Proceeding of International Joint Conference on UNESA
Homepage : https://proceeding.unesa.ac.id/index.php/pijcu
p-ISSN : XXXX-XXXX; e-ISSN : XXXX-XXXX

PIJCU, Vol. X, No. X, XXXX 2023
Page XX-XX

Integration of Autoregressive Distributed Lag (ARDL) Modeling with Google Search Console Data Analysis for SEO Performance Evaluation

Figo Gymnastiar Farhaan Pratama[1], Asmunin[2], Hafizhuddin Zul Fahmi[3], I Gde Agung Sri Sidhimantra[3], Dodik Arwin Dermawan[5]
[1, 2, 3, 4, 5] State University of Surabaya, Surabaya, Indonesia

Keywords:
Search Engine Optimization
Google Search Console
Autoregressive Distributed Lag
Web-based Application
CTR

ABSTRACT

The author designed a website that can be used to analyze Google Search Console performance results with Python and Django implementations to take the Autoregressive Distributed Lag method calculation function with the aim of drawing conclusions about the significance between variables in the long term. Web users will be asked to fill out a form to upload Google Search Console data to be analyzed in CSV format. Then the system will perform three stages of testing, namely stationary tests, optimum lag tests, and regression tests. After all tests are done, the system will display the test results on the same page. The result of this analysis is that the website can analyze Google Search Console datasets that can be uploaded via the CSV document form and will be saved by the system to the database. The conclusion is in long-term relationship of the Clicks variable is valid and variable Impressions, Ranking, CTR, have a positive and significant effect on increasing the number of Clicks. And the results of regression testing with variable dependent CTR is valid. It shows the const and Clicks variable in the long run have a positive and significant effect on the average increase in CTR.

INTRODUCTION

In the last decade, there has been a significant evolution of marketing to become more digitalized with the presence of online platforms and digital marketing technologies. Of all the modern marketing techniques and tools, digital marketing resources, channels, and social media are the most interactive, engaging, and important [3]. Digital promotion is a very innovative and new idea in the 21st century. Although marketing products or assistance via the internet is very efficient but expensive than traditional product marketing [8]. Digital marketing is the latest and most effective way to market products in the digital era. With the potential to gain buyer reach and increase brand awareness which aims to increase product sales, this is the reason why business people should consider digital marketing strategies in their sales. Search Engine Optimization (SEO) is one of the digital marketing strategies that utilizes the Google search engine algorithm because in several website page views, sellers can already provide various product information needed by buyers. Google also plays an important role in sorting the information displayed on buyer searches based on the ranking of the most relevant websites with search keywords, search locations, and interesting content [9].

This study will provide an overview of how digital promotion will be more efficient with some data presented through Google Search Console, such as the number of website visitors calculated through clicks and impressions, the average Click Through Rate generated by website visitors, and the average website search ranking after being indexed by the Google search engine. The application of the Search Engine Optimization concept will be able to help develop the MSME market by increasing the visibility of swallow bird sound products to potential customers throughout Indonesia [1]. Google Search Console

analyzed data to map the company's digital marketing strategy that is being and will be implemented [6]. The method used in this study is the Marketing Funnel, a marketing concept that explains the customer experience before getting to know the product, getting to know it, making decisions, to making purchases and evaluating the company's products. The purpose of this study is to measure the effectiveness of digital marketing owned by the Graha Office company in the period August-December 2023.

With increasing awareness of the benefits of modern digital marketing, business actors are required to understand the correlation of digital marketing strategies applied to micro business products with the visibility of their websites on search engines [7]. However, in reality, many MSME business actors are not yet aware of the benefits of digital technology for their business. Lack of understanding of the benefits and how to do business on the internet is an obstacle that will be faced during the research. MSME business people need to develop their business to be able to survive in this digital industry by applying the concept of digital marketing, especially SEO, to their product websites. Through this research, we design a website that can analyze business web performance and provide conclusions on short-term & long-term strategies that must be used to improve business web performance in order to gain wider customer reach. By having a website that reaches a wide traffic area, it will have high potential to become a good business [4], so that it can expand the market reach so that the scope of potential customers will also be greater.

