ANALISA EFEKTIVITAS KAMPANYE IKLAN DIGITAL TERHADAP MINAT PASAR PADA PROGRAM PELATIHAN RENANG CLUB BIATHLON DENGAN METODE BINARY LOGISTIC REGRESSION

Ratih Permanasari[1], Rachmat Adi Chandra[2], Agung Fitriansyah[3], Jerry Heikal[4]
libiaquire@gmail.com[1], ra.chandra87@gmail.com[2], agung.fitriasyah2984@gmail.com[3], jerry.heikal@bakrie.ac.id[4]
Universitas Bakrie Jakarta

ABSTRACT

The Biathlon Club is a part of the swimming, swimming and physical training, middle distance swimming and running, swimming and physical training, middle distance swimming and running, athletics, physical training, Swimming, running intermediate and physical training which are all the basis of biathle training (middle distance swimming and running). Because the number of club members is still below 150 and has not yet reached the expected income target, the management will market the Swimming Club under the Biathlon Club to increase the number of participants and make it more widely known. There are many methods used for marketing, one of which is the digital advertising campaign method, but before it is carried out within a certain target time and periodically, its effectiveness must first be analyzed in attracting customer interest. The method taken is to advertise one short videos with a maximum duration of 53.85 seconds regarding the swimming training offered. Digital advertising is carried out on two social media, namely Tiktok and Instagram. Achieving a high target market reach through "clicks on advertisements" is a measure of market interest in the advertisements displayed.value propositionresulting from the effective use of digital advertising campaigns in marketing the services offered by the club. The theme of this research is: Evaluating Advertising Campaign Performance: Binary Logistic Regression Framework - Assessing the effectiveness of digital advertising campaigns by predicting the likelihood of user engagement (market interest) through binary logistic regression. The research method was carried out to analyze the effectiveness of digital advertising on interest. The market regarding participation in the Swimming Club is using the Binary Logistic Regression Method with Python and obtained two market results based on Tiktok and Instagram ads. The reach of Instagram ads is 1148 people for 2 days and Tiktok ads is 3021 people for 2 days, all areas are set in Jakarta with a specific target market, namely people who like sports, swimming, fitness and physical exercise. In Instagram ads, Variable With this method, club managers can find out the level of effectiveness of using digital advertising to achieve market response in a short time to fulfill the Club's marketing needs. Along with improving the quality of the club and meeting the maximum training needs of customers, in this case it isshared value (benefits involving shared values) that can be achieved jointly between club managers, coaches, parents and students.

***Keyword** : Binary Logistic regression, binary logistic model, ads campaign, ad campaign, digital ads, tiktok ads, Instagram ads, digital adevertising, shared value, value proportion, demographic segmentation, swimming training, physical training, biathlon.*

INTRODUCTION

Swimming, athletics (running), and physical exercise are basic sports skills that should be possessed and taught to children from an early age, not only to form a healthy body, but also to build good discipline in children. Increasing interest and participation in various sporting activities is the main focus in the world of coaching and athlete development. In this context, the Biathlon Club, which initially only taught swimming and

physical training, will now expand the training features to include swimming, swimming and physical training, middle distance swimming and running, swimming and physical training-swimming and middle distance running, athletics, physical training, swimming, middle distance running and physical training which are all the basis of biathle training.

Biathlon is taken from the word biathle, a sport discipline. Modern Pentathlon which combines elements of middle distance running and swimming, which is a basic discipline that athletes must master from an early age, has emerged as an attractive option for many individuals who are looking for a challenge and a comprehensive physical activity. In its development, the club needs more development, especially in improving the service features offered. The current swimming training held by the Biathlon Club, the number of participants has not reached the club's expected market target, namely having a minimum center covering locations in the Jakarta area with each center having a minimum of 100 participants. The only centers available are two centers in Kelapa Gading, North Jakarta, as well as a special Center for Institutions (Private).

Therefore, the Club management wants to expand the market reach of those who want to join swimming training around Jakarta, in addition to implementing the B 2 B marketing method, the management wants to carry out digital marketing using the ads campaign method via social media. However, beforehand, managers must analyze the effectiveness of digital marketing through ad campaigns on social media using the binary logistic regression method by measuring the high market interest in swimming training for respondents in the Jakarta area. The aim of preparing this journal is to analyze in depth the effectiveness of digital advertising campaigns on market interest in swimming training or classes which is one of the Club Biathlon programs.

The method used is to use binary logistic binary using python. To find out market interest in the Jakarta area, an effective online marketing strategy is needed to position the product (service offered) as an option. Prioritize the audience so they make a CTA (call to action). By understanding the effectiveness of digital advertising campaigns in the context of product marketing via social media, this is an effective first step towards engagement and repeat orders for customers who have joined.

