Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

"Emerging Technologies in the Computer World", January -2017

Sales Predictor: An online Product Sales Prediction Using User Online Behavior To Increase Overall Productivity

Prof. Shilpa Pimpalkar, Harshpriya Gaikwad[1], Shivani Parab[2], Snehali Patil[3], Pradnyashree Punase[4]

Department of Computer Engineering,
All India Shree Shivaji Memorial Society's
Institute of Information Technology,
Kennedy Road, Pune- 411001.

Abstract — *The proposed approach in this project, presents a web based efficient application for creating an efficient user friendly online sales predictor system. In this study, we have presented a model to analyse user history of e-customers and extract information and make predictions about their shopping behavior on a digital market place. We have used a server-side program to collect clickstream data and users history , at the same time another java script program has been used to collect data from client side, It introduces new design for data mining system which combines app usage and content mining. Nowadays, An electronic market place has the advantage of offering more choices, lower prices, easy search and access to online customers. That is why Internet market share is expanding every other day . So customers behavior patterns are also gaining importance in terms of buying or not buying. Application usage and clickstream data may reveal the behavior of users. This is also true for other software usage if it is connected to the internet and a corporate server. So beside customers behavior, a detailed customer profile may also be extracted through analysis .*

Keywords-– *Cloud Computing, Machine Learning, Client/server, Distributed applications, Centralized Database, Android application systems, Security and reliability issues in distributed applications*

I. INTRODUCTION

In a regular market place, when a customer visits a shop, customers tell the shopkeeper the description of required product what does he want, how does he want. They may also tell their budget. The shopkeeper will show them the required products according to given requirements. The shopkeeper may predict that the customer will buy product or not by customers behaviour and facial expressions also. So the shopkeeper will give preference to those customers who really intend to buy the product. Therefore the shopkeeper can provide the service efficiently may give discounts (offers) on the particular category where the customer is interested so that the customer will easily buy the product. But in case of online shopping, there should be a system to predict the behavior of customers. In this case the attributes should differ as facial expressions can not be seen.

Nowadays, in everyday life, smartphones have become an increasingly popular medium for customer to find ideas on products and services. Everyone uses smartphone for purchasing too. So our aim is to develop an android application for smartphone users to provide better service. Our sales offer will be sent only to those customers who are really interested for purchasing product(s).We are designing a sales predictor android app to predict customers behavior collect the customers who are really intending to buy products in order to provide service efficiently to increase the sale.

In this project , we are developing our own android app and taking log files and all databases of users(customers) from the server. First mining is performed on the databases taken for analysis. For analysis, mining algorithm used is Naive Bayes and on the basis of mining result, generate the report of recommendation on the admin side. According to report of recommendation, admin will send the offers to customers for their wish list products or recommended products through server.

The paper is organized as follows, section 2 describes the existing system and its loopholes, in section 3 we discuss every of that aspect responsible for performance deterioration in detail. In section 3 we observe the proposed system and how all the drawbacks are quarantined along with implementation of certain additional features.

II. EXISTING SYSTEM

A. Web Recommender System

A web recommender models user habits and behaviors by constructing a knowledge base using temporal web access patterns as input. Fuzzy logic is applied to represent real-life temporal concepts and requested resources of periodic pattern-based web access activities.

B. Periodic Access Patterns

Some systems mines periodic access patterns, which occur frequently in a particular period, such as every morning between 9 and 10am, directly from semantically enriched web usage logs. With such knowledge, it is possible to deduce only the web resources that the user is most probably interested in during a specific time period, which might change as it is unpredictable.

C. IP Address Oriented

When any user search for any product online via a computer system, ip address is saved by existing web resources of online shopping. When user visits that web site again on the same system, the user will see the offers and advertisements of previous searched products. The user may be different but if he/she using same system for product search or purchasing, that web sources will give offers and pop messages of advertisement related to product based on previous searched data irrelevant of user surfing on that web site of online shopping. These offers are delivered by tracking respective system's ip address, by web resources hence the existing system are ip address based. In case of mobile phones MAC address is tracked in online shopping applications.