Search Engine Optimization is an important concept in digital marketing because of its ability to increase website visibility in the Search Engine Result Page (SERP). A search engine is an application programmed to help search for information on the internet by entering search keywords. Search engines are usually accessed through pages on websites that allow users to search billions of website content [10]. According to Janner Simarmata, keywords will be matched by the search engine with its database [11]. Researchers who study SEO have various points of view and opinions in defining SEO itself, citing in their journal who argue that SEO is "a science applied by webmasters, web developers, and website content writers to get top ranking on search engine pages." [2]. They also argue that the high percentage of internet traffic that is done through search engines is directly proportional to the importance of a website's ranking on search pages. Google Search Console is a free service from Google that has tools and analytics to improve website optimization on the Google search engine. GSC provides insight into how a website's performance interacts with Google search. Google Search Console can facilitate the work of web developers by analyzing errors, providing website performance data, and correcting SEO weaknesses on the website to be developed for the better. Google Search Console can also sort out which pages to display or hide from Google search pages, this will be useful if there is information about the website that web visitors do not need to know or if you want to form a certain branding for the product being sold.

RESEARCH METHOD

The method section of this report will provide a comprehensive overview of the research design, detailing the systematic approach and procedures employed to conduct the study. It will outline the data collection methods, specifying the tools and techniques used to gather relevant information, including surveys, interviews, and observational

strategies. Additionally, the section will explain the data analysis process, describing how the collected data was processed, interpreted, and evaluated using various analytical techniques and statistical tools. This thorough explanation will ensure clarity and transparency in understanding the methodologies utilized, enabling replication and validation of the research findings.

This research uses a quantitative approach. Meaning, this is approach used in a study that uses data in the form of numbers and an analysis process using statistical tools [13]. This study uses a graphical approach and statistical tests, where as much as 4 data obtained from the Suarawalet.id website will be analyzed using the Autoregressive Distributed Lag model on the website designed by the researcher to determine the advantages and disadvantages of website performance in the short and long term. The characteristics of quantitative research are systematically arranged, planned, and clearly structured from the data collection stage to data analysis [12]. This research uses time series data. Time series data is data that is chronologically arranged based on time used to see the influence within a certain time span [5]. In this study, researchers took a time limit of 6 months, starting from April 1, 2023 to October 1, 2023. The data variables taken are the number of clicks, the number of impressions, the average percentage of Click Through Rate (CTR), and the average ranking. The variables mentioned on Table 1 are explained operationally with the following definitions.

Table 1. Variable Definitions

Variable	Definition	Long-term Hypotesis (α_x)	Short-term Hypotesis (β_x)
Clicks	The number of user clicks to your site is affected by the type of search results and how they are tabulated	Dependent Variable	Dependent Variable
Impressions	The frequency with which users find links to websites in search results varies, with records for image search and other results determined by the number of impressions on the user's screen.	Positive	Positive
Ranking	A website's average rank in search results is determined by its highest rank each time it appears in SERPs. The rank of each page is detailed in different tables on Google Search Console.	Negative	Negative
CTR	It represents the percentage of impressions that convert into clicks.	Dependent Variable	Dependent Variable
e	Error Term	-	-

From the web-based application that will be developed by researchers, application users can analyze website performance data obtained from Google Search Console by entering the value of each data. There are 3 tests of data analysis that will be carried out by the application to get maximum conclusions and results

Stationarity test will be the first test performed by users in analyzing Google Search Console data. This test will need to be done so that the value of each variable under study is around the average value with fluctuations that are independent of time and variance. In other words, the stationary test will identify influential variables in time series analysis so that the resulting short-term and long-term estimation calculations become consistent and can be interpreted appropriately. In the context of this research, researchers will conduct a stationary test on the number of clicks, number of impressions, average rank, and CTR (Click Through Rate) data. This stationarity test will also use the ADF (Augmented Dickey Fuller) method which is useful for showing whether a variable is stationary or not. The variable will be declared stationary if the ADF probability level of the variable touches the Mackinnan Critical Value (<5%).