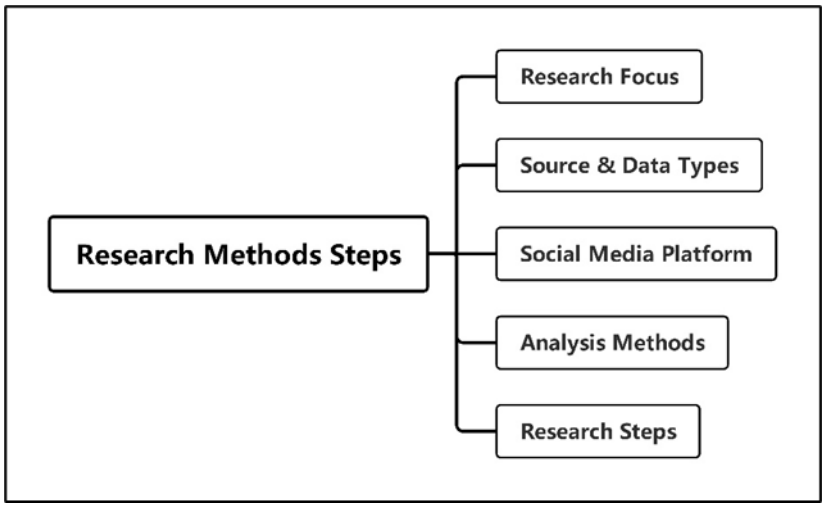
Through this journal, readers are invited to explore key findings from analyzing the effectiveness of digital advertising campaigns on market interest using the binary logistic regression method, while understanding the importance of identifying and understanding potential market needs and preferences in sustainable sports development.

Based on the background described above, this research can formulate the problem of what the consumer market segment of Biathlon club participants is. However, in this case we do not discuss segmentation in detail because we has created customer personas directly into the dataset so that we can directly operate on binary logistics using Python.

METHODS

The research method used is the Binary Logistic Regression Method using Python as an analysis to measure the effectiveness of digital advertising campaigns, in this case using social media Tiktok ads and Instagram ads. The effectiveness measured is market interest in digital advertising campaigns on both platforms, although one of them has a more effective value, but we do not discuss which platform will be used.

We have five stages in the research method which are presented in the following picture :



Picture 1. Research Methods Stages

1) Research Focus

The focus of this research is to analyze the effectiveness of the Binary Logistic Regression method in digital advertising campaigns aimed at people targeted by Club Biathlon as "new market share" to join the Club, by measuring market interest through CTA (call to Action) 'Clicked on ads '.

2) Source and Type of Data

The subject of the ad is a video campaign' which is a data source which produces data which we call a 'Dataset'. The type of data in the CSV file is numerical data derived from the respondent's demographic data.

3) Social Media

used are Instagram and TikTok ads, with several considerations, namely:

4) Analysis Method

Is using binary logistic regression. Binary logistic regression is a statistical method used to describe the relationship between a response variable (y) which is binary in nature and a predictor variable (x) which is qualitative in nature and vice versa. Binary logistic regression models are used to analyze the relationship between one response variable (y) and several independent variables. Binary logistic regression works well for binary classification problems that have only two values on the dependent variable, such as yes and no or 0 and 1.

There are three main types of logistic regression that differ in execution and theory, namely binary logistic regression, multinomial logistic regression, and ordinal logistic regression.

The advantages of using the Binary Logistic Method are:

Interpretability

The results of binary logistic regression can be interpreted relatively easily. Regression coefficients can provide information about the extent to which each independent variable influences the probability of success of an advertising campaign.

Probability of Success

Binary logistic regression provides an estimate of the probability of success of an advertising campaign. This can help in determining how likely an ad is to be successful.

Adjustment to Category Variables

Binary logistic regression can handle categorical independent variables well. This is useful when we want to include category information, such as ad type or product type, in an effectiveness analysis.

Model Capability Measurement

Logistic regression provides good model evaluation metrics, such as likelihood ratio tests, deviance, and pseudo R-squared. This helps in measuring how well the model fits the data.

Handling of time dependent Independent Variables

If there are independent variables that depend on time, binary logistic regression can be used with several modifications to handle this situation.

Flexibility in Assumptions

Binary logistic regression does not have strict assumptions about the distribution of dependent or independent variables, so it can be used in various conditions.

