International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 06 | June-2018 www.irjet.net p-ISSN: 2395-0072

User Profile Based Behavior Identificaton Using Data Mining Technique

Prof. Nilesh B Madke[1], Mr. Yogesh Kulkarni[2], Mr. Rushikesh Mule[3] ,Mr. Akash Lakade [4], Mr. Subodh Kulkarni[5]

[1,2,3,4,5]*Department of Computer Engineering*
[1,2,3,4,5]*Sinhgad Academy of Engineering, Pune, Maharashtra, India*

Abstract - *In regular retail shop the shopkeeper may predict the behaviour of customers using their facial expressions which can results in increase in their sell. However, while considering online shopping it's not feasible to see and make analysis customer behaviour such as facial expressions, products they view or touch etc. In this case, click streams and shopping patterns of E-Customers may provide some hints about their buying behaviour and interest. Here we have presented a model to collect, analyse click streams of E-Commerce Customers and extract information from data and make predictions about their shopping behaviour on an online shopping market. The model we present predicts category of most likely bought and viewed or clicked products on a online shopping market by the customer and according to that it gives recommendations of products to the E-customers. We have also provided offers on items added in basket of most likely bought category by customer through basket analysis. For analysis and prediction, we have used Naive Bayes algorithm. Result of this analysis can be used in Customer Relationship Management and Business Intelligence.*

Key Words: data mining, Naïve Bayes, clickstream, user profiling, online shopping market.

1. INTRODUCTION

One of the basic advantages of using the digital market is that it offers more number of choices at minimum prices and also provide easy access to online customers. Hence, the digital market is expanding daily. As a result of this customers behaviour analysis and prediction are gaining more importance. It is important to study the E-customers behaviour, so we can predict about their behaviour in digital market. The behaviour of customer is the study of when, why, how and where customers buy a product or not. Understanding about what customer actually need is important while building an online e-commerce application.

Data mining is the process of discovering meaningful pattern through huge amount of datasets which are stored in data warehouses or data marts. The key idea behind use of data mining technique is to classify the customer's data with respect to the posterior probability. So in this model the data mining concept performed for the classification on training data set and also use for prediction.

1.1 Advantages of Data mining in E-commerce:

A. Customer profiling:

Basically, customer profiling is customer focused vision in E-commerce site. This strategy motivates online vendors to use business intelligence through the mining of customer's data to make plan for their business operations. It also helps to develop new research on products or services for e-commerce or business aspect. Analysing and classifying the customer's data can help companies to review the sales price. Companies can also use user's history data to find out individual interest. As a result of this companies can plan about their strategy and improve their sales.

B. Personalization of services:

Personalization is the technique of providing services and contents to individuals on the basis of history data available in the database.

C. Basket Analysis:

Market Basket Analysis is analytic and business intelligence tool which helps online vendors to know their customers behaviour in better way. This helps them in their future strategies.

D. Sales Forecasting:

Sales forecasting is the process that involves the aspect of an individual customer spend time on online site to buy product and in this process, it is trying to predict if the customer will buy an item again or not.

E. **Market Segmentation**:

Market segmentation is the best use of data mining technique. The data which has got from various sources, it can be broken down in various and meaningful segments such age, gender, name, phone no., occupation of customers etc. Segmentation of the database of a retail company will also improve the conversion rates. With the help of that company can focus their promotion on a close and highly wanted market.

Finally, in our model we will be dynamically generating the web data of customers and analysis will be performed based on some attributes that are mentioned in the section V. According to that analysis, prediction will be performed.This study introduces a model on prediction of customer behavior using click stream data which is helpful for further business implementations.

2. MOTIVATION

Click streams are the mouse clicks or mouse activities a user makes when they are surfing on internet can tell us a lot about their behaviour if gets analysed in a right way. By analysing users click patterns and their relationship with web contents one can redesign or rebuild a website or e-business along with the behaviour of the online users of it. There are a number of studies going on collecting and analysing, data mining and also click stream data analysis. Online marketing intelligence can be carried out with data mining techniques such as classification, clustering, analysis and prediction etc. Companies can yield a lot by analysing the relation between customers and products on the online shopping market by data mining. Classification models can be used for this purpose in a better way.