After the stationarity test has been carried out, it will proceed to the next test, namely the optimum lag test. This is important to do at an early stage before the calculation is carried out because it will reveal that the optimum lag test is carried out to determine the most optimal lag criteria that will be used for further analysis. The selection of the right lag optimum can produce accurate calculations in the process of long-term identification of each variable. There are several criteria used in determining the optimal lag, namely the Likelihood Ratio (LR), Final Prediction Error (FPE), Akaike Information Criterion (AIC), Schwarz Information Criterion (SC), and Hannan-Quinn Information Criterion (HQ). The way to determine the most optimal lag criteria is by looking at the highest number of stars contained in each of these criteria and the optimum lag will be used in the next test.

The Autoregressive Distributed Lag (ARDL) model is the last test in a series of data analysis on Google Search Console. Regression test will be adopted as a data analysis technique in this study to determine the effect of variables in one period by using the optimum lag of the dependent variables (ΔClick & ΔCTR) simultaneously.

$$\boldsymbol{clicks_t = \gamma_1\, impressions_t + \gamma_2\, ranking_t + e} \quad (1)$$

where $clicks_t$ is total clicks variable $impressions_t$ is total impressions variable, and e is error term. Based on equation 1, the ARDL calculation formula equation is as follows.

$$\Delta \boldsymbol{clicks_t} = \sum_{k=1}^{n1} \boldsymbol{\alpha_1 \Delta impressions_{t-1}} + \sum_{k=1}^{n2} \boldsymbol{\alpha_2 \Delta ranking_{t-1}} + \boldsymbol{\beta_1 \Delta impressions_{t-1}} + \boldsymbol{\beta_2 \Delta ranking_{t-1}} + \boldsymbol{e} \quad (2)$$

Δ_t is the first difference of each variable, coefficient $\alpha_1 - \alpha_2$ show the long dynamic relationship between variables, coefficient $\beta_1 - \beta_2$ show the short-term dynamics between variables. From equation 2, the calculation of error connection can be defined as follows

$$\Delta \boldsymbol{clicks_t} = \sum_{k=1}^{n1} \boldsymbol{\alpha_1 \Delta impressions_{t-1}} + \sum_{k=1}^{n2} \boldsymbol{\alpha_2 \Delta ranking_{t-1}} + \boldsymbol{\beta_1 \Delta impressions_{t-1}} + \boldsymbol{\beta_2 \Delta ranking_{t-1}} + \boldsymbol{\delta EC_{t-1}} + \boldsymbol{e} \quad (3)$$

Where EC_{t-1} is defined as error connection and δ can be defined as error connection parameter. Based on the results of long-term regression testing, it is found that the variable C (error term) is negative and insignificant to the number of clicks. While the two variables number of impressions and average rating are positive and significant to the number of clicks. To get accurate and consistent results, this study also uses an approach to the effect of average rank with average CTR.

$$\boldsymbol{CTR_t = \alpha_1\, ranking_t + \beta_1\, ranking_t + e} \quad (4)$$

where CTR_t is the Click Through Rate (CTR) variable, Type equation here. is the ranking variable, and e is the error term. Based on equation 4, the ARDL calculation formula equation is as follows

$$\boldsymbol{\Delta CTR_t = \sum_{k=1}^{n1} \alpha_1 \Delta ranking_{t-k} + \beta_1 \Delta ranking_{t-k} + e} \quad (5)$$

Based on the results of long-term regression testing in equation 5, it is found that the variable C (error term) is positive and significant to the average CTR. While the average rank is positive and significant to the average CTR.

RESULTS AND DISCUSSION

This section presents the research results and findings, showcasing the data collected and analyzed throughout the study. The findings are meticulously displayed using a combination of words, tables, figures, and photographs to ensure clarity and comprehensibility. There are 3 tests of data analysis that will be carried out by the application to get maximum conclusions and results, as if:

In this stationarity test shows the system will take the variables of number of clicks, number of impressions, average rank, and CTR (Click Through Rate) and then the analysis process will be carried out using the ADF (Augmented Dickey Fuller) method which aims to determine the stationarity or not of a variable. The variable will be declared stationary if the ADF probability level of the variable touches the Mackinnan Critical Value (less than 5% or 0.05). The following are the results of the variable stationarity test. It can be seen in Figure 1 that all variables tested have a P-value of less than 5% or 0.05, so all variables can be declared stationary.

