

ISBN : 978-1-5090-1434-7



THE 3rd INTERNATIONAL CONFERENCE
ON INFORMATION TECHNOLOGY, COMPUTER,
AND ELECTRICAL ENGINEERING

“**GREEN TECHNOLOGY**
STRENGTHENING IN INFORMATION TECHNOLOGY,
ELECTRICAL AND COMPUTER ENGINEERING
IMPLEMENTATION”

PROCEEDINGS

SEMARANG, 19TH - 21ST OCTOBER 2016



**UNIVERSITAS
DIPONEGORO**
BECOMES AN EXCELLENT RESEARCH UNIVERSITY



IEEE

Proceedings

**2016 3rd International Conference on Information Technology,
Computer, and Electrical Engineering (ICITACEE 2016)**

October 19-21st, 2016, Semarang, Indonesia

Editor:

Mochammad Facta
Munawar Agus Riyadi
Eko Didik Widiyanto
M. Arfan

Proceedings

2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE 2016)

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For reprint or republication permission, email to IEEE Copyrights Manager at pubs-permissions@ieee.org. All rights reserved.

Copyright ©2016 by IEEE.

Publisher :

Department of Electrical Engineering
Diponegoro University

ISBN : 978-1-5090-1434-7 (PRINT, Part Number : CFP1689Z-PRT)

ISBN : 978-1-5090-1433-0 (DVD, Part Number : CFP1689Z-DVD)

ISBN : 978-1-5090-0890-2 (XPLORE COMPLIANT, Part Number : CFP1689Z-ART)

Additional copies may be ordered to:
Department of Electrical Engineering
Diponegoro University,
Jl. Prof. H. Soedarto, S.H., Tembalang
Semarang, Indonesia 50275

Pattern Discovery of Indonesian Customers in an Online Shop: A Case of Fashion Online Shop

Rianto
Dept. of Electrical Engineering &
Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
rianto@mti.ugm.ac.id

Lukito Edi Nugroho
Dept. of Electrical Engineering &
Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
lukito@ugm.ac.id

P. Insap Santosa
Dept. of Electrical Engineering &
Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
insap@ugm.ac.id

Abstract—The amount of information displayed on an online shop could distract consumers' focus that might result in customers leaving the website or canceling transaction. Recommendation that provides customers with relevant products is required to make a decision easy for a wide variety of products. Relevancy of products for a customer can be achieved through user patterns discovery. This study aims to develop a user model to provide personalization on e-commerce based on customers group. This study involved 100 participants consisted of 25 girls, 25 adult females, 25 boys, and 25 adult males. Data collection was performed using an e-commerce website i.e. <http://www.erstore.biz> to record users' activity. Data analysis was done using the RapidMiner Studio 5 to provide a decision tree as user model grouped by age and gender. This study found that male and female users among different age groups have different patterns in recognizing process. Female users have a shorter process in the identification of age group than male users.

Keywords—*pattern discovery; Indonesia; Indonesian customer; online shop; fashion*

I. INTRODUCTION

Website is the most popular technology in the growth of informatics. In the past, the objective of a website was to help researchers of CERN (Conseil Européene pour la Recherche Nucléaire) to complete their task. Traditionally, a website is defined as a space to publish documents using a tool called browser to access it [1]. One of the reasons for the popularity of website is the unlimited availability of information and communication [2]. Originally, the technology of website is only load information by text mode. However, the improvement technology of website makes possible to load images, audio, and video. The complete facilities in website technology makes possible to use them in several domains such as news, entertainment, education, government, organization, and business [3].

Lee and Koubek [4] divided a website into four categories i.e. entertainment, information, communication, and business (commerce). YouTube is an example of a website in the entertainment category; BBC News is one of information category; Facebook is of communication category; while amazon.com is a kind of business category. The highest level of the growing website is the business (commerce) category. Commerce or E-commerce reaches its remarkable development because it gives a lot of benefits to users [5]. For vendors, e-

commerce is an extensive marketing medium, where customers could easily find a product [6]. Nonetheless, the diversity of information in website produces another problem for users i.e. information flooding.

The amount of information displayed on an online shop could distract customers' focus that may make customers leaving the website. Therefore, the information displayed should be effective for users when they use a website to get successful transactions. The challenge of information displayed is how to present considerable information on a single page or multi pages using a link [7]. Another issue in the successful transaction in an online shop is a time-consuming process of a complete task of the user [8]. Lesser time consuming browsing is better because it can increase the trust of users when using an online shop. This situation suggested that information provided by an online shop should be relevant to users needs. Therefore, it can help a user to make a decision in an extensive variety of products [9]. To provide relevant information for users required user profile data to generate pattern discovery of users. Consequently, users could be identified when they interact with the website. However, the heterogeneity of users, creates another problem in the process of users identification. This case is a new challenge in the online shop development [10].