D. User Interactivity

User can view the products and add it to cart without registration. When user proceeds to next step to buy product then log in or registration window is displayed after choosing the products. That means user can view products without account. Sometimes users do not log in or sign up to that online shopping application and exit after checking required products. Due to this, web resources are unknown of information about specific user's history of searched item and and user profile.

III. LITERATURE SURVEY

1. Analysis and Prediction of E-Customers' Behavior by Mining Clickstream Data

Authors: Gokhan Silahtaroglu, Hale Donertasli

Data mining is a technology to extract useful information from large data sets. It is an application area of machine learning. Classification, clustering and association rule discovery are the main models of data mining. In order to perform data mining application a special data set is needed. This data set must be built in accordance with the aim of the data mining application. This special data set is called data warehouse or data mart.

Clickstream is a record of a user's activity on the internet, these mouse clicks a user makes when he or she is surfing may tell us a lot about the behavior of the user if it is analyzed in an appropriate way. In a regular retail shop the behavior of customers may yield a lot to the shop assistant. However, when it comes to online shopping it is not possible to see and analyze customer behavior such as facial mimics, products they check or touch etc. In this case, clickstreams or the mouse movements of e-customers may provide some hints about their buying behavior. In this study, we have presented a model to analyze clickstreams of e-customers and extract information and make predictions about their shopping behavior on a digital market place.

2. Page Interest Estimation Based on the User's Browsing Behavior

Authors: Yan Li Bo-qin Feng, Yan LI Feng WANG

A page interest estimation method based on analyzing the users' browsing behaviors is designed in this paper. Different from existing methods, present system avoids users' feedback to the browser and doesn't collect the information which may produce privacy issues, e.g. users' browsing history, bookmarks, and so on. Because the information recorded by server in access logs will be utilized to evaluate the page interest, a reference-based data preprocessing method is carried out to improve the reliability of the access data and extract the necessary information for interest estimation. An interest

estimation model which mainly considers the reference length, the size, the visiting time and the visiting times of each accessed page is then used to determine the page interest. The proposed method is tested by analyzing the practical web access logs for validation. The page interest estimation can be efficiently accomplished and the data set is well prepared for further machine learning algorithms to construct user profiles.

3. Query recommendation based on snippet click model

Authors: Yiqun Liu , Junwei Miao, Min Zhang, Shaoping Ma, Liyun Ru

Query recommendation helps users to describe their information needs more clearly so that search engines can return appropriate answers and meet their needs. State-of-the-art researches prove that the use of users' behavior information helps to improve query recommendation performance. Instead of finding the most similar terms previous users queried, we focus on how to detect users' actual information need based on their search behaviors. The key idea of this paper is that although the clicked documents are not always relevant to users' queries, the snippets which lead them to the click most probably meet their information needs. Based on analysis into large-scale practical search behavior log data, two snippet click behavior models are constructed and corresponding query recommendation algorithms are proposed. Experimental results based on two widely-used commercial search engines' click-through data prove that the proposed algorithms outperform practical recommendation methods of these two search engines. To the best of our knowledge, this is the first time that snippet click models are proposed for query recommendation task.

4. User Behavior Based Enhanced Protocol (UBEP) for Secure Near Field Communication

Authors: Vinay Gautam, Vivek Gautam

The UBEP uses two phase user verification system to authenticate a user. Firstly the acquisition phase is used to acquire and store the user interaction with NFC device and the same information is used in future to detect the authenticity of the user. The Near Field Communication (NFC) is a technology for contactless short-range communication based on the Radio Frequency Identification (RFID). It uses magnetic field induction to enable communication between electronic devices. When user interacts with any NFC enable device then different factor of a user can be used to distinguish between authorized or unauthorized usersfour different factors are cosidered: touch, time, angle and distance to know authenticity of the user. y, Therefore, well known and easy to apply key agreement techniques without authentication can be used to provide a standard secure channel. This resistance against Man-in-the-Middle attacks makes NFC an ideal method for secure pairing of device.