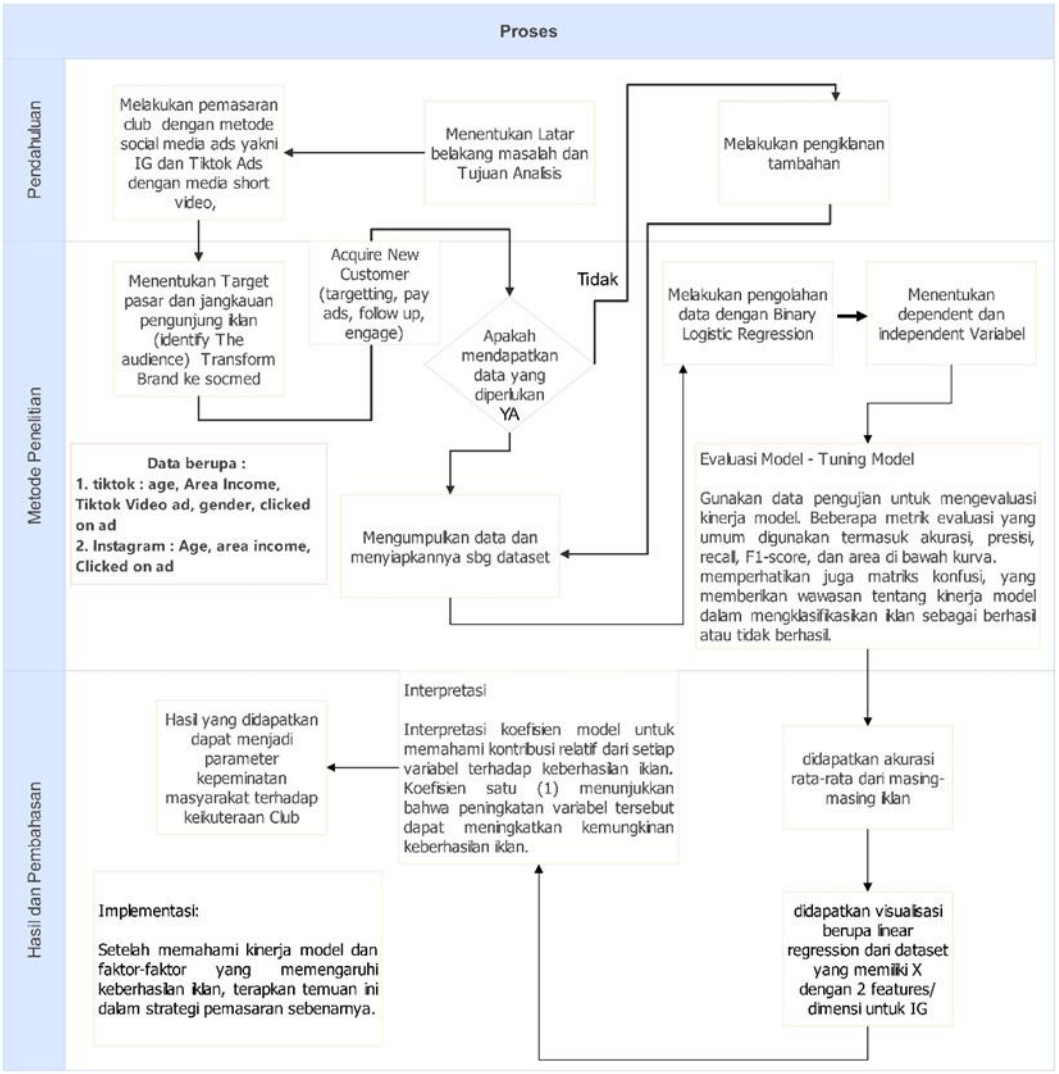
Easy to Apply

Binary logistic regression is relatively easy to implement and understand. This makes it a good choice for initial analysis or when model interpretability is important.

5) Research Method

The steps taken in this analysis method determine the accuracy of the research target from the initial method, namely determining the background of the problem and objectives to interpreting the research results. In the table below, in the data explanation section, it is stated that the data is in the form of short video ads, where the steps for creating short video ads must also fulfill several steps to provide maximum results.

This digital ad uses a short video, which is a Video Cmpaign ad, which has a duration of under 55 seconds, namely 53.85 seconds to be precise. Appropriate and strategic steps can be seen in the following chart:



Picture 2. Research Methods Steps Chart

The process we carry out in conducting digital video campaigns is:

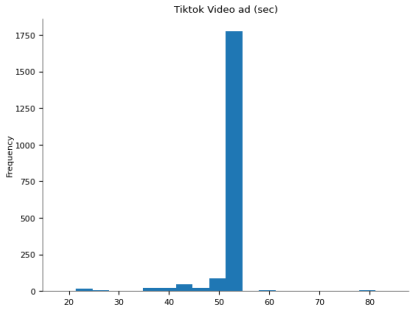
1. Identify the campaign objective, namely attracting as many respondents as possible to view the video,
2. Get to know the target audience, namely the audience aged 18 to 65 years who enjoy sports, especially swimming and physical exercise. Respondents are Instagram and TikTok users who live in Jakarta, male and female, with a certain income range.
3. Interesting Stories, interesting stories here are packaged in the form of short videos of less than 55 seconds by showing random videos from beginner to advanced level, this can attract the interest of respondents.
4. Quality audio visuals, our own videos are taken with an internal camera which is then processed with Tiktok Video editing where the result is a short video compilation
5. Consistent Brand Identity, namely the videos we display consistently display random and compiled swimming training videos from basic to advanced level.
6. Choosing the Right Platform: We use the digital platforms Tiktok and Instagram which have high reach.
7. Optimizing video duration, namely the videos we display are quite short and concise, namely under 55 seconds (53.85 seconds)
8. A clear Call to Action (CTA), namely 'Play Video'

RESEARCH RESULT AND DISCUSSION

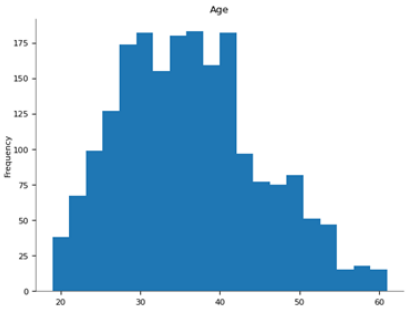
The respondents in this research are the target market for the Tiktok and Instagram digital advertising campaigns, namely Instagram and TikTok users, namely the audience aged 18 to 65 years who enjoy sports, especially swimming and physical exercise. Respondents are Instagram and TikTok users who live in Jakarta, male and female, with a certain income range.

