International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Predicting Click-Through Rates using Data Mining Technique on Digital Advertisements

Ivan Diryana Sudirman, Mulyani, Iston Dwija Utama

***Abstract:** In times of increasingly busy use of social media, placing advertisements on social media such as Facebook is an attractive alternative. With the various advantages of advertising on social media, making it very suitable for MSMEs. The use of video as a format for delivering advertising messages is also rife because of the faster internet speed. However, the cost of making ads in the form of videos is relatively more expensive so it needs a lot of consideration when making it be efficient. One thing that advertisers often pay attention to on social media is the click-through rate. This variable becomes one of the measures of the effectiveness of an advertisement. Hence predicting click-through rate is become important nowadays for advertisers, especially to those who have budget constraints. This research tries to predict the click-through rate using data mining techniques. This paper use CRISP method. The dataset was taken from a Facebook advertisement from a small-medium enterprise in Indonesia. Video watches at 25%, 50%, 75%,95% and 100% is use as predictors. The results show that data mining can be used to predict the click-through rate using video watches percentage. Deep learning is the most suitable model for this prediction. The interpretation of the results from data mining is done and managed to find the variables that support the predictions and contradict the predictions.*

***Keywords:** data mining, advertising, click-through rate, social media, deep learning*

I. INTRODUCTION

Social media is currently growing very rapidly and become part of the everyday life of modern society. According to the statistica survey [1], Indonesia is one of the largest social media markets in the world. In the 3rd quarter of 2018, Youtube has the highest penetration with 88% penetration. WhatsApp penetration percentage is 83% and Facebook is 81%. High numbers of users and fast penetration of social media users in Indonesia have made social media an attractive platform for advertising corner of the paper.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author

Ivan Diryana Sudirman*. Entrepreneurship Department BINUS Bandung School of Creative Technology, Jl Pasir Kaliki No.25-27, Bandung, 40181, Indonesia. ivan.sudirman@binus.edu. +6281321531133.

Mulyani, SSN, MM. Entrepreneurship Department BINUS Bandung School of Creative Technology, Jl Pasir Kaliki No.25-27, Bandung, 40181, Indonesia. mulyani@binus.edu

Iston Dwija Utama, SE, MAB. Entrepreneurship Department BINUS Bandung School of Creative Technology, Jl Pasir Kaliki No.25-27, Bandung, 40181, Indonesia. iston.utama@binus.edu

In the case of failure, According another survey from statistica [2], Indonesia is the third largest Facebook user, as many as 130 million residents use facebook. More than 270 million Facebook users in India make it the largest Facebook audience nation. A large number of users is certainly very attractive for companies to place advertisements on Facebook social media.

There are several advantages of using social media like facebook as advertisement media such as a flexible budget and a more targeted audience. Companies can arrange expenses for advertising according to their budget. They can also adjust their target audience with enough precision. Companies can manage the appearance of ads by several variables such as audience region, age, sex and so forth according to their marketing target.

Internet speeds are becoming increasingly faster allowing advertisers to use high-quality images or videos. For Facebook ads, marketers can choose images or videos for advertising purposes that will impact ad quality. As a consideration according to a study, for Indonesian audience video has a higher rate for engagement instead of pictures. [3]. Relatively cheap combine with large audience reach and the opportunity to use video, also other several advantages makes media social is suitable for small-medium enterprise (SME) who want to communicate their message. Social media such as facebook becomes an attractive alternative for advertising media especially nowadays where competition is getting tougher. Besides the advantages that already explain above, another advantage of advertising using Facebook is its ability to record and provide data from advertisements that are served. Lots of important data such as reach, impression, engagement, also click-through rate are available. One of the important data in social media ads is click-through rate.

According to Facebook, click-through rate or CTR is the percentage of times people saw your ad and performed a link click. CTR signifies how many links clicks received on the ad compared to how many impressions the ad received. It is a common measurement used by online advertisers to understand how ads are pulling traffic to websites and other destinations. CTR is calculated as the number of links clicks divided by the number of impressions[4].

The commercial value of advertisement on the Web depends on whether users click on the advertisement[5]. Thus making the CTR is an important variable for the advertiser and predicting the CTR has become important for the marketers. This research tries to build a model that could find general factors and predict the CTR of the ads using video views percentage as the predictor. As explained above, using videos for advertising is more effective but creating videos for advertising on social media is not easy and relatively more expensive.