The main contribution of paper is in two-fold:

• To propose a new protocol User Behavior based Enhanced Protocol (UBEP) for secure access of NFC devices.

• Increase the Quality of Service (QOS) to the user in respect to Near Field Communication

5. Web Browsing Behaviors Based Age Detection

Authors: Misha Kakkar, Divya Upadhyay

Users basic attributes like age, gender location etc… plays an essential role in today's web applications. Previous research shows that there is relationship between users' browsing behavior and their basic characteristics. In this paper we made an approach to detect a user's age depending on his web browsing history**.** The browsing history is defined as a record set having 16 columns which contains the probability of different web pages viewed by the corresponding user. This data is collected by conducting a survey among internet user of age. In this the difficulty of guessing the website users' age based on their browsing actions, in which the type of website viewed is treated as a concealed variable to predict the age of different users. The solution consists of information taken from two different categories of users adult and youngsters which is analyzed to predict the user's age based on their profile and browsing behavior and then, a supervised neural network model is trained to predict a webpage user's age i.e. the probability distribution of the ages of a given Webpage's readers. Based on the error analysis, the prediction model resulted from gives a good accuracy.

6. Web Content Recommender System based on Consumer Behavior Modeling

Authors: A. C. M. Fong, Baoyao Zhou, Guan Y. Hong and The Anh Do.

Personalized web content recommendation aims at minimizing ambiguity and unwanted information that is presented to the consumer, thereby reducing the effect of information overload that is often encountered by web surfers. An approach for constructing a user behavior knowledge base, which uses personalization. By performing computer-intensive behavior analysis, modeling and knowledge base construction in advance, future application may include real-time

recommendation on portable web-enabled devices that are becoming increasingly popular among consumers of electronics products, e.g. cellular phones that have limited processing power relative to a personal computer would be well suited to this kind of asymmetric approach. In the future, consumers who choose to use this system on their web enabled mobile devices will be able spend less time and effort searching what they want, but have more of what they are likely to want recommended to them at the appropriate times.

IV. PROPOSED SYSTEM

The proposed system contains three main components, Admin, Server, Android Application to collect the user history and predict the users behavior.

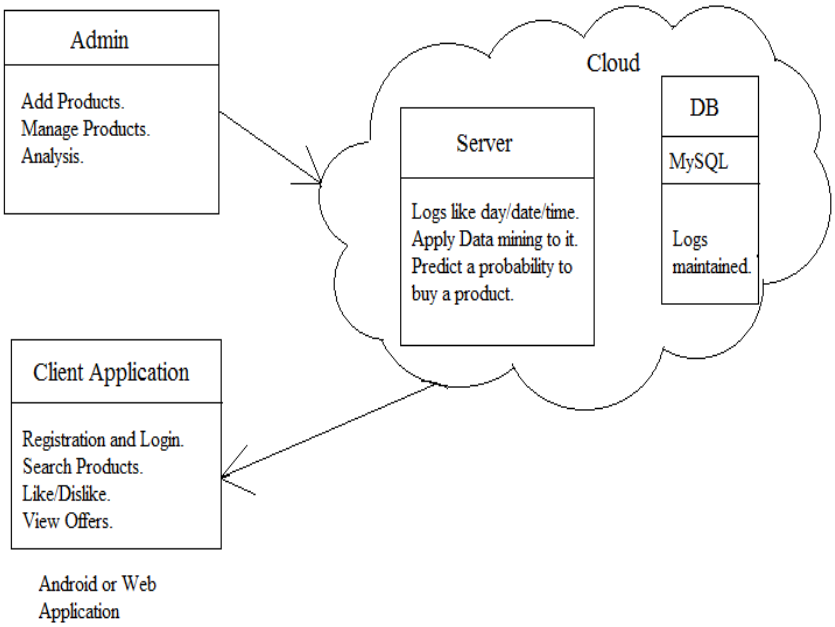

Figure 1. Proposed System Architecture

1. ADMIN

The system admin adds and manages products on the server. It does the analysis of the user history that is stored as a dataset on the server. The analysis of the dataset will be done. Admin will act upon the data after the data mining technique has been applied.