So, returning again to the steps listed in picture 2, the initial step specified is to import all libraries, then perform a description of the data set resulting from advertising coverage and obtain the following distribution results:

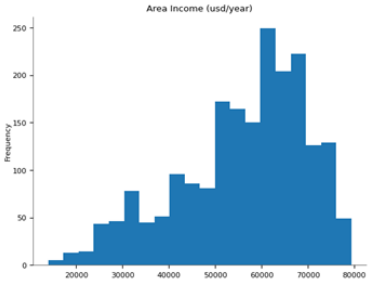
TIKTOK ADS CAMPAIGN



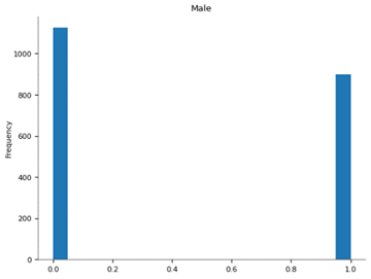
Picture 3. Tiktok Video ad Statistic



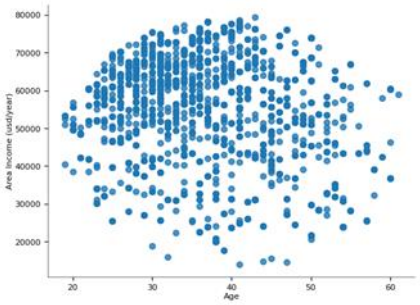
Picture 4. Tiktok Age Statistic



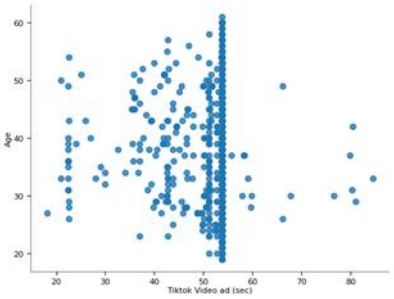
Picture 5. Tiktok Area Income Statistic



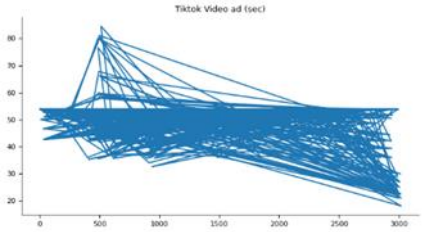
Picture 6. Tiktok Gender Statistic



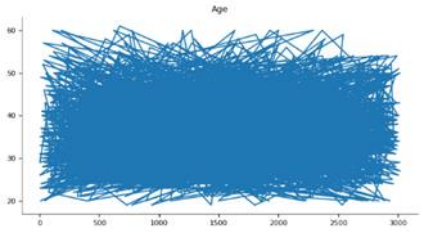
Picture 7. Tiktok Area Income vs Age Distribution



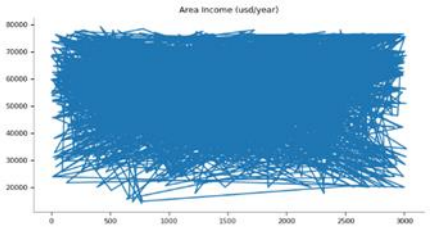
Picture 8. Tiktok Age vs Tiktok Video ad Distribution



