6%
No	42	32.6%
Maybe	54	41.9%

Are you getting the exact item you need on the first search?



Yes	20	15.5%
No	109	84.5%

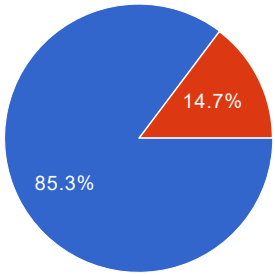
Do you share your taste in social media?



Yes	61	47.3%
No	43	33.3%
Maybe	25	19.4%

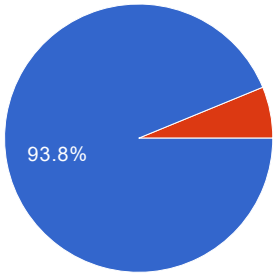
Do you like to access the items that exactly belong to you taste directly at ecommerce websites?

Yes	110	85.3%
No	19	14.7%

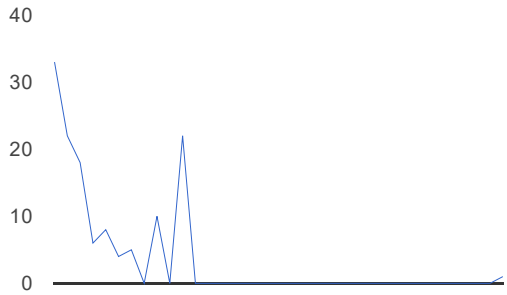


Will it be helpful if the e-commerce sites are suggesting the exact items according to user needs?

Yes	121	93.8%
No	8	6.2%



Number of daily responses



7.2. Gantt Chart

ID	Task Name	Start	Finish	Duration
1	Feasibility survey	2/15/2017	2/28/2017	10d
2	Project Pre-assessments	2/13/2017	2/15/2017	3d
3	Refferencing Research Papers	2/15/2017	2/21/2017	5d
4	Identify Methods and Technologies	2/20/2017	2/22/2017	3d
5	Study on Tools and Technologies	2/23/2017	3/2/2017	6d
6	Develop work breakdown	3/1/2017	3/2/2017	2d
7	Plan roles and Responsibilities	3/3/2017	3/6/2017	2d
8	Project Charter- submission	3/6/2017	3/6/2017	1d
9	Prepare Project Proposal Report	3/8/2017	3/17/2017	8d
10	Proposal Presentation	3/20/2017	3/22/2017	3d
11	Prepare SRS Document	4/14/2017	4/21/2017	6d
12	SRS Presentation	4/24/2017	4/28/2017	5d
13	System Design	5/1/2017	5/19/2017	15d
14	Design DB Architecture	5/4/2017	5/10/2017	5d
15	Design demo e-commerce web site	5/4/2017	5/12/2017	7d
16	Social Media mining and Analyzing Agent	5/9/2017	10/23/2017	120d
17	User Opinion Predictor Agent	5/9/2017	10/23/2017	120d
18	Trend Predicting Agent	5/9/2017	10/23/2017	120d
19	Validation Agent and Customer behavior Analyzer	5/9/2017	10/23/2017	120d
20	Prepare for the presentation	6/26/2017	6/27/2017	2d
21	Progress Presentation I	6/28/2017	6/28/2017	1d
22	Prepare for the presentation	9/4/2017	9/7/2017	4d
23	Progress Presentation II	9/8/2017	9/8/2017	1d
24	Implement Whole System	9/11/2017	9/29/2017	15d
25	System Testing	9/25/2017	10/30/2017	26d
26	Prepare Final Report (Draft)	9/25/2017	10/6/2017	10d
27	Prepare Feedback Submission (Draft)	10/5/2017	10/11/2017	5d
28	Prepare Final Presentation	11/10/2017	11/16/2017	5d
29	Final Presentation and Viva	11/20/2017	11/20/2017	1d
30	Final Report Submission (Soft bound)	11/16/2017	11/16/2017	1d
31	Prepare Research Paper	8/10/2017	8/14/2017	3d
32	Submit the Research Paper	8/18/2017	8/18/2017	1d
33	Website Assessment	10/27/2017	10/27/2017	1d