So, in the sense of customer behaviour, a detailed customer profile can be created through an analysis on click stream data. So, before starting a click stream analysis, analysts need to build a model with a proper database or data warehouse. The data warehouse will play a major role in data mining model. This study covers important works on data mining technology applied to e-commerce.

3. OBJECTIVES

The objectives of customer behavior are that we can improve our sales if we study the customers. We can change the way to sell our products depending on the ways that customers buy them. Continuous observation of customer behavior can enable you to find out their interest which can in turn help you to recommend products of their interested category to the ultimate satisfaction of e-customers. As the trend in market shifts, a customer analysis will be the first indicator of the same. Whether it is demand forecasting or sales forecasting, both of them are possible and therein lay the importance of customer or consumer buying behavior. The primary objective of this study is to increase sells of e-commerce through customer behavior analysis by finding loyalty or interest of customers in specific category of products and providing recommendations of products of interested category. We are also going to provide offer on product added in cart by e-customer of interested category. This way we are going to provoke the e-customer to buy the products which will result in increase in sells of e-commerce.

4. LITERATURE SURVEY

1.” Analysis and prediction of e-customers’ behaviour by mining clickstream data”, GokhanSilahtaroglu, Hale Donertasli, 2015 IEEE international conference on big data (big data)- In this paper, author have presented a model to analyze clickstreams of e-customers and extract information and make predictions about their shopping behaviour on a digital shopping market. The model predicts whether customers will or will not buy their items added to shopping baskets on a digital shopping market. For the analysis and prediction decision tree and multi-layer neural network data mining models have been used.

2.” Analysis of the internet using behaviour of adolescents by using data mining technique”, ChonnikarnRodmorn, MathurosPanmuang, KhuanwaraPotiwara, 2015 7th international conference on information technology and electrical engineering (ICITEE)-The author has investigated the association rule of upbringing of parent to affect behaviour and experience of internet using of teenagers by the apriori algorithm.

3.” Efficient association rule mining algorithm based on user behaviour analysis for cloud security auditing”, Chunye Zhao, Shanshan Tu, Haoyuchen, Yongfeng Huang, 2016 IEEE-apriori and fp-growth algorithm are used for finding associations between product and customer transaction.

4.” users profiling using clickstream data analysis and classification”, WedyanAlswiti, Ia'farAlqatawna,Bashar A l-Shboul, Hossam Faris,HebaHakh, 2016 cybersecurity and cyberforensics conference-Author proposed a model to extract features based on the sequences of API calls and frequency of appearance and identifies malware by using k-nearest neighbor algorithm.

5. “The analysis and prediction of customer review rating using opinion mining”, WararatSongpan, 2017 IEEE sera 2017, June 7-9,2017, London, UK-This paper proposes the analysis and prediction rating from customer reviews who commented as their opinion using probability's classifier model. This model has used classifier to calculate probability that shows value of trend to give the rating using Naive Bayes techniques.

5. DATASETS USED IN STUDY

Dataset has a server-side program which is used to collect data from client side and store it in organization’s database or company’s database. Data attributes used in the study are given as follows:

Day and date: This attribute represents the day of week days and date on which user visit the site.

Time slot: This attribute represents the time slot of day such as 0 represents morning, 1 for afternoon and 2 for evening/night.

Category id: Products on the site have been categorized into electronics, cloths, shoes, etc. And one unique id is provided for each category such as electronics as 1, cloths as 2, shoes as 3, etc.

Cart count: This attribute includes the number of different products in the cart. The products in the cart may be of same category but they may differ in their size and colour.

Buy count: This show how many products or item of specific category bought by customers.