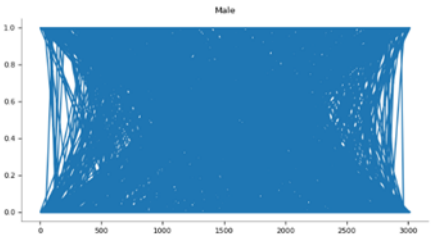
Picture 9. Tiktok Video Ad Values



Picture 10. Tiktok Age Values



Picture 11. Tiktok Area Income Values

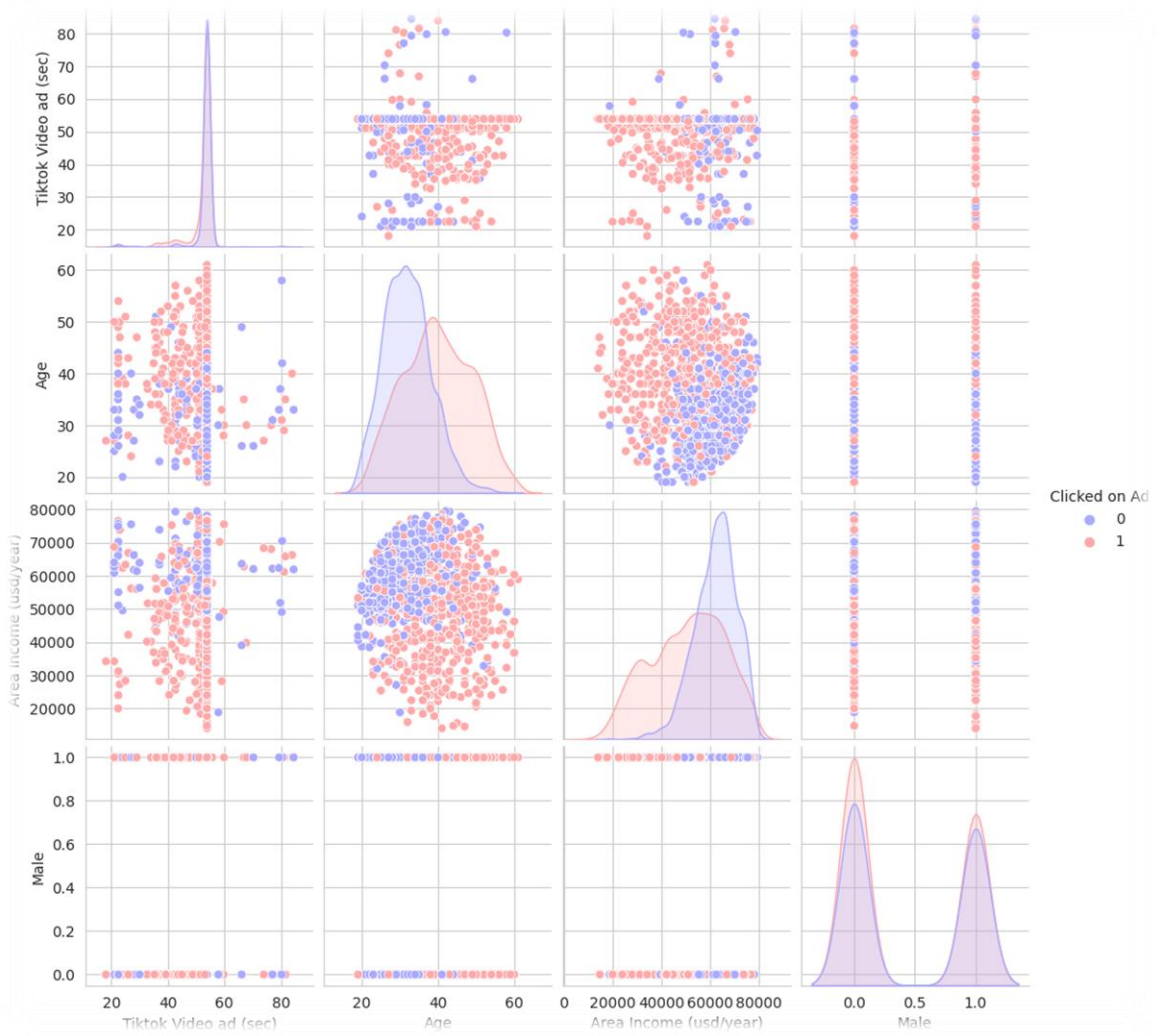


Picture 12. Tiktok Gender Values

	Tiktok Video ad (sec)	Age	Area Income (usd/year)	Male	Clicked on Ad
count	3020.000000	3020.000000	3020.000000	3020.000000	3020.000000
mean	52.679225	36.161258	55959.886748	0.445364	0.532450
std	4.949253	8.763585	13715.199675	0.497088	0.499028
min	18.110000	19.000000	13996.500000	0.000000	0.000000
25%	53.850000	29.000000	48344.530000	0.000000	0.000000
50%	53.850000	35.000000	58526.040000	0.000000	1.000000
75%	53.850000	42.000000	66281.460000	1.000000	1.000000
max	84.530000	61.000000	79484.800000	1.000000	1.000000

Picture 13. Tiktok Dataframe Result

From the description above, data is obtained that 25% of respondents are around 29 years old with an average income of 48,344 dollars/year with the majority being female, 50% are 35 years old, with an income of around 58,526 dollars/year with the majority being male. and 75% of respondents' age range is 42 years, income is around 66,281 dollars/year with the majority being male.



Picture 14. Tiktok CTA vs Distribution Features Result

The data above shows statistics on age vs Tiktok Video ads (length of watching video ads) for TikTok users who make a CTA (call to action) in the form of 'Clicked on ads' on the distribution of age, income range, gender and duration of watching videos.
From this diagram you can see an estimate of how many 'clicked on ads' and who didn't click.

'Clicked on ads' is marked with the number 1 and the color orange, then 'not clicked on ads' is marked with the number 0 and the color purple. For example, in the explanation of Area income (y) vs Tiktok Video ads (x), that of respondents with a viewing duration of 40 – 60 seconds, the majority in the income range of 20,000 to 60,000 USD/year 'Clicked on ads'. This explanation applies to other diagram functions.