Variable	ADF Statistic	p-value
Clicks	-10.838	0.000
Impressions	-4.112	0.006
Ranking	-12.345	0.000
CTR	-6.379	0.000

Figure 1. Stationerity Test

The optimum lag test is conducted to determine the most optimal lag criteria that will be used for further analysis. The selection of the right optimum lag can produce accurate calculations in the long-term identification process for clicks in Figure 2, impressions in

Figure 3, and ranking in Figure 4 to find the most optimal lag calculation for each variable. The results of the most optimal lag will be used as variables in the regression test.

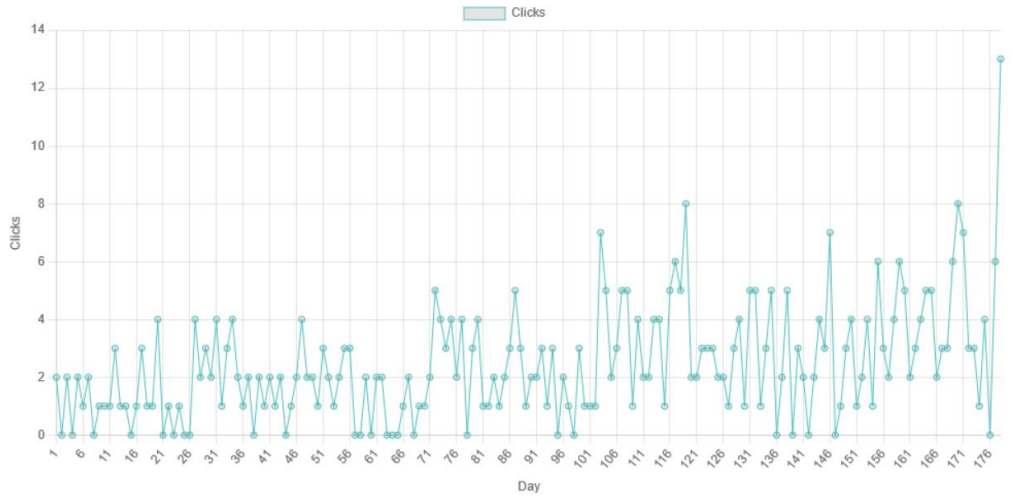

Figure 2. Optimum Lag Test Results for Variable Clicks

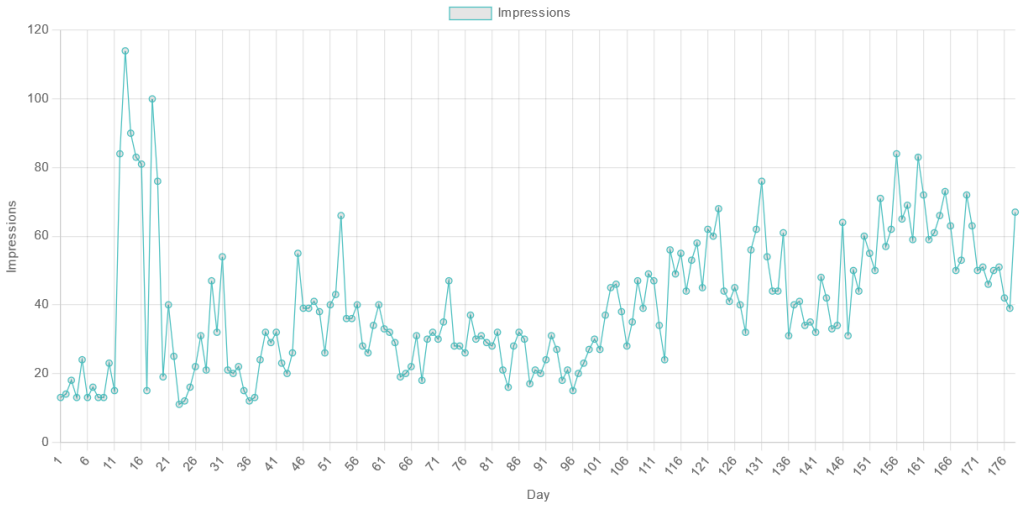

Figure 3. Optimum Lag Test Results for Variable Impressions

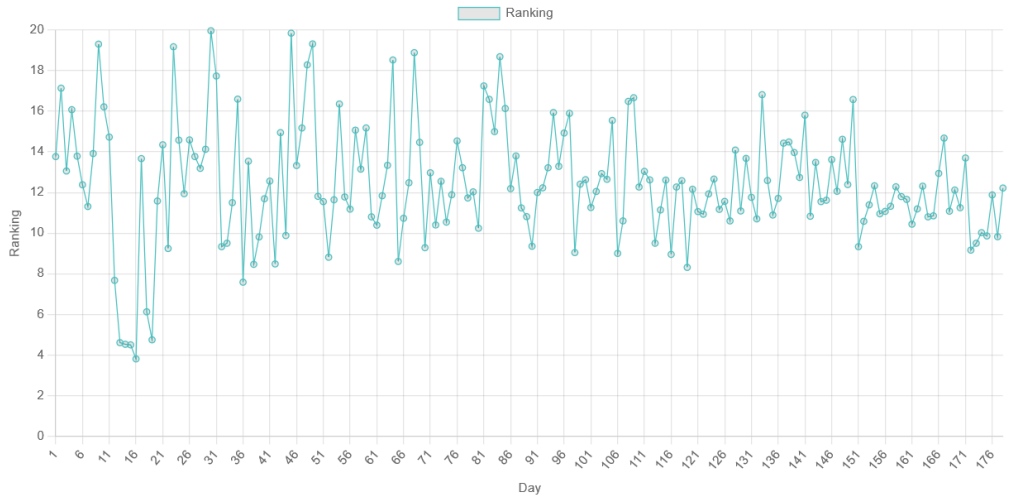

Figure 4. Optimum Lag Test Results for Variable Ranking

Regression tests are useful for seeing the long-term effect of a variable in a certain period of time by using the optimum lag of the dependent variable. A variable is declared to have an influence on the dependent variable if the coefficient value (coef) is positive. Then the significance of a variable on the dependent variable can be seen from the t-statistic (t) which is positive. This is the results from regression test with the dependent variable Clicks according to GSCount. Based on the results of long-term regression test in Figure 5, the coef and t-statistics of the variables C (error term) and Impressions are positive, indicating significance to the dependent variable Clicks. Meanwhile, the coef of Ranking is negative and the t-statistics of Ranking is negative, which means there is no significance to the dependent variable Clicks. To get accurate and consistent results, this research also uses analysis with the average CTR dependent variable.