Therefore advertisers must have some kind of clue so that their video ad can be effective and efficient.

There are three major advertisement billing schemes used by search engines namely pay-per-impression (PPM), pay-per-click (PPC), and pay-per-action (PPA)[6], [7]. When a PPM billing system is used, the advertisers are billed when each ad appears, whether or not the client clicks on the ad. The advertiser is only charged under PPC billing if your announcement or URL is clicked on, and the advertiser pays for PPA billing only when a user action such as registration or purchase happens. Social media advertising also uses the same schemes. In this paper, we analyzed the advertising that uses pay per click schemes. Pay-per-click is a scheme that more suitable for the small-medium enterprise because it is less risky.

As mentioned above, video is more attractive for the audience. The video ad length plays an important role in video advertising and can affect engagement. According to a study, long ads enhance recognition, mid-roll ads lead to better brand name recognition than pre-roll and post-roll ads because of attention spillover[8]. Although the mentioned study was conducted on Youtube video ads and video programs, nevertheless the result shows that the length of the video watched can affect how the audience will react. Facebook measures how long viewers watch the video. It divided into five categories namely video watched at 25%, video watched at 50%, video watch at 75%, video watch at 95% and video watch at 100%. This paper uses those categories as independent variables to predict the CTR using the data mining technique. As far as researchers know, there are no studies that study how CTR is predicted using these variables.

II. THEORY AND METHODOLOGY

In this section, we review several studies and research dedicated to media social advertising and data mining. There are several studies that predict click-through rate namely research done by Gao and Gao [9] using a dataset in which each of the records in the data represents an impression session instead of an impression. An impression session incorporates one or more than one impression with one or more than one click. To aim at each advertisement instead of each impression session, it causes the process more adaptable when developing the model using the data mining technique. Other studies present a Multiple Criteria Linear Programming (MCLP) prediction model as the solution. The experiment uses datasets provided by a leading Internet company in China and can be downloaded from track2 of the KDD Cup 2012 datasets [5].

Richardson, Dominowska, and Ragno [10] argues that for ads that have been showed repeatedly, predicting CTR is empirically considerable, but for new ads, another requirement must be employed. They present that they are able to use features of ads, terms, and advertisers to develop a model that accurately estimates the click-through rate for new ads. They also show that employing the model rises the convergence and performance of an advertising system. As a result, the model improves both revenue and user satisfaction.

Tikno [3] conducted research to measure the performance of media types used on Facebook advertising platforms such as photos and videos. The investigation used three control variables namely gender, age group, and product type as the interest group. The findings of the research showed that Indonesians have a higher engagement rate for ads when the videos are used instead of images. This higher engagement rate gives more advantages for online shop owners/advertisers by decreasing their advertising cost, influencing more audiences, and rising potential conversion into a further transaction.

Gauzente [11] measures the effect of two different categories of message content. The first category of message content is descriptive and provides facts and product characteristic information. The second category is commercial. A second major finding is the moderating function of price perception. Low-price consumers tend to click on commercial advertising messages more quickly whereas high-price consumers obviously prefer concise content. The disparity between descriptive sponsored ads and commercial sponsored ads rises significantly if the customers are highly price-conscious. Interestingly, this gap is less significant in the case of low price- consciousness.

Van den Broeck et al [12] study the intention of respondents to avoid ads placed in the message stream was significantly greater than to avoid ads positioned in the sidebar. Using multiple moderation analyses, they discover that Facebook motivations and product involvement are important moderators of the effect of ad placement on ad avoidance intent. The results point to the important role of the degree of product involvement when directing Facebook ads to the correct audience and choosing the suitable ad placement. They conclude that ad position is a significant factor to consider when advertising on Facebook.

This section also presents the methodologies adopted for this study from the data collection to processing data using rapidminer. A sequence of steps and directions in the data analytics process are provided by the CRISP model [13]. The CRISP model consists of six phases: business and data understanding, data preparation, modeling, evaluation, and deployment.

The business understanding process focuses on the client perspective's interpretation of the project goals and translates this information into the description of data mining problems. It is important for data mining professionals to understand the business for which the knowledge is to be processed and how to find a solution.

The data comprise phase begins with the collection of initial data. The analyst then increases knowledge of the data, identifies data quality issues, finds initial data insights, or detects interesting subsets in order to form hypotheses about hidden information. The data understanding phase consists of four steps including initial data collection, data description, data exploration, and data quality verification.