Tabel 1. Logistic Regression Model

Logit Regression Results

Dep. Variable:	Clicked on Ad	No. Observations:	2023
Model:	Logit	Df Residuals:	2018
Method:	MLE	Df Model:	4
Date:	Fri, 12 Jan 2024	Pseudo R-squ.:	0.3029
Time:	00:19:06	Log-Likelihood:	-975.27
converged:	True	LL-Null:	-1399.1
Covariance Type:	nonrobust	LLR p-value:	3.706e-182

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.6537	0.725	0.902	0.367	-0.766	2.074
Tiktok Video ad (sec)	-0.0129	0.012	-1.094	0.274	-0.036	0.010
Age	0.1464	0.008	17.600	0.000	0.130	0.163
Area Income (usd/year)	-8.64e-05	5.28e-06	-16.365	0.000	-9.67e-05	-7.6e-05
Male	-0.0938	0.113	-0.834	0.404	-0.314	0.127

Optimization terminated successfully. Current function value: 0.482089 Iterations 7, is a result that typically appears after running an optimization algorithm. By using this formula,

$$Y= 1/(1+〖\exp〗^{(-(C+\beta1X1+\beta2X2+\beta3X3+\beta4X4))})$$

It indicates that the optimization process has converged to a solution, and the value of the objective function at the solution is 0.482089. The "Iterations 7" part indicates that the optimization algorithm took 7 iterations to converge to this solution. This message is commonly seen in the context of numerical optimization, such as in logistic regression or other optimization problems.

Tabel 2. Logistic Regression accuracy Result

	precision	recall	f1-score	support
0	0.72	0.84	0.78	457
1	0.84	0.73	0.78	540
accuracy			0.78	997
macro avg	0.78	0.78	0.78	997
weighted avg	0.79	0.78	0.78	997

accuracy 0.7803410230692076

The statement "precision recall f1-score support" along with the provided values is a part of the classification report generated to evaluate the performance of a classification model. Here's the meaning of each term:

Precision: It is the ratio of correctly predicted positive observations to the total predicted positives. In the given report, the precision for class 0 is 0.72 and for class 1 is 0.84.

Recall: It is the ratio of correctly predicted positive observations to the all observations in the actual class. In the given report, the recall for class 0 is 0.84 and for class 1 is 0.73.

F1-score: It is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is a good way to show that a classifier has a good value for both recall and precision. In the given report, the F1-score for class 0 is 0.78 and for class 1 is 0.78.

Support: It is the number of actual occurrences of the class in the specified dataset. In the given report, the support for class 0 is 457 and for class 1 is 540.

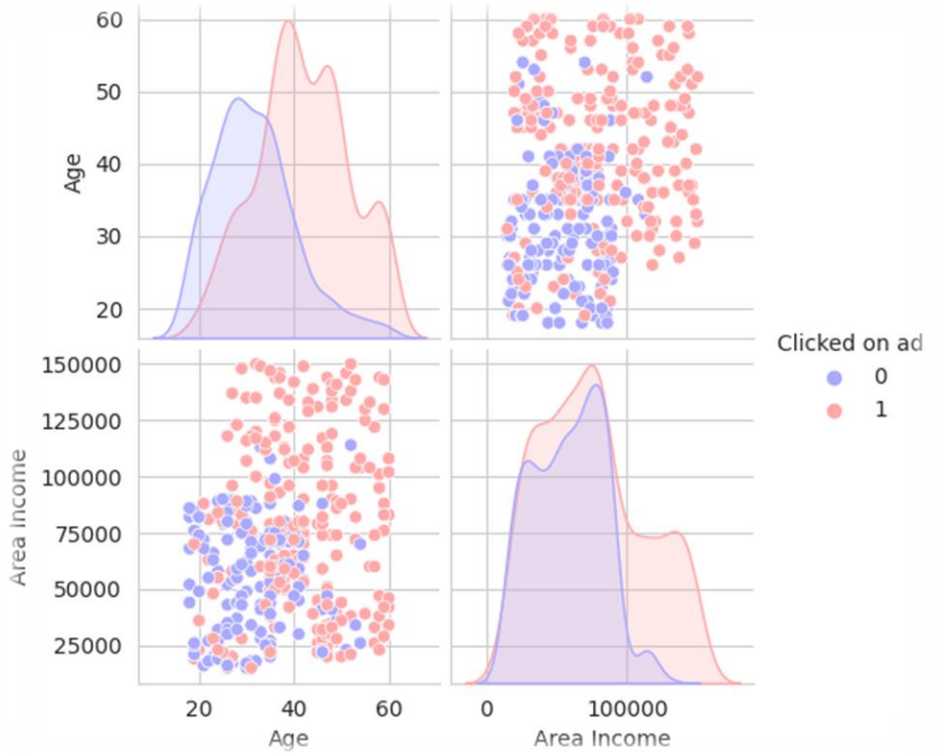
The "accuracy" line indicates the overall accuracy of the model, which is 0.78 in this case. The "macro avg" and "weighted avg" lines provide the average precision, recall, and F1-score across all the classes, with and without considering the class imbalances.