Admin will give special offers to the interested users as per the result of the analyzed data.

2. SERVER

Cloud server stores the database related to the all the users. The database consists of logs of the user history which will be generated each time the user logs in. The dataset includes time duration, click count, search count, ratings, whether the customer adds the item to the cart or not and the buy count.

- Database will be stored using MySQL techniques.
- Data Mining techniques will be applied to the database like Naive Bayes algorithm and Interpolation.
- Admin receive the generated analyzed data after the data mining is performed and accordingly perform actions.

3. USER

The user will register and then login into the application.

- When first time any user uses the system, user will have to register.
- Afterwards, when system starts user needs to log in to the system using username and password provided at the time of registration.

- For the security of the user SHA3 algorithm will be used.

User will view the products, search for any particular products, he/she might add the product to the cart or might buy the products or return without buying. He/she can rate the products, and also comment.

V. METHODS

A. NAIVE BAYES ALGORITHM

What is Naive Bayes ?

- In machine learning, naive bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes theorem with strong (naive) independence assumptions between the features.
- Probabilistic classifiers : A probabilistic classifier is a classifier that is able to predict, given a sample input, a probability distribution over a set of classes, rather than only outputting the most likely class that the sample should belong to. Probabilistic classifiers provide classification with a degree of certainty.

Why Naive bayes?

Naive Bayes or Bayes' Rule is the basis for many machine-learning and data mining methods. The rule (algorithm) is used to create models with predictive capabilities. It provides new ways of exploring and understanding data.

Why preferred naive bayes implementation:

1) When the data is high.
2) When the attributes are independent of each other.
3) When we want efficient output, as compared to other methods output.

Bayes rule:

A conditional probability is the likelihood of some conclusion, *C*, given some evidence/observation, *E*, where a dependence relationship exists between *C* and *E*.

This probability is denoted as P*(C |E)* where:

$$P(C|E)=\frac{P(E|C)P(C)}{P(E)}$$

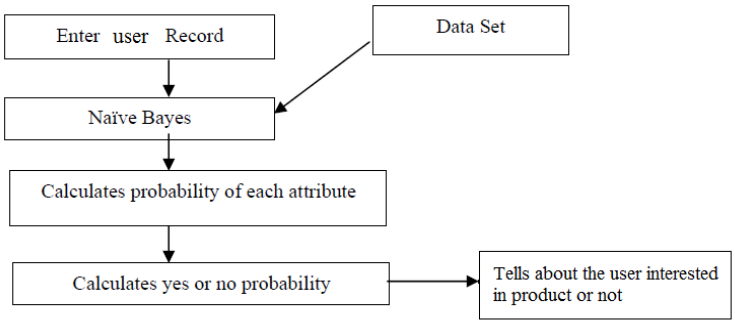

***Figure 2. Implementation of Naïve Bayes algorithm on User* Data**

B. SHA3

The most important aspect of a user account system is how user passwords are protected. User account databases are hacked frequently, so you absolutely must do something to protect your users' passwords.

Hash algorithms are one way functions. They turn any amount of data into a fixed-length "fingerprint" that cannot be reversed. They also have the property that if the input changes by even a tiny bit, the resulting hash is completely different This is great for protecting passwords, because we want to store passwords in a form that protects them even if the password file itself is compromised, but at the same time, we need to be able to verify that a user's password is correct.