Modeling, various modeling techniques are selected and applied during this phase and their parameters are calibrated to the optimum values. In general, for the same type of data mining problem, there are several techniques. Certain techniques have specific data form requirements. It may, therefore, be appropriate to return to the data preparation phase. Modeling phases include modeling technique choice, test design development, prototype construction, and project evaluation.

Evaluation, It is important to assess the model more thoroughly and to review the model construction to ensure that the business goals are properly achieved before we begin to implement the model that has been constructed by the data analyst. In particular, the development of a Deployment Model is not the project start. In order to use the acquired knowledge, the customer needs to organize and present them.

III. III. RESULT AND DISCUSSION

In this research, the data was collected from a small-medium enterprise that uses Facebook for their marketing campaign. The SME studied used two formats for delivering the ad namely photo and video but we extract the data from the video format only. Data was breakdown into daily so 196 records collected. CTR, video watched at 25%, video watched at 50%, video watch at 75%, video watch at 95% and video watch at 100% were successfully downloaded.

The day data type was date and the rest of the data was number. In this study day data only use to breakdown the dataset and not used in the data mining process, therefore the day column was deleted. Afterward, the rest of the data was normalized in order to have the same range's value. Only if all data attributes are brought into the same range, their weighting for the analysis's purpose is consistent[14].

There were not many steps that need to be taken in this data preparation stage because the data from facebook is quite representative. The table below represents a few of the datasets used after transform and cleaning.

Table I. Few Row of Dataset Used

CTR (all) Number	Video watches at 25% Number	Video watches at 50% Number	Video watches at 75% Number	Video watches at 95% Number	Video watches at 100% Number
-0.333	-0.993	-0.939	-0.827	-0.770	-0.748
-0.591	3.953	3.629	3.750	3.505	3.485
-0.673	3.128	2.793	2.840	2.593	2.589
-0.583	2.752	2.664	2.874	2.906	3.176
-0.544	2.002	1.863	1.971	2.056	2.240
-0.461	2.106	1.954	2.140	2.098	2.307
-0.688	3.660	3.352	3.444	3.335	3.417
-0.614	2.140	1.875	1.897	1.787	1.897
-0.300	0.551	0.568	0.683	0.798	0.902
-0.771	-0.600	-0.537	-0.320	-0.617	-0.706

For the modeling phase, the auto model function was used. CTR (all) was set as the column predicted. The classification was not set and video watched at 25%, video watched at 50%, video watch at 75%, video watch at 95% and video watch at 100% were selected as the input column. Afterward, the models used in this research were generalized linear model, deep learning, decision tree, random forest, gradient boosted trees, and support vector machine.

The result of data mining using Rapidminer we found that the lowest Root Mean Squared Error model is deep learning as shown in table 2 below with the standard deviation from deep learning is only 0.1.

Table II. The Models Root Mean Squared Error

Model	**Root Mean Squared Error**	**Standard Deviation**	**Total Time**	**Training Time (1,000 Rows)**
Generalized Linear Model	0,7	0,1	6189,0	22666,7
Deep Learning	0,6	0,1	5759,0	20427,4
Decision Tree	0,8	0,2	838,0	42,7
Random Forest	0,6	0,2	5963,0	1666,7
Gradient Boosted Trees	0,6	0,2	63625,0	641,0
Support Vector Machine	0,6	0,2	28217,0	769,2

The deep learning model is better for predicting click-through rate compared to other models such as the Decision tree or Generalized linear model. Other models have the same root mean squared error value but Deep learning has the lowest standard deviation.

Correlation results can be seen in the following table 3. Deep learning has the highest correlation among other models with a 0,061 standard deviation.

Table 3. Models Correlation

Model	Correlation	Standard Deviation
Generalized Linear Model	0,8	0,0
Deep Learning	0,8	0,1
Decision Tree	0,7	0,1
Random Forest	0,8	0,2
Gradient Boosted Trees	0,8	0,1
Support Vector Machine	0,8	0,1

Based on the results above, we can find out that deep learning is the best model for predicting CTR by using video watches percentage as variables. Deep learning has an R Square value of 0.5345336 with MSE = 0.47106266. This shows that the video watches percentage is good enough to explain the CTR variables in the regression model. Many things can influence someone to want to click on an ad, with the R square above 50% we can state that it is good enough.