OLS Regression Results

Dep. Variable:	Clicks	**R-squared:**	0.790
Model:	OLS	**Adj. R-squared:**	0.786
Method:	Least Squares	**F-statistic:**	217.7
Date:	Wed, 10 Jul 2024	**Prob (F-statistic):**	1.18e-58
Time:	08:29:22	**Log-Likelihood:**	-233.62
No. Observations:	178	**AIC:**	475.2
Df Residuals:	174	**BIC:**	488.0
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-3.0157	0.437	-6.907	0.000	-3.878	-2.154
Impressions	0.0562	0.004	14.299	0.000	0.048	0.064
Ranking	0.0834	0.026	3.253	0.001	0.033	0.134
CTR	33.2854	1.428	23.313	0.000	30.467	36.103

Omnibus:	29.193	**Durbin-Watson:**	1.334
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	183.701
Skew:	0.255	**Prob(JB):**	1.29e-40
Kurtosis:	7.951	**Cond. No.**	960.

Figure 5. Regression Test Results Variable Clicks

Based on the results of long-term regression testing on Figure 6, it is found that the coef variable C (error term) is positive and the t-statistics are also positive, which indicates significance to the CTR average. While the Ranking coef is negative and the Ranking t-statistics are negative, which means that it is not significant to the CTR average dependent variable.

OLS Regression Results

Dep. Variable:	CTR	**R-squared:**	0.766
Model:	OLS	**Adj. R-squared:**	0.762
Method:	Least Squares	**F-statistic:**	189.9
Date:	Wed, 10 Jul 2024	**Prob (F-statistic):**	1.23e-54
Time:	08:29:22	**Log-Likelihood:**	415.02
No. Observations:	178	**AIC:**	-822.0
Df Residuals:	174	**BIC:**	-809.3
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.0928	0.011	8.588	0.000	0.071	0.114
Clicks	0.0228	0.001	23.313	0.000	0.021	0.025
Impressions	-0.0014	0.000	-12.999	0.000	-0.002	-0.001
Ranking	-0.0022	0.001	-3.224	0.002	-0.003	-0.001

Omnibus:	31.707	**Durbin-Watson:**	1.746
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	61.278
Skew:	0.847	**Prob(JB):**	4.94e-14
Kurtosis:	5.322	**Cond. No.**	277.