The approach to solve the issues that arise in websites such as time consuming, information flooding, poor usability and others is by using website personalization [8]. Personalization is defined as a tool to facilitate users according to their personal needs [11]. Vora and Bojewar [12] defined that personalization is an action that allows users to have a personal experience by user preferences. On the other hand, personalization also suggests providing a specific recommendation accurately [13], increases using intensity of website [14], and helps a user to make a quick decision in a variety of products selected [6, 15]. In the business domain, personalization has been used as a medium of direct marketing based on users need to increase successful transaction [16].

Website personalization can be developed when user's data are available for using as the user model [17]. The process of collecting data in user activity is called user's profiling. Users profiling is defined as user identification process through user interests when they interact with a website. In the future, users profiling will evolve data mining and machine learning [13]. Users profile is almost similar with the personal of users

preferences i.e. interest, hobby, and favorite. Moreover, to make easy process in identification, users can be classified into age, gender, and profession or educational level [10]. User classification can provide a website personalization easier. Meanwhile, website personalization is not only for individual users but also for group users namely group personalization. The differences between personal and group personalization are summarized in Table 1.

TABLE I. THE DIFFERENCES BETWEEN INDIVIDUAL AND GROUP PERSONALIZATION

| Variables | Individual | Group |
|-------------------|---|--|
| Identification | Talent, ability, or preferences | The similarity of age, gender, or profession. |
| Data requirement | Requires many data according to the number of users [18]. | Reducing the amounts of data based on similarity users profile [18]. |
| Data Completeness | Incomplete individual information collected [19]. | Data was collected based on similarity profiles [19]. |
| Target | Individual target [20] | Public target [20] |

Personalization of a website is achieved when a website could identify the users. Accordingly, a website should have knowledge to recognize the users. One of the methods to provide knowledge of a website is by observing user behaviors while they are interacting with the website. An output of the observation is a pattern that will be used as a user model to recognize users. In the present study, user model was developed using a decision tree that was analyzed using RapidMiner Studio 5.

II. METHOD

The current study is a part of a research project on adaptive e-commerce based on group personalization. The project was conducted using a survey method to collect the data. Overall, the steps of the study were 1) identification of products recommended for an online shop; 2) shopping pattern discovery of websites; 3) an adaptive e-commerce website. The aim of step 1 was to identify products recommended for an online shop among Indonesian customers. The study found that one of the recommended products was fashion category. Therefore, the adaptive e-commerce website that will be provided was an online shop in the fashion category. Fashion category was aligning with a study that found fashion was the fastest growing category in an online shop [21]. Step 2 i.e. shopping pattern discovery aimed to provide a user model as pre-knowledge to identify website users. The user model was developed based on visited links when the users interacted with a website. The adaptability of the website will be evaluated in step 3 through a survey method to determine the level of accuracy of the user model in identifying the users.

This paper is the result of the study on step 2 that aimed to identify patterns of user interaction according to gender and age groups. The result was a decision tree through users profiling in www.erstore.biz. Audience overview from Google Analytic indicated that 100 users have been visiting www.erstore.biz in the period of 1 - 4 July 2016. The detail analysis by Google i.e. session: 126, pageviews: 431, pages/session: 3.42, bounce rate: 48.41%, and new session: 87.30%. Data collection involved 100

respondents which were divided into four groups: 25 girls and 25 boys aged 12-23 years, and 25 adult males and 25 adult females aged 26 – 45 years. Age category was selected based on the category recommended by the Ministry of Health of the Republic Indonesia [22] which divides ages into the following ranges (Table 2).

TABLE II. AGE CATEGORY BY MINISTRY OF HEALTH OF THE REPUBLIC INDONESIA

| Number | Category of age | Range |
|--------|-------------------|-------------|
| 1 | Toddlers | 0-5 years |
| 2 | Childhood | 5-11 years |
| 3 | Early adolescence | 12-16 years |
| 4 | Late adolescence | 17-25 years |
| 5 | Early adulthood | 26-35 years |
| 6 | Late adulthood | 36-45 years |
| 7 | Early elderly | 46-55 years |
| 8 | Late elderly | 56-65 years |
| 9 | Senescence | => 65 years |

Demographic data of internet users in Indonesia show that most of the users were in the age of 18 - 35 years old i.e. 82,8% [23]. Therefore, age category of adolescent and adult were chosen. Data was collected from an online shop in fashion category namely ErStore at <http://www.erstore.biz>. Respondents were asked to fill the personal data including name, occupation, and date of birth before using ErStore.