The general workflow for account registration and authentication in a hash-based account system is as follows:

1. The user creates an account.

2. Their password is hashed and stored in the database. At no point is the plain-text (unencrypted) password ever written to the hard drive.
3. When the user attempts to login, the hash of the password they entered is checked against the hash of their real password (retrieved from the database).
4. If the hashes match, the user is granted access. If not, the user is told they entered invalid login credentials.

Steps 3 and 4 repeat everytime someone tries to login to their account.
In this project, we are using SHA3(Secure Hashfunction Algorithm 3) cryptographic hash function for users' password authentication.

SHA3 a hash function formerly called *Keccak*, chosen in 2012 after a public competition among non-NSA designers. It supports the same hash lengths as SHA-2, and its internal structure differs significantly from the rest of the SHA family.

C. INTERPOLATION

Interpolation:

- Interpoation is the prediction of values within the data range using the model
- It is the method of constructing new data points within the range of a discrete set of known data points.
- The unknown value at a particular point can be found using many interpolation formula's

Linear interpolation

A method of estimation of the co-orddinates of points lies between two given data points.

Llinear interpolation takes two data points, say (x1,y1) and (x2,y2), and the interpolant is given by:

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{x_2 - x_1}$$

Newton's forward difference formulae :

Let the function **f** is known at **n+1** equally spaced data points $\mathbf{a = x_0 < x_1 < \ldots < = x_n = b}$ in the interval **[a,b]** as $\mathbf{f_0, f_1, \ldots f_n}$. Then the **n** the degree polynomial approximation of **f(x)** can be given as

$$\mathbf{f(x) \ @ \ P_n(x) = \sum_{i=0}^{n} \binom{r}{i} D^i f_0}$$

where $\mathbf{r = (x-x_0) / n \ Þ \ x = x_0 + r\,h \ Þ \ 0 \le r \le n}$

and $\binom{r}{i}$ are the binomial coefficients defined as $\binom{r}{0} = 1$, $\binom{r}{i} = \frac{r(r-1)\ldots(r-i+1)}{i!}$ for any integer $\mathbf{i > 0}$

VI. CONCLUSION

In this study we have introduced a novel approach so as to extract customers shopping behavior from the user history generated at the server after the user views the products. We aim to provide a mobile application which will use the Naive bayes classifier algorithm for predicting the e-customers behavior which will help the shopkeepers to provide special offers to only those customers who are actually going to buy the product.
This way the marketing will become efficient for the companies and even the customer will get satisfactory offers and will be happy to shop.
The interpolation technique used in this approach will also help the shopkeeper to predict the future sales by using the past history data.

VII. REFERENCES

[1] Gokhan Silahtaroglu, and Hale Donertasli "Analysis and Prediction of E-Customers' Behavior by Mining Clickstream Data" IEEE International Conference on Big Data(Bigdata) 2015.

[2] Yan LI Bo-qin Feng, and Yan LI Feng Wang "Page Interest Estimation Based on the User's Browsing Behavior", Second International Conference on Information and Computing Science 2009.

[3] Yiqun Liu *, Junwei Miao, Min Zhang, Shaoping Ma, and Liyun Ru, "How do users describe their information need: Query recommendation based on snippet click model", Expert Systems with Applications 38 (2011) 13847–13856.

[4] Vinay Gautam, and Vivek Gautam, "User Behavior Based Enhanced Protocol (UBEP) for Secure Near Field Communication", International Journal of Computer, Information, Systems and Control Engineering Vol:7 No:11, 2013.

[5] Misha Kakkar, and Divya Upadhyay "Web Browsing Behaviors Based Age Detection", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-1, March 2013.

[6] A. C. M. Fong, Baoyao Zhou, S. C. Hui, Guan Y. Hong, and The Anh Do, "Web Content Recommender System based on Consumer Behavior Modeling" IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, May 2011.