Click count: When user click on particular item, its entry will be calculated as click count. We are going to count only the click count made on the item or products.

Search count: This variable represents what the customer search on site and the customer can search a certain product on the site by just typing the keywords.

6. METHODOLOGY

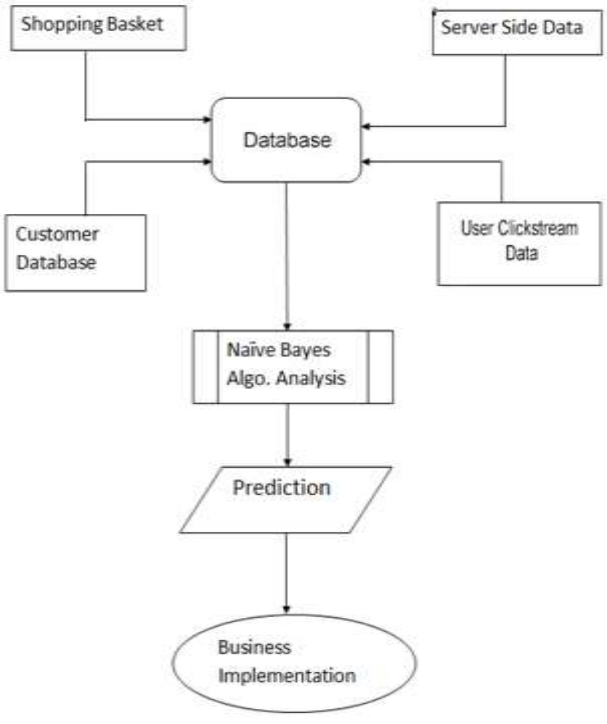

Fig.1. Overall model for applied analysis.

7. PROPOSED SYSTEM

Figure shows that there are three models such as the admin module, the client module and the server module. The customers can open site and perform various functions such as making registration, login, and search products, like/dislike, click and view products. The server module maintains the activity log of the user. The admin module is used to give offers based on the analysis and prediction performs.

Fig. 2. Architectural Overview

7.1 Admin Module:

Admin module consists of two phases. The first phase consists of add and manage products where admin is able to add, delete and manage products. And second part is analysis part where the actual algorithm is being implemented. There is connectivity between the admin module and the database which is used in the system. Based on the prediction made offers will be given to the individual interested customers. Admin at the admin site will be able to execute queries on the database for managing the products such as add, delete, update.

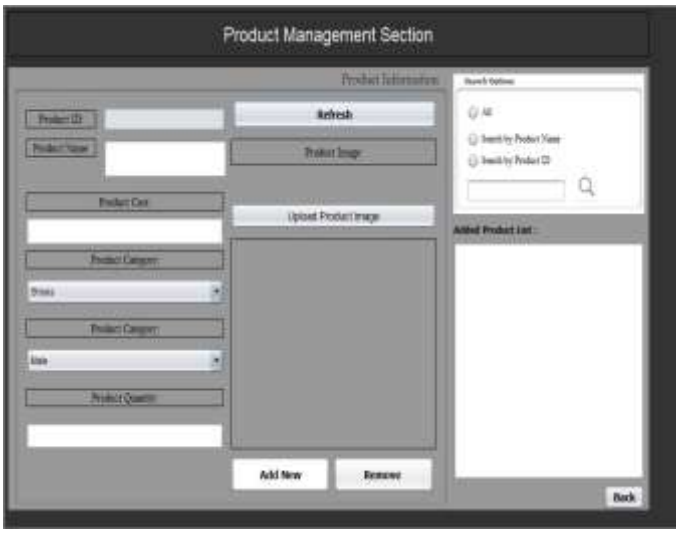

Fig.3. Admin Module

7.2 Client Module:

Client module is used for the customers activities such as registration, searching, viewing, etc. Firstly, Clients will get registered to the application. After that when they get logged in into the application, they will start searching for products. The customers will search the products based on the name, category, like/dislike, rating of a particular product.