ErStore website had a signifier to identify the users based on the visited link. The signifier was a design of menu to categorize product that appropriates with users who are representing each age groups. Category of "formal" represented adult group, while "casual" represented adolescence group. Further, casual category was divided into casual classic which representing adults, and casual modern which representing adolescences. Method of users profiling was done by recording the visited links in a database expressed in binary values i.e. 0 (not visiting) and 1 (visiting). Links that a user has been visited for more than 2 minutes were given a value of 1, whereas link that a user has not been visited or has been visited for less than 2 minutes were given a value of 0. The minimum number of menus was 3 to get value of 1, otherwise the activity will not be recorded. The algorithm of user profiling process are shown as below:

Step 1: identity of user session (u)

Step 2: while (i <= m)

Step 3: check if (t < 2 minute) value=0 else value=1

Step 4: check if count ((value=1) < 3) record=false

Notes: u= user; i= counter (minimum 3 menus); m= number of menu; t= time spent in the website;

Examples of a dataset as the result of user profiling process is shown in Table 3.

TABLE III. EXAMPLES OF DATASET

| User (U) | Menu(1) | Menu(2) | Menu(3) | Menu(4) | Menu(n) |
|----------|---------|---------|---------|---------|---------|
| User 1 | 1 | 1 | 0 | 1 | 1 |
| User 2 | 1 | 0 | 0 | 1 | 1 |
| User 3 | 1 | 1 | 1 | 0 | 0 |
| User n | 1 | 0 | 0 | 1 | 1 |

Thereafter, the dataset was analyzed by using RapidMiner Studio 5 to provide a decision tree as a user model to classify website users by age group.

III. RESULT AND DISCUSSION

A trial website namely ErStore at <http://www.erstore.biz> was used to determine user profiling. Initially, users were divided into two groups to make easier analysis of data i.e. males and females. The activities of users will be automatically recorded into a database as a dataset that will be analyzed to develop a user model.

A decision tree was used to predict classification or regression. The main function of classification in decision tree is to classify objects based on attribute values. Classification tree has been successfully used to solve some problems such as financial, marketing, software engineering and medicine [24]. Among the reasons of using classification tree were its simplicity and effectiveness in identifying users [25]. Some algorithms used to provide a decision tree were including ID3, C4.5, CART, CHAID, and QUEST. However, some researchers suggested to use C4.5 algorithm for several reasons: 1) removes branches that do not contribute to the accuracy; 2) allow attribute value to be missing; 3) handles continuous attributes by splitting attributes value range into two subsets.

The results indicated that there were differences in shopping patterns between male and female users. Male and female shopping patterns are shown in Fig. 1 and 2, respectively.

