



Plagiarism Checker X Originality Report

Similarity Found: 3%

Date: Thursday, March 31, 2022

Statistics: 73 words Plagiarized / 2426 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

JurnallmKomerInma (Journ ComerScice d foraton 1 5 /(202 2), 33 - 38 . hp: // dx . doi . org / 10 . 21609 / ji . v15i . 1024 33 Pictg Analysiof se ' s tre rom WebLog a ine - Commerce ug Clascat Algoh Sa wadar , Ahmd TrHiya 1 1. Introduction The increasing rise of e-commerce is the primary driver of faster and more convenient commercial transactions. Users always expect the applications to provide whatever they want, quick and convenient e-commerce features for consumers to access, and information about products relevant to consumer interests.

This information is precious for building loyal consumers and can improve the business strategies of e-commerce. Many businesses have understood the value of offering goods and services tailored according to customer needs. It **can be achieved by** defining the interests of the user. The document's substance also determines the user's interest **that the user has** read [1]. Clickstream data analysis can also select the people involved, help businesses recognize customer satisfaction, encourage productivity, and develop marketing campaigns by identifying consumer preferences.

Clickstream data is the essential piece of information for firms looking to tailor their offerings to their clients[1]. It has a lot of value and may provide you with **a lot of information** about the people who visit your website. The document's substance also determines the user's interest **that the user has** read. In other research, user interest may also be established by overall access time, pages on the website that are most often viewed, or the pages that have been accessed the most recently.

Several other specialists looked into user interests and discovered that they are expressed in the posts they visit and the ones they respond to [2]. To develop a more trustworthy model of consumer behaviour, pattern discovery with the proper methodology is also necessary. Several techniques have been employed., such as

KNN[2], Naïve Bayes[3], Decision Tree[4] [6] Large-scale data expansion and the use of network data mining to explore customer preferences and user interests accompanied the e-commerce company's growth.

Several studies have created models to forecast consumer preferences based on user purchase habits, web page visit counts, and web surfing paths. Several user attributes are used to measure market interest[6][7] the order in which pages are discussed, as well as the length of time spent on each page and the frequency with which they are seen, are all factors to. Zheng et al[9] conducted a comprehensive analysis of user browsing time with user interaction.

However, they did not take into account the number of visits or the order in which they occurred. It is different from the research of Yong Li et al[8] that used the time visit and the sequence visit. This study proposes an interest stickiness measurement algorithm and reveals the reliability and accuracy of the algorithm. Kim et al[10] was utilizing the PIE (Page Interest Estimation) to perform a research to determine user interest approach, which measures user interest by using information related to browsing time, frequency of visits, and pages visited.

The findings of this analysis suggest that the PIE approach can be used to assess user interest. The dynamic Mining method is used to measure consumer interest[11]. In several studies, it is reported that the Dynamic Mining method is capable of determining user interest. In a variety of studies, it is reported that the Dynamic Mining approach is capable of determining user interest. In this paper, we will measure user interest by using PIE and Dynamic Mining approaches and compare the two methods that are best used in the case study that we have.

This article uses two techniques to characterize user interest based on numerous indications such as user id, page-time browsing, tool used, and page visited, namely, Dynamic Mining and Page Interest Estimation to predicted potential user and not potential. The potential users are users who have a high-interest value. The high-interest value used in this article for the PIE approach if user interest > 2 and in Dynamic Mining user interest > 600. The information regarding potential users can be used as part of product marketing targets for the company. We have performed comparisons with multiple classification algorithms to identify patterns.

2. Material And Method

2. Page Interest Estimation

Page Interest Estimation is a method used to measure user interest on a website. Manstes et al [3] used this approach to assess user interest on the e-commerce website.

Yan eal [12] devi tqe foriming patestat buids at user ofie thdiurg thuserand n
equirperalinmaon. Then of theferce, i. thimt on th ablgeludine hs l one pa seen e
equenof sis,anthte ween thfir and navit, e ltht o epr usertestdinte bove eoit, e iner ge my
aafolows: $Interest_i = RL_i \cdot C_F \cdot CR_i \cdot VT_i$) Whe $1 = i = m$, m denotes the sum of
pages visited by visitors, and m is the total number of visitors.

Fr (2) RL_i (3) (4) Where RL_i is **the total amount of** time spent on the website by the user
throughout the given access log period, $byte$ denotes the size of the website page,
visits. For HTML files, CF and CR are weight coefficients of 1 and CT coefficients of 1. 1)
Browse Time: The user's choice for alluring locales on which they spend more time is
uncovered by Browse Time. There may, in any case, be a brief switch to another page. In
this manner, the page estimate is influenced by pageview time.