The result shows that deep learning is the more appropriate model for predicting CTR through video watches percentage. Hence the next step is to study the model and interpreted the result thus new knowledge concerning the object of this research can be found.

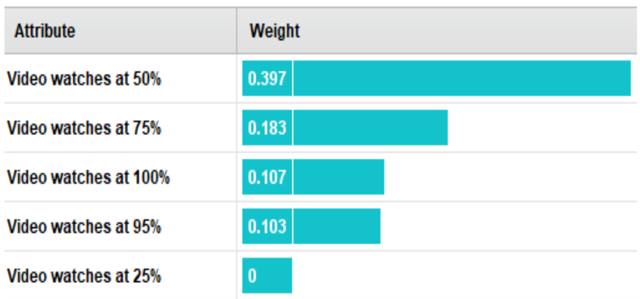

Fig 1. Deep Learning Weights

Weights on Fig 1 above show which columns have in general most influence on the predictions for this specific model. An interesting thing is shown from the results above, video watches at 50% is the most influential variable in predicting the CTR.

Followed by video watches at 75%, video watches at 100% and video watches at 95%. Other interesting discoveries are that the video watches at 25% have 0 weight on predicting the click-through rate. The Simulator function of the Rapidminer result can be seen in Fig 2 below.

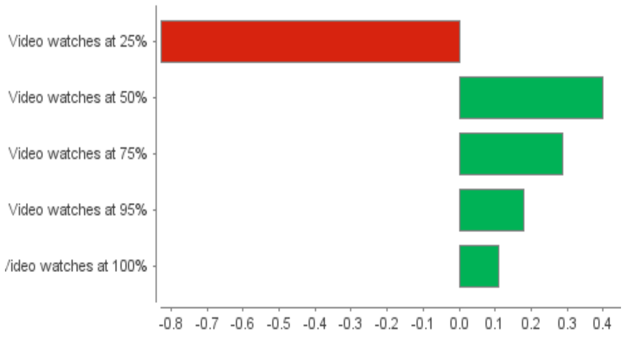

Fig 2. Important Factors for Prediction

The red bar on the above figure shows a variable that contradicts prediction and the green bar shows variables that support prediction. The video watches at 25% is the variable that contradict the prediction.

In the course of this work, we discovered that predicting click-through rate using the data mining technique was possible. A suitable model for this prediction is deep learning with the RMSE value of 0,6 and a standard deviation of 0.1. Correlation using deep learning was the highest among other models which are 0.8.

From the deep learning model, we found that video watches at 50% have the highest weight at 0.397 followed by video watches at 75% at 0.183. Video watches at 100% were at 0.107 higher than video watches at 95% which was 0.103.

IV. CONCLUSION.

This research aim is to predict the click-through rate from the percentage of video watched by the audience who use facebook. From the literature study, how long an ad video is watched can impact viewer decisions. Based on the results of data processing it can be seen that CTR prediction using the percentage of video watched was possible. Deep learning comes out as the most suitable model in predicting CTR.

The results of deep learning show that a video ad on Facebook must be at least 50% of its duration so that the possibility of the ad being clicked is higher. Based on this information, advertisers should really utilize 50% of the initial duration of their ad videos. Advertisers must be able to create attractive ad content in the initial 50% of the duration of their ads.

Other findings from this study are that videos watched under 25% of their duration actually reduce the chances of the ad being clicked on by viewers. Generally, the duration of an ad on social media is between 20 seconds and 60 seconds, therefore advertisers must be able to retain their viewers to be able to watch videos more than 5 seconds for 20 seconds video ad duration or 15 seconds for video ad with 60 seconds length.

In this research the video length was 60 seconds, thus the first 15 seconds is the crucial moment. After the viewer watches the video for 30 seconds (50% of the duration) the chance for viewers to click on these ads is higher.