INSTAGRAM ADS CAMPAIGN

	Age	Area Income	Clicked on ad
count	1147.000000	1147.000000	1147.000000
mean	37.721011	68176.983435	0.596338
std	10.646170	33743.425090	0.490845
min	18.000000	15000.000000	0.000000
25%	29.000000	42000.000000	0.000000
50%	37.000000	66000.000000	1.000000
75%	46.000000	86000.000000	1.000000
max	60.000000	150000.000000	1.000000

Picture 15. Instagram Dataframe Result

From the description above, data is obtained that 25% of respondents are around 29 years old with an average income of 42,000 dollars/year with the majority being female, 50% are 37 years old, with an income of around 66,000 dollars/year with the majority being male. and 75% of respondents' age range is 46 years, income is around 86,000 dollars/year with the majority being



Picture 16. Instagram CTA vs Distribution Features

The data on the side shows statistics on age vs Intagram video ads (length of watching video ads) for TikTok users who make a CTA (call to action) in the form of 'Clicked on ads' in the distribution of age and income areas.

From this diagram you can see an estimate of how many 'clicked on ads' and who didn't click.

'Clicked on ads' is marked with the number 1 and the color orange, then 'not clicked

on ads' is marked with the number 0 and the color purple.
For example, in the explanation of Area income (y) vs age (x), the most Clicked on ads were among respondents with an age range of 25 to 60 years with an income range of 25,000 to 150,000 USD/year.

Tabel 3. Logistic Regression Model Result

Logit Regression Results

Dep. Variable:	Clicked on ad	No. Observations:	768
Model:	Logit	Df Residuals:	765
Method:	MLE	Df Model:	2
Date:	Thu, 11 Jan 2024	Pseudo R-squ.:	0.2322
Time:	14:53:57	Log-Likelihood:	-397.08
converged:	True	LL-Null:	-517.20
Covariance Type:	nonrobust	LLR p-value:	6.821e-53

	coef	std err	z	P>|z|	[0.025	0.975]
const	-5.1358	0.448	-11.476	0.000	-6.013	-4.259
x1	0.1201	0.010	11.765	0.000	0.100	0.140
x2	1.768e-05	2.99e-06	5.918	0.000	1.18e-05	2.35e-05

Tabel 4. Logistic Regression Accuracy Result

	precision	recall	f1-score	support
	0	0.00	0.00	155
	1	0.59	1.00	224
	accuracy		0.59	379
	macro avg	0.30	0.50	0.37	379
	weighted avg	0.35	0.59	0.44	379

\

0.5910290237467019

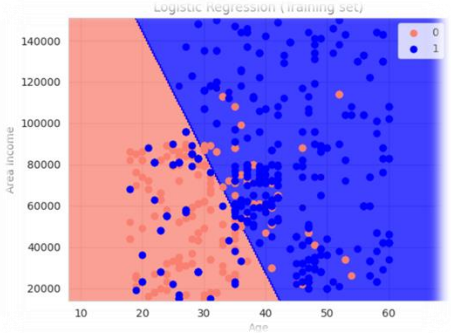
Optimization terminated successfully.
Current function value: 0.517029 Iterations 6

"Precision, recall, f1-score, and support" are performance metrics used to evaluate the performance of a classification model. The given values in the statement are the values of these metrics for two classes, 0 and 1.
Precision: It is the ratio of correctly predicted positive observations to the total predicted positives. In the given report, the precision for class 0 is 0.00 and for class 1 is 0.59.
Recall: It is the ratio of correctly predicted positive observations to the all observations in the actual class. In the given report, the recall for class 0 is 0.00 and for class 1 is 1.00.
F1-score: It is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is a good way to show that a classifier has a good value for both recall and precision. In the given report, the F1-score for class 0 is 0.00 and for class 1 is 0.74.

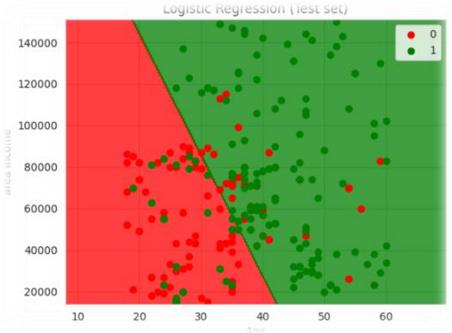
Support: It is the number of actual occurrences of the class in the specified dataset. In the given report, the support for class 0 is 155 and for class 1 is 224.

The "accuracy" line indicates the overall accuracy of the model, which is 0.59 in this case. The "macro avg" and "weighted avg" lines provide the average precision, recall, and F1-score across all the classes, with and without considering the class imbalances.

The search results provided some additional information about these metrics and their interpretation. Precision and recall are important metrics in classification problems, and F1-score is a measure that combines both precision and recall. The F1-score is often more informative than the raw accuracy, especially if you have an uneven class distribution.



Picture 17. Instagram CTA Logistic Regression (Trainging Set)



Picture 18. Instagram CTA Logistic Regression (Test Set)

Evaluation of model performance, including confusion matrices and accuracy, can depend greatly on the context and specific characteristics of the classification problem at hand. Therefore, there is no exact number or threshold that can be considered a “good result” for accuracy or other evaluation metrics. Good or bad results will depend largely on the nature of the data, the class distribution, and the business or scientific objectives involved.