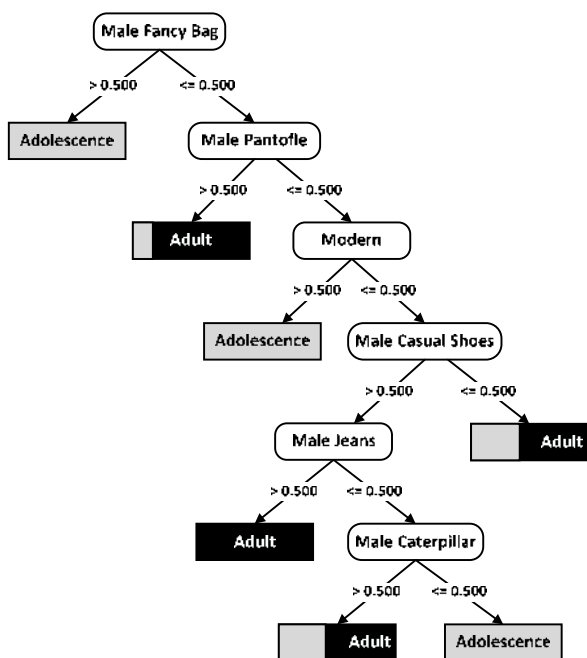


Fig. 1. Male shopping pattern

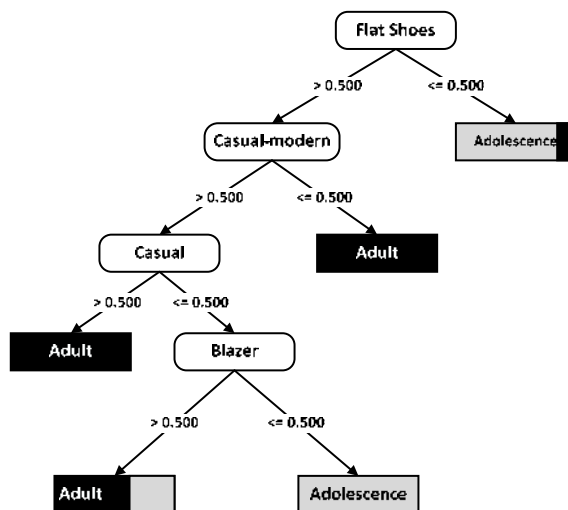


Fig. 2. Female shopping pattern

The decision tree in Fig. 1 explains that the root node of the decision tree is "male fancy bag." A value that is higher than 0.500 indicates that the menu has been visited by users who are of adolescence group, while a value that less than 0.500 is reported that it has not been visited by an adolescence user, but it may be visited by users of other age groups. When a user does not visit "male fancy bag," system will automatically evaluate "male pantofle." A user will be identified as adult when he visits "male pantofle," or will continue into "modern" menu if the user does not visit "male pantofle." In "modern" menu a user will be identified as adolescence when he is visiting this menu. Otherwise, system will continue to evaluate "male casual shoes" to identify the user. When a user does not visit this menu, he will be identified as user of adult group. However, when a user does not visit "male casual shoes" system will proceed to evaluate "male jeans" and "male caterpillar." If a user visits "male jeans" and "male caterpillar," accordingly he will be identified as adult instead adolescence.

Based on female shopping pattern that are displayed in Fig. 2, it is obvious that male and female users have different patterns in fashion shopping. The differences were evident in recognizing process to classify users by age group. Female users have a shorter process than male in recognizing of users by age group. Fig. 2 shows that when a user does not visit "flat shoes" it can be confirmed that she is of adolescence group. A user who has visited "flat shoes" but did not visit "casual modern," she would be identified as adult group. However, when a user visited "casual modern" system would continue to check "casual." When a user visits "casual" she was identified as a user of the adult group. Finally, if a user did not visit "casual" system would check "blazer" and user visited this site will be identified as adult despite of adolescence.

Overall, this study provides rules of website user classification based on gender and age to improve website personalization. The decision tree as the results of this study can be used as a set of branches in a programming language. However, this rule has several limitations to be implemented into the website in general because it is designed specifically for an online fashion store.

IV. CONCLUSION

This present study focused on pattern discovery of users when they are doing online shopping. In the future, these patterns will be useful as a basic knowledge of recognized users in an online shop to provide personalization based on users group. We conclude that:

1. Male and female have a different pattern in recognizing process to be classified based on age group;
2. Female users have a shorter process in the identification of age group than male users.
3. In female patterns, when a user visits "flat shoes" and "casual modern" and "casual" or "blazer," she will be identified from adult group;
4. Adolescence in female pattern is recognized by visiting "flat shoes" and "casual modern" and "casual";
5. Adolescence in male patterns is identified when visits "male fancy bag" or "modern" or "male casual shoes." While, adult group can be recognized when visiting "male pantofle" or "male jeans" or "male caterpillar."

V. FUTURE RESEARCH

Future studies need to focus on how to implement pattern discovery to identify users when using a website and to provide an adaptive mechanism to develop an adaptive website. It is an important study to provide relevant information that suitable with users need particularly in e-commerce.

ACKNOWLEDGMENT

We are thanks to Universitas Teknologi Yogyakarta, Universitas Negeri Yogyakarta, OkeTiket.com, and RuangKerja Software Engineering for their contributions to the present study. Their participation in the present study was appreciated.