= ????????? (??) ????????? ?????? ?? ????????????????? (??) where is **the total amount of** time
the user spends on page i , **and The average amount of time** a person spends on a page i
is called average. 2) The strategy utilized: Clients who are interested in enrolling with
web destinations should, for the most part, spend **a lot of time** on web pages and use
the HTTP POST method. They will not make advantage of the POST technique if they
have no desire to register on websites. A method POST and GET can be established by:
If (6) then else S. Di and AT.

day , Prectng Analys of r 's re fWeLog a 35 Where is the method used for page . 3) Page
Visited: Page visits represent the average number of pages accessed on a location
during a session . The page's escalation will be much larger if the client is in
terested in or needs to visit the page. The formulas below can be used to compute it
[1] . (7) Where visit (i) represents a collection of all pages that have been viewed (i), and
represents a collection of all pages that have been visited. 4) Interest: Three factors
influence the level of interest : browse time, the strategy utilized, and recurrence.

We create intrigued on each page by using: ????????????????? (??) = ?? . ?? (??) + ?? . (8)
Where is the weight of interest on page i . a , b and c represents the user's interest to ,
and . The total of $?? + ?? + ?? = 1$. Therefore, this research used $a = 0,33$, 33 and $?? =$
 $0,33$. The choice of weight values for parameters a , b , and c is 0.33 because of the
consideration to divide the three parameters BrowTime, MU, and Freq, equally. 2.3
Pattern Recognition 1) Naive Bayes is simple to build and can be used for massive data
sets without the **need for complicated iterative parameter estimation** techniques.

Although it may not be the ideal classifier for every circumstance, it is usually
trustworthy and successful. 2) A Decision Tree may be a more effective calculation for
the purpose of examining the link between free factors and subordinate factors due to

the tree looking pattern. 3) SVM is a data classification method that includes a pre-processing phase. The benefit of this technique is that it properly categorizes the data into two groups while also making a minor generalization error. 3. Result And Discussion The dataset used was a web log file from the komputermurahjogja.com which has moved to ascomputer.co.id e-commerce website. The number of weblog files included in this analysis was 91456 log files, created by 243 user.

The data collected has complete information, consisting of Number of Session Identification, Original Referrer, referrer page, current page, mouse event, users IP, user agent, and browser. Table 1 shows an example of the log file format that was used. Table xx . xx Uer ' ipry 1862 xx . xx Uer ' iclint Molla.0nw N.1 rv:3.0eck001 Firfo3 Uer ' ae Winw Ffo BrosndS 3.1 Data Pre-processing The pre-processing phase of the Dynamic Mining system was carried out in several stages, namely : a. Cleaning of information: Data cleansing removes the web crawler robot. Examples of web crawler robots are Googlebot, Bingbot, or Soguo. b. Identification of the user.

Integration of IP addresses and additional information, such as user agents or referral websites, can identify users and their sessions correctly. 36 Jurnalllu putr Inform(Journalof puteScnceand ati vume 1 5 , i 1, bruary 2 c. Browsing time. It indicates the user's preferences for a certain page of interest on which the user spends more time. d. Method used. If a person is interested in registering on the website, they spend a lot of time on the website and use the HTTP POST method. e. Page visited: pages/visits that indicate the average number of pages visited on the web during one session. In comparison to the Dynamic Mining method, the pre-processing phase of the Page Interest Estimation consists of four stages, namely: a. Data cleaning.

The data cleaning process is the same as the data cleaning process in the Dynamic Mining method b. Page Identification. Identification of what pages visited by users c. Page size: measure the size of each page accessed by the user d. Identify the session id 3.2 Application of Dynamic Mining and Page Interest Estimation Methods Thpr - prnphain ac g gentuser ctit s prt in Taes and 3. Table RL VT F Leo 2. 1. 0. 1 As 0. 0. 0. 1 Saun 0. 0. 0. 1 In Table 2, the formation of user activity is influenced by the browsing time and the number of pages visited by each visitor.

In contrast to the case in Table 3, in the formation of user activity, the influencing factors are the frequency, the time of the visit, which is recapitulated from the first visit to the end in one session, as well as the time taken spent by each visitor on each visit. In Table 4, descriptive statistics of the dataset used. Table 4 . Ds crivStatsf Dtat Method Min Max Avg Std Page Interest Estimation Freq 0.2 1 0.924 0.190 RL 0.002 1.34 2 0.631 0.421 VT 0 1.10 1 0.099 0.221 Interest 0.004 2.84 1.266 0.853 Dynamic Mining BrowTime 0 7890

1366 1556 Number of visited 1 897 33.97 82.02 4 Method used 0 0 0 0 Interest 0 2625
461.67 523.2 1 3.3

Comparison Dynamic Mining and Page Interest Estimation In analyzing pattern discovery, the researcher used the Rapid Miner application to process data with the algorithms that had been described. The cross-validation technique (k-fold = 10) used by the process can evaluate the performance of SVM, Decision Tree, and Naive Bayes. The results of this stage are presented in the table below. All datasets totaling 243 rows were processed using k-fold = 10, and the average accuracy, precision, and recall using k-fold = 10 folds were calculated. The results are presented in tables 5 and 6.