Fig.4. Client Module

7.3 Server Module:

The server which we have use is MySQL server. JDBC connectivity is being use for connecting the databases. Server is responsible for user's authentication and the server also provides the services requested by the users. It also maintains the users logs based on the clicks and the activity. Admin will perform analysis on logs to analyse the user behaviour and will predict the individual customer interested category of products.

Fig.5. Server Module

8. ALGORITHM

Data classification process includes two steps. First step consists of learning process where the training data is analyzed by a classification algorithm. Then after that second process is classification where test data is tested against classification algorithm to evaluate the accuracy of the classification algorithm. When learning is completed this model is then using to classify data into different classes. As in our case the e-customers are classified into classes of interested categories such as electronics, clothing and accessories, sports collectibles, beauty products, books etc. For achieving this first we have to make analysis of e-customer's behavior from their history data. Here for analysis and prediction of class labels for e-customers we are going to use Naïve Bayes classifier as specified follow:

A. Bayes Rule:

A conditional probability is the possibility of some conclusion C, given some observation E, where a dependency exists between C and E.

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

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

B. Naive Bayesian Classification Algorithm:

Bayesian classifiers perform statistical classification. They are used to predict class membership probabilities, such as the probability that a given tuple belongs to which particular class. Bayesian classification is basically based on Bayes theorem or rule.

The Naive Bayesian classifier works as follows:

1) Let D be a training set of tuples and their associated class labels. Each tuple is represented by an n-dimensional attribute vector, X = (x_1,x_2,…,x_n) shows n measurements on the tuple from n attributes, respectively, A1, A2..., An.

2)Let's assume that we have m classes C_1, C_2,…, C_m. Given a tuple, X, the classifier will predict class of X which is having the highest posterior probability, conditioned on X. That is, the Naïve Bayesian classifier predicts that tuple X belongs to the class C_i if and only if

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

3) As P(X) is constant for all classes, only P (X|C_i) P (C_i) need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the probability of classes are equally likely, that is, P(C_1)=P(C_2) =…=P(C_m), and we would therefore maximize P(X|C_i). Otherwise, we maximize P(X|C_i)P(C_i). Note that the class prior probabilities

may be estimated by P (C_i) = |C_i, D|/|D|, where |C_i, D| is the number of training tuples of class C_i in D.

4)If the given data sets have many attributes, it may take expensive computation to compute P (X|C_i). In order to reduce computation in evaluating P (X|C_i), the naïve assumption of class conditional independence is made. This assumes at first that the values of the attributes are conditionally independent of one another, given the class label of the tuple. Thus,

$$P(X|C_i)=\prod_{k=1}^{n}(X_k|C_i)$$

$$=P(X_1|C_i)* P(X_2|C_i)*\dots *P(X_m|C_i)$$

We can easily estimate the probabilities P($X_1|C_i$), P($X_2|C_i$),… ,P($X_m|C_i$) from the training data. Here X_k refers to the value of attribute A_k for tuple X. For each attribute, we look at whether the attribute has categorical or continuous-valued. For instance, to compute P (X|C_i), we consider the following:

(a) If A_k is categorical, then P($X_k|C_i$) is the number of tuples of class C_i in D having the value X_k for A_k, divided by |C_i,D|, the number of tuples of class C_i in D

(b) If A_k has continuous value, but the calculation is pretty straight forward. An attribute having continuous-value is typically assumed to have a Gaussian distribution with a mean μ and standard deviation σ, defined b.

$$g(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

So that,

$$P(X_k|C_i)=g(X_k,\mu_{C_i},\sigma_{C_i})$$

We should compute μ_{C_i} andσ_{C_i}, which are the mean and standard deviation, of the values of attribute A_k for training tuples of class C_i. After that we are going to put these in above equation

5)In order to predict the class label of X, P(X|C_i)P(C_i) is evaluated for each class C_i. The classifier predicts that the class label of tuple X is the class C_i if and only if P(X|C_i)P(C_i)>P(X|C_j)P(C_j) for 1 ≤ j ≤ m, j ≠ i

In other words, the predicted class label is the class C_i for which P(X|C_i)P(C_i) is the maximum.