REFERENCES

- [1] G. Song and G. Salvendy, "A framework for reuse of user experience in Web browsing," *Behaviour & Information Technology*, vol. 22, pp. 7990-2003, 2003.
- [2] B. Suryavanshi, N. Shiri, and S. Mudur, "Adaptive web usage profiling," in *Advances in Web Mining and Web Usage Analysis*. vol. 4198, O. Nasraoui, O. Zaïane, M. Spiliopoulou, B. Mobasher, B. Masand, and P. Yu, Eds., ed: Springer Berlin Heidelberg, 2006, pp. 119-138.
- [3] M. Butkiewicz, H. V. Madhyastha, and V. Sekar, "Understanding website complexity: measurements, metrics, and implications," presented at the Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, Berlin, Germany, 2011.
- [4] S. Lee and R. J. Koubek, "The effects of usability and web design attributes on user preference for e-commerce web sites," *Computers in Industry* vol. 61, pp. 329-341, 2010.
- [5] Y. Purwati, "Standart Features of e-commerce user interface for the web," *Journal of Arts, Science & Commerce*, vol. 2, pp. 77-87, 2011.
- [6] F. Al-Qaed and A. Sutcliffe, "Adaptive decision support system (ADSS) for B2C e-commerce " in *ICEC 2006*, Fredericton, Canada, 2006.
- [7] P. I. Santosa, "Student Engagement with Online Tutorial: A Perspective on Flow Theory," *International Journal of Emerging Technologies in Learning*, vol. 10, pp. 60-67, 2015.
- [8] M. F. Dhivya Delphina and S. G. Winster, "Personalized content optimization using tensor segmentation," in *Computer Communication and Systems, 2014 International Conference on*, 2014, pp. 136-141.
- [9] Y. Yang, "Web user behavioral profiling for user identification," *Decision Support Systems*, vol. 49, pp. 261-271, 6// 2010.
- [10] I. Garrigós, J. Gomez, and G.-J. Houben, "Specification of personalization in web application design," *Information and Software Technology*, vol. 52, pp. 991-1010, 2010.
- [11] C. Adolphs and A. Winkelmann, "Personalization research in e-commerce- a state of the art review (2000-2008)," *Journal of Electronic Commerce Research*, vol. 11, pp. 326-341, 2010 2010.
- [12] D. Vora and S. Bojewar, "Design of a tool using statistical approach for personalization and usability improvement," presented at the Proceedings of the International Conference & Workshop on Emerging Trends in Technology, Mumbai, Maharashtra, India, 2011.
- [13] S. Kanoje, S. Girase, and D. Mukhopadhyay, "User Profiling Trends, Techniques and Applications," *International Journal of Advance Foundation and Research in Computer*, vol. 1, 2014.
- [14] S. R. Alpert, J. Karat, C.-M. Karat, C. Brodie, and J. G. Vergo, "User Attitudes Regarding a User-Adaptive eCommerce Web Site," *User Modeling and User-Adapted Interaction*, vol. 13, pp. 373-396, 2003/11/01 2003.
- [15] B. Mobasher, R. Cooley, and J. Srivastava, "Creating adaptive web sites through usage-based clustering of URLs," in *Knowledge and Data Engineering Exchange*, Chicago, IL, 1999, pp. 19 - 25.
- [16] S. Zorzo, R. Gotardo, P. M. Cereda, B. L. Kimura, R. Rios, and R. Grande, "An Approach to Treat the User's Preferences About Personal Data," in *Proceedings of the European Computing Conference*. vol. 28, N. Mastorakis, V. Mladenov, and V. T. Kontargyri, Eds., ed: Springer US, 2009, pp. 383-388.
- [17] A. Hawalah and M. Fasli, "Dynamic user profiles for web personalisation," *Expert Systems with Applications*, vol. 42, pp. 2547-2569, 4/1/ 2015.
- [18] J. Teevan, M. R. Morris, and S. Bush, "Discovering and using groups to improve personalized search," presented at the Proceedings of the Second ACM International Conference on Web Search and Data Mining, Barcelona, Spain, 2009.
- [19] H. Hang-bo and Z. Dan-run, "Personalized search based on groupization," *Journal of Henan Polytechnic University(Natural Science)*, vol. 01, 2010.
- [20] K.-T. Leung, D. Lee, and Y. Liu, "Combining Personalization and Groupization to Enhance Web Search," in *Conceptual Modeling*. vol. 8217, W. Ng, V. Storey, and J. Trujillo, Eds., ed: Springer Berlin Heidelberg, 2013, pp. 372-386.
- [21] N. Rajani, K. Rohanimanesh, and E. Oliveira, "Identifying Interestingness in Fashion E-commerce using Pinterest Data," in *6th International Workshop on Social Recommender Systems (SRS 2015)*, Sydney, Australia, 2015.
- [22] D. K. Indonesia, *Profil Kesehatan Indonesia*. Jakarta: Departemen Kesehatan Indonesia, 2009.
- [23] APJII, *Profil Pengguna Internet Indonesia 2014*. Jakarta: APJII, 2014.
- [24] L. Rokach and O. Maimon, *Data Mining With Decision Trees* vol. 81. Singapore: World Scientific, 2014.
- [25] H. Dahan, S. Cohen, L. Rokach, and O. Maimon, *Proactive Data Mining with Decision Trees*: Springer, 2014.