Classifier	Accuracy	Precision	Recall	F1 Score
Naïve Bayes	95.43 %	90.27%	100%	0.94
Decision Tree	100%	100%	100%	1
SVM	96.68%	93.03%	100%	0.94

User Interest Indicator
User id Session id Browsing Time (mins) Method Used (POST/ GET) No of page s visite d
Intere st U1 11161 815 U2 11162 053 11 G 9 U3 11162 191 U4 11162 204 11 U5 11162
248 U10 9 U24 3 S. Di and AT. day , Prectng Analys of r ' s re fWeLog a 37 Table 6 .
AcraReloTeeAlitmUng g InteEmaopc Classifier Accuracy Precision Recall F1 Score Naïve
Bayes 99.64 % 100% 99.38% 0.99 Decision Tree 100% 100% 100% 1 SVM 98.60% 97.77%
100% 0.97 .

Detming usertests focused one iitof e ges sied thuser. aton te frcy of pasis, the of ea pa
vitwaaso e te merin detminusertest Botapprchused ffert amerfor gaer bute decionee
althm inthPatdiy se d e m aavaue.Taes 5 shttof e ee agoris iled,naelnve yes,decion tr an
thve Baetod haower l of aahe othwo.parson, the decision tree approach has a greater
level of accuracy. Unlike in Table 6, the accuracy of the Naive Bayes method is higher
than that of the SVM algorithm. However, suppose you look at tables 4 and 5. In that
case, the decision tree algorithm has the highest accuracy in the Dynamic Mining and
Page Interest Estimation approach, reaching 100% in the determination of user interest.

It can also be seen if the decision tree algorithm has a better performance in
determining user interest in the Dynamic Mining or Page Interest Estimation approach.
Several variables impact the decision tree algorithm's performance, including that a
decision tree requires less data preparation during preprocessing and does not require
data normalization. Creating a decision tree is unaffected by missing values in the data.

The lower level of accuracy in the SVM & Naïve Bayes was influenced by the amount of
data that is not much, as it is known if SVM and Naïve Bayes are an algorithm that can
produce a high level of accuracy if the dataset used is in large quantities while in this
study after going through the data preprocessing process, the data user formed is 243.
4. Conclusion Thrd opmtof - cmce ia chlge npeopl npeopl shd ae o in omtest avari of

ys tt eir panes come th oterAnayzng vi en linwi e - cmce wil hp busine ipreir comii antge.

38 Journallu putr Inform(Journalof puteScnceand ati vume 1 5 , i 1, bruary 2 " Reeah etdogfor nasi f - comcuer avitbaeouineunweugmig, y. [6 S. indai nA. T. iya, " Copaso ClasicaonPforne sd n ym ing f Uer nteNvitiPaerie - Cmerc Webt,

INTERNET SOURCES:

<1% - <https://bbamantra.com/e-commerce-introduction/>

<1% - <https://www.sciencedirect.com/science/article/pii/S1110866515000341>

<1% -

<https://2012books.lardbucket.org/books/marketing-principles-v2.0/s13-gathering-and-using-informatio.html>

<1% -

<https://www.universalclass.com/articles/computers/javascript/user-input-and-output-in-javascript.htm>

<1% -

<https://iwaponline.com/jh/article/17/4/584/3570/Efficient-joint-probability-analysis-of-flood-risk>

<1% - <https://www.thomsonreuters.com/en/artificial-intelligence/publications.html>

<1% - <https://essayhelpp.com/algebra-assignment/>

<1% - <https://www.wired.com/story/your-time-on-facebook-digital-wellness/>

<1% - <https://onstrategyhq.com/resources/strategic-planning-process-basics/>

<1% -

<https://www.ijert.org/research/selection-of-the-best-classifier-from-different-datasets-using-weka-IJERTV2IS3226.pdf>

<1% -

<https://www.certificationanswers.com/en/category/google-digital-garage-exam-answers/>

<1% - <https://journals.sagepub.com/doi/full/10.1177/0312896219877678>

<1% -

<https://quizlet.com/210418095/information-systems-management-c724-set-2-flash-cards/>

<1% -

<https://www.aionlinecourse.com/tutorial/machine-learning/k-fold-cross-validation>

<1% - <https://www.hindawi.com/journals/jhe/2021/1382559/>

1% -

<https://michael-fuchs-python.netlify.app/2019/11/30/introduction-to-decision-trees/>