9 RESULT AND ANALYSIS

Naïve Bayes works efficiently or gives better accuracy on average to large datasets. Its performance decreases if we have less size of dataset. As we have maintained click and activity logs of E-Customers it is massive data to be analyze and classify them based on customer behaviour and make prediction by using Naïve Bayes algorithm.

Accordingly, it gives recommendation of products of interested category to individual customers. It gives the accuracy of almost 93% on the classification of customers in different categories of products. According to previous research it shows better accuracy over decision trees and neural networks for classification.

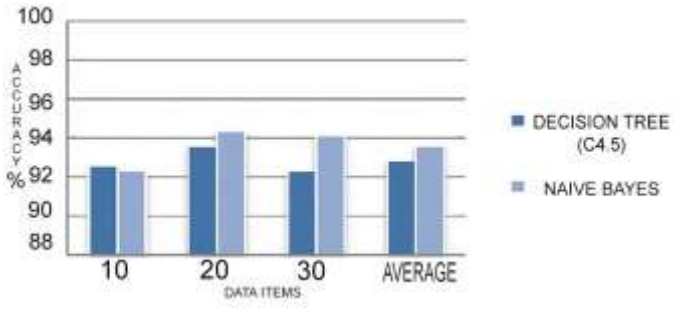

Fig.6. % Accuracy of Naïve Bayes

10 CONCLUSION

In this study, application analyse the massive volume of customer data and classify them based on the customer behaviours and Naive Bayes algorithm helps for considering different attributes which required for analysis. Also, it gives accurate results for large amount of datasets. This kind of customer behavior analysis will directly produce increase in sells of e-commerce application. This application helps shopping effective and easy.

11 FUTURE WORK

The application which we have built can also be implemented on Android platform as well because most of the customers use smart phones for daily purposes. Also, the algorithm for carrying out the same task can be done using a hybrid approach. We can also provide location base offers to customers. If the number of users for such application goes on increasing this also can be implemented on Hadoop platform.

ACKNOWLEDGEMENT

We would like to take this opportunity to thank all the people who were a part of this seminar in numerous ways, people who gave an un-ending support right from the initial stage. In particular, we wish to thank our internal guide prof. N. B. Madke who gave his co-operation timely and precious guidance without which this seminar would not have been a success. We thank him for reviewing the entire seminar with painstaking e orts and more of his unbanning ability to spot the mistakes.

We would like to thank our H.O.D. Prof B. B. Gite for his continuous encouragement, support and guidance at each and every stage of development of this project.

REFERENCES

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

[2] Chonnikarn Rodmorn, Mathuros Panmuang, behavior of Khuanwara Potiwara, "Analysis of the Internet Using behavior of Adolescents by Using Data Mining Technique." in 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thai

[3] Chunye Zhao, Shanshan Tu, Haoyuchen, Yongfeng Huang, "Efficient Association rule mining algorithm based on user behavior analysis for cloud security auditing", in 2016 IEEE.

[4] Wedyan Alswiti, Ja'far Alqatawna,Bashar Al-Shboul, Hossam Faris,Heba Hakh,"Users profiling using clickstream data analysis and classification" in2016 Cybersecurity and Cyberforensics Conference

[5] Wararat Songpan, "The Analysis and Prediction of Customer Review Rating Using Opinion Mining." In 2017 IEEE.

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

[7] Fauzan Burdi,Anif Hanifa Setianingrum,Nashrul Hakiem,"Application of the Naive Bayes Method to a Decision Support System to provide Discounts" in 2016 6th International Conference on Information and Communication Technology.

[8] Huma Parveen,Shikha Pandey," Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm" in 2016 IEEE.