Figure 6. Regression Test Results Variable CTR

In this section, system will define what is the conclusion of all three tests as you can see in Figure 7, Figure 8, and Figure 9 so users can get the digital marketing suggestion on how and what to improve on Google Search Console. Hoping, the micro businesses can easily improve their digital marketing and search engine optimization strategies with the help of GSCount suggestion on test conclusion on Figure 10.

STATIONERITY TEST

In the first test, namely the stationary test, the variable is declared stationary if the P-Value is less than 5% or 0.05.

Conclusions
Clicks is stationary
Impressions is stationary
Ranking is stationary
CTR is stationary

Figure 7. Stationerity Test Conclusion

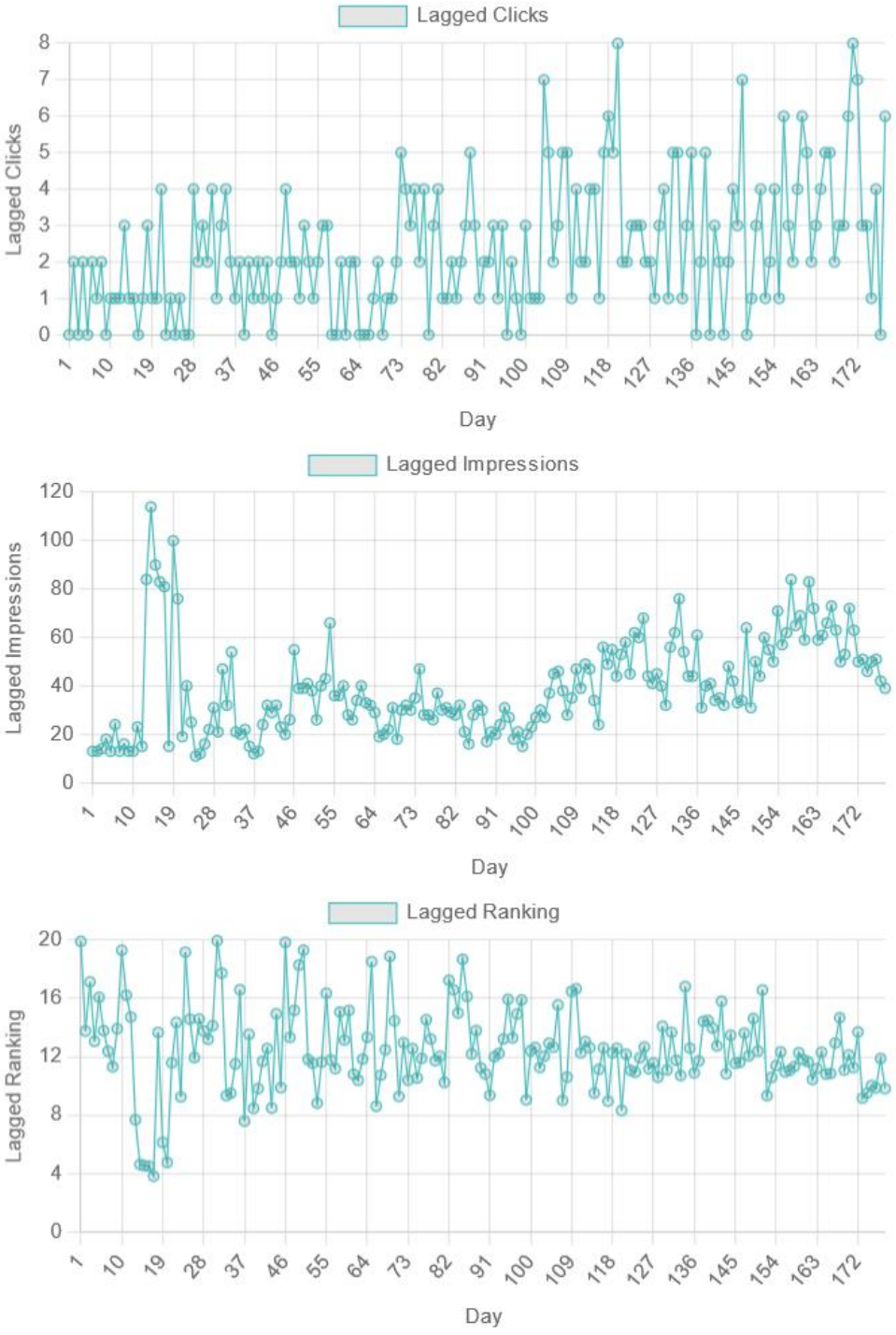

Figure 8. Optimum Lag Test Conclusion

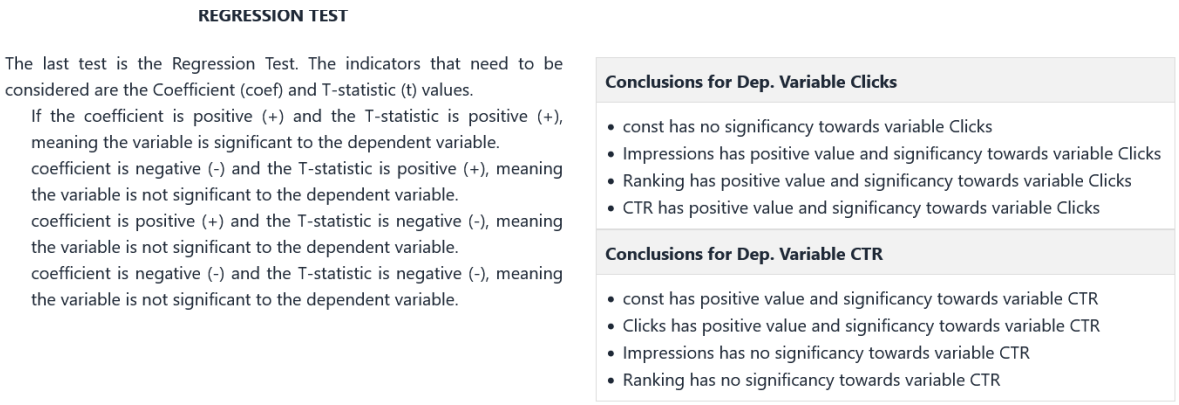

Figure 91. Regression Test Conclusion

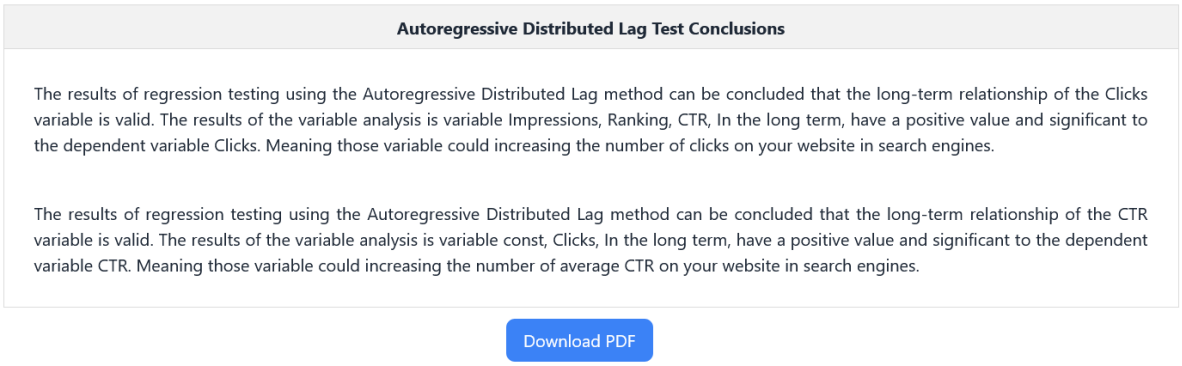

Figure 10. Test Conclusion and Suggestion

CONCLUSION

The conclusion of the Development of Website-Based Google Search Console Data Analysis Application Using Autoregressive Distributed Lag Method is that the GSCount web can analyze Google Search Console datasets that can be uploaded via CSV document form and will be saved by the system to the GSCount database.