This discovery can help advertisers, especially SMEs to be able to map what must be made in the initial duration of the video to mid-duration. So that it can increase the chances of the ad being clicked on and achieve the desired goal of advertisers on social media. Social media users have tendency to observe carefully on learning the content message especially for youth [15]

The limitation of this research is that in this study predicting CTR only uses variable video watches. It is also important to study other variables that can predict the value of CTR. Also in this study, the dataset used only came from an SME. The use of bigger datasets is expected to improve the prediction results. Different video and business may impact the prediction, it is recommended to conduct similar research on different businesses

REFERENCE

1. "Indonesia: social media penetration," *Statista*. [Online]. Available: https://www.statista.com/statistics/284437/indonesia-social-network-penetration/. [Accessed: 29-Oct-2019].
2. "Facebook users by country 2019," *Statista*. [Online]. Available: https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/. [Accessed: 29-Oct-2019].
3. Tikno, "Measuring performance of facebook advertising based on media used: a case study on online shops in indonesia," *Procedia Comput. Sci.*, vol. 111, pp. 105–112, 2017.
4. "CTR (link click-through rate)," *Facebook Ads Help Centre*. [Online]. Available: https://en-gb.facebook.com/business/help/877711998984611. [Accessed: 31-Oct-2019].
5. F. Wang, P. Zhang, Y. Shang, and Y. Shi, "The Application of Multiple Criteria Linear Programming in Advertisement Clicking Events Prediction," *Procedia Comput. Sci.*, vol. 18, pp. 1720–1729, 2013.
6. M. Mahdian and K. Tomak, "Pay-Per-Action Model for On-line Advertising," *Int. J. Electron. Commer.*, vol. 13, no. 2, pp. 113–128, Dec. 2008.
7. "Moran & Hunt, Search Engine Marketing, Inc.: Driving Search Traffic to Your Company's Website, 3rd Edition | Pearson." [Online]. Available: https://www.pearson.com/us/higher-education/program/Moran-Search-Engine-Marketing-Inc-Driving-Search-Traffic-to-Your-Company-s-Website-3rd-Edition/PGM50706.html. [Accessed: 31-Oct-2019].
8. H. Li and H.-Y. Lo, "Do You Recognize Its Brand? The Effectiveness of Online In-Stream Video Advertisements," *J. Advert.*, vol. 44, no. 3, pp. 208–218, Jul. 2015.
9. Z. Gao and Q. Gao, "Ad-centric Model Discovery for Prediciting Ads's Click-through Rate," *Procedia Comput. Sci.*, vol. 19, pp. 155–162, 2013.
10. M. Richardson, E. Dominowska, and R. Ragno, "Predicting Clicks: Estimating the Click-through Rate for New Ads," in *Proceedings of the 16th International Conference on World Wide Web*, New York, NY, USA, 2007, pp. 521–530.
11. C. Gauzente and Y. Roy, "Message content in keyword campaigns, click behavior, and price-consciousness: A study of millennial consumers," *J. Retail. Consum. Serv.*, vol. 19, no. 1, pp. 78–87, Jan. 2012.
12. E. Van den Broeck, K. Poels, and M. Walrave, "An experimental study on the effect of ad placement, product involvement and motives on Facebook ad avoidance," *Telemat. Inform.*, vol. 35, no. 2, pp. 470–479, May 2018.
13. Shearer, Colin, "The CRISP-DM Model: The New Blueprint for Data Mining," *J. Data Warehous.*, vol. 5, no. 4, pp. 13–22, 2000.
14. R.-C. Härting and A. Sprengel, "Cost-benefit considerations for Data Analytics - An SME-Oriented Framework enhanced by a Management Perspective and the Process of Idea Generation," *Procedia Comput. Sci.*, vol. 159, pp. 1537–1546, 2019.
15. Aryanto, R and Mulyani, "Website User Interface of Marine Bio-Tourism Destination for Today Digital Environment," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 12, pp. 2064–2067, Oct. 2019.

AUTHORS PROFILE

Ivan Diryana, graduated from doctoral of business management, he has an experience in industry for 2 years then focusing in education field as a lecturer. He also have a business in culinary area for more than eight years. Interest area of research and lecturing in information system, entrepreneurship, marketing, and business.

Mulyani, graduated from Doctoral of economic program. Mulyani have a passion in research about art and humanity, management and leadership. She is a lecturer of Entrepreneurship program of BINUS University. Her concern about teaching and research made her become part of lecturer association namely Ikatan Dosen Republik Indonesia.

Iston Dwija Utama, graduated from master in business management program, passionate in entrepreneurship, marketing, and management field. Experienced in industry for more than 8 years and having own business in culinary. Beside teaching, he is also active in professional organization that concern in SMEs namely LUNAS (Layanan UMKM Naik Kelas) as a business mentor