Good results from the accuracy of the confusion matrix can vary depending on the application context. In general, accuracy is one of the important evaluation metrics in assessing the performance of classification models. However, accuracy alone may not be sufficient to provide a complete picture, especially when there is class imbalance in the dataset. For example, in the case of class imbalance, where one class may be much more dominant than the others, accuracy can be biased towards the majority class. Instead, other metrics such as precision, recall, and F1-score also need to be considered.

Precision measures how many of the predicted positive classes are actually positive, while recall measures how many of the positive classes are actually predicted correctly. The F1-score is the harmonic average of precision and recall, and is often more appropriate than accuracy in cases of class imbalance. Thus, while high accuracy is generally desirable, it is important to consider other evaluation metrics and the specific context of the classification problem at hand.

Compare model results with other models or baselines to get perspective on how well the model performs.

It is important to remember that no metric is perfect and that performance assessment models should always be considered in a broader context.

However, in this case we consider that the results of the logistic regression diagram show that digital advertising will be used as a marketing method. This can be seen from the positive factor, namely 1, which indicates the number of markets that "Clicked on ads" which shows interest as an initial parameter. It is positive to continue marketing using ads campaigns via social media, which is EFFECTIVE.

CONCLUSION

Binary logistic regression is used in measuring the effectiveness of online advertising campaigns because it can predict binomial variables, such as the gender of internet users, and can be used to estimate multinomial variables, such as brand or product category preferences.

In the context of online advertising campaigns, binary logistic regression can help in understanding consumer preferences and behavior, as well as help in determining the right target audience for campaign advertising.

In addition, binary logistic regression can also be used to model the relationship between independent variables and dependent variables, such as the influence of advertising on consumer behavior.

In this case, binary logistic regression can help in measuring the effectiveness of campaign advertising by predicting the likelihood that consumers will take certain actions, such as buying a product or visiting a company website, if in this research certain actions are reflected by "Clicked on ads".

Overall, binary logistic regression is a useful tool in measuring the effectiveness of online campaign advertising as it can help in understanding consumer preferences and behavior, as well as help in determining the right target audience for campaign advertising.

From the results of the logistic regression diagram, digital advertising is EFFECTIVE when used as a marketing method. This can be seen from the positive factor, namely 1, which indicates the number of markets that "Clicked on ads" which shows interest as a positive initial parameter for continuing marketing using ads campaigns. via social media.

It is important to remember that interpretation of binary logistic regression results should be done with caution, and the results should be viewed as probabilities. Overall, binary logistic regression is a powerful tool for evaluating and understanding advertising effectiveness, but it needs to be used with the right context and considering the unique characteristics of each advertising campaign.

DAFTAR PUSTAKA

Edvidel Arda, Arif Aulia, Jerry Heikal. (2024). Analysis Of Factors Influencing The Performance Of Payakumbuh City Public Housing And Residential Area Employees Using Binary Logistic Regression. Vol. 01 No. 03 Edition January - March 2024 Pg. 379-384https://jurnal.ittc.web.id/index.php/jebd/article/view/547/493.

O.Cahya, R.A & Cahyono, P.R (2022) Analysis of Service Satisfaction at Mitra Husada Pringsewi Hospital Using the Logistic Regression Method. Ilmudata.org Volume 2 (12), 2022.

Sugeng Santoso. Model Logit (Logistic Regression) Sebagai Model Probabilitas Linier Alternatif, Dalam Contoh Aplikasi Teknik Industri. doi: 10.22219/JTIUMM.Vol2.No1.59-64.

Zabrinna Zenitha Zahroh, dan Ismaini Zain. Analisis Regresi Logistik Multinomial Pada Faktor-Faktor Yang Mempengaruhi Sumber Air Bersih Rumah Tangga di Jawa Timur. Jurnal Sains Dan Seni ITS. Vol. 7, No. 2 (2018), 2337-3520 (2301-928X Print).

Dian Yulis Wulandari, Ismaraidha, Ocdy Amelia. Analisis Model Regresi Logistik Biner Dalam Menentukan Faktor – Faktor Yang Mempengaruhi Masa Studi Mahasiswa. Journal of Mathematics Education and Science. Vol. 8, No. 1, Oktober 2022.

Ali Ilham Sofiyat , Awaluddin Tjalla , Mahdiyah. Pemodelan Regresi Logistik Biner Terhadap Penerimaan Pegawai di PT XYZ Jakarta. Matematika Sains Volume 1 No.1 Tahun 2023.

Hendra H. Dukalang. Perbandingan Regresi Logistik Biner dan Probit Biner Dalam Pemodelan Tingkat Partisipasi Angkatan Kerja. Jurnal Ilmiah Matematika, Sains dan Teknologi Desember 2019, Vol. 7, No. 2, Hal. 62-70 doi: https://doi.org/10.34312/euler.v7i2.10355.