The results of regression testing using the Autoregressive Distributed Lag method can be concluded that the long-term relationship of the Clicks variable is valid. The results of the analysis of the variables Impressions, Ranking, CTR, in the long run both on the dependent variable Clicks have a positive and significant effect on increasing the number of clicks on your website in search engines. The results of regression testing using the Autoregressive Distributed Lag method can be concluded that the long-term relationship of the CTR variable is valid. The results of the analysis of the const variable, Clicks, in the long run on the dependent variable CTR have a positive and significant effect on the average increase in CTR. This means that if you intend to increase sales (CTR), then you must increase the number of clicks (Clicks) of visitors. In addition, the analysis results can also be downloaded in pdf format to be accessed without the internet.

ACKNOWLEDGEMENTS

I would like to thank PT Lentera Alam Nusantara for their support and assistance throughout the course of this research and the writing of this report. The resources provided by PT Lentera Alam Nusantara have been vital in the execution and completion

of our study. I also thank the lecturer for any contributions and guidance have been instrumental in ensuring the success of this project.

References

[1] A. S. Attia and N. Morsy, “Analyzing the Role of Search Engine Optimization (SEO) as an E-Marketing Supportive Tool (Tourism Companies' Websites),” *The Scientific Journal of the Faculty of Tourism and Hotels,* pp. 6-7, 2023.

[2] A. Panchal, A. Shah and K. Kansara, "Digital Marketing - Search Engine Optimization (SEO) and Search Engine Marketing (SEM)," *International Research Journal of Innovations in Engineering and Technology (IRJIET),* pp. 17-21, 2021.

[3] A. Kovalenko and K. Yuliya, "Online Marketing Impact on Micro-Enterprises: An Insighgt through Visibility in Search Engines," *Sciendo,* pp. 38-58, 2020.

[4] A. Warokka, H. Sjahruddin, S. Sriyanto, E. Noerhartati and K. Saddhono, "Digital Marketing Support and Business Development Using Online Marketing Tools: An Experimental Analysis," *International Journal of Psychosocial Rehabilitation,* pp. 1475-7192, 2020.

[5] D. Mladenovic, A. Rajapakse, N. Kozuljevic and Y. Shukla, "Search Engine Optimization (SEO) for Digital Marketers: Exploring Determinants of Online Search Visibility for Blood Bank Service," *Online Information Review,* pp. 1468-4527, 2023.

[6] H. Hasyim, J. Simarmata and N. Nasirwan, "Building marketing performance through digital marketing and database-based networking capability in Indonesian SMEs.," *International Journal of Data and Network Science,* vol. 6(4), pp. 1125-1134, 2022.

[7] H. Ayad, S. E. Sari-Hassoun and M. Usman, "The impact of economic uncertainty, economic growth and energy consumption on environmental degradation in MENA countries: Fresh insight from multiple thresholds NARDL approach," *Environ Sci Pollut Res 30,* pp. 1806-1824, 2023.

[8] S. Long and J. Liang, "Asymmetric and nonlinear pass-through of global crude oil price to China’s PPI and CPI inflation.," *Economic research-Ekonomika istrazivania,* vol. 5(2), pp. 240-251, 2018.

[9] L. A. Abdillah, "FinTech E-commerce payment application user experience analysis during COVID-19 pandemic.," *arXiv preprint arXiv:2012.07750.,* 2020.

[10] D. Wijaya and B. G. Y. Daniawan, "Search Engine Optimization (SEO) As A Promotional Media on Google Search," *Bit-Tech,* vol. 4(1), pp. 31-39, 2021.

[11] L. J. Duckett, "Quantitative Research Excellence: Study Design and Reliable and Valid Measurement of Variables," *Journal of Human Lactation,* vol. 37, no. 3, 2021.

[12] M. F. Bockarie, E. K. Duramany-Lakkoh, H. Jalloh and E. Udeh, "STABILITY OF MONEY DEMAND IN SIERRA LEONE USING AUTOREGRESSIVE DISTRIBUTED LAG (ARDL) APPROACH 1980 – 2018," *European Journal of Economic and Financial Research,* vol. 8, no. 1, pp. 34-35, 2024.

[13] K. Alexopoulos, I. Anagiannis, N. Nikolakis and G. Chryssolouris, "A quantitative approach to resilience in manufacturing systems," *International Journal of Production Research,* vol. 60, no. 24, p. 7178–7193, 202.