 ISSN 2714-7533

Deciphering the Visibility of Higher Education Institutions: A Statistical Analysis of Google Search Console Data

Ikhwan Arief[a,1,*]

[a] Industrial Engineering Department, Faculty of Engineering, Universitas Andalas, Padang, West Sumatra, Indonesia
[1] ikhwan@eng.unand.ac.id
* corresponding author

ARTICLE INFO

Article history
Received November 24, 2022
Revised January 27, 2023
Accepted February 18, 2023

Keywords
Webometrics
google search console
website visibility
predictive model
education sector

ABSTRACT

The research presented here delves into the connection between data from Google Search Console (GSC) and the Webometrics visibility score of a specific public university's web presence. The study scrutinized GSC parameters such as clicks, impressions, click-through-rate (CTR), and average position to assess their impact on the university's digital visibility. The results indicate that impressions and average position play a critical role in determining the Webometrics visibility score, underscoring the significance of search engine optimization for learning establishments. The research also pinpointed the most effective search queries that drive substantial visitor traffic to the university's website, underlining the need for precise content targeting to optimize search performance. In this study, a predictive model was developed using multiple linear regression analysis to accurately predict the Webometrics visibility score based on GSC metrics, suggesting that strategic efforts to enhance these parameters could boost a university's online prominence. Additionally, a theoretical model was proposed to clarify the dynamic relationship between impressions, positions, and clicks in shaping the overall web visibility. Although this study provides valuable insights, it is based on data from a single university, which calls for further investigation using more varied datasets. Ultimately, the study emphasizes the immense potential of leveraging GSC data to bolster a university's online footprint, suggesting that strategic enhancements of vital parameters can greatly improve a university's online visibility according to Webometrics. As the academic world becomes increasingly digital, implementing these findings to guide search engine optimization strategies is a crucial element of institutional administration.

1. Introduction

The swift progression of digital technology has revolutionized various aspects of our society, including the educational sector. The ubiquitous nature of the internet has heightened the importance of maintaining a strong online presence for educational institutions. Universities across the globe are increasingly appreciating the benefits of a solid online presence, recognizing its potential to bolster their educational agendas, outreach activities, and international reputation [1]. However, developing and sustaining a powerful online presence is no easy feat. It entails optimizing website content, ensuring its accessibility and visibility, and comprehending how this virtual representation connects with internet users.

There exists a conspicuous void in quantifying the effectiveness of these online strategies and comprehending the correlation between this online presence and widely acknowledged metrics of academic prestige and influence. Conventional methods for assessing academic influence, such as

research publication and citation counts, academic reputation, graduate employment rates, and so on, fail to adequately capture the digital clout of these institutions [2][3]. It is within this void that the current study finds its position.

The Webometrics ranking of universities, a globally respected academic ranking system, introduces a crucial "Visibility" parameter that quantifies the digital influence of universities [4]. It is calculated by counting the distinct external networks linking back to the institution's web pages. However, the intricacies of how different website elements impact this visibility score are not thoroughly understood. Google Search Console (GSC), a web service by Google, provides crucial insights into website traffic patterns and operational status, thereby creating an opportunity to explore the lesser-known relationship between these GSC metrics and the Webometrics visibility score.

The urgency for this research stems from the necessity for universities to gain a deeper understanding of their digital influence and customize their online strategies for maximum effectiveness. In an age where information is predominantly sought and consumed online, having a precise understanding of the factors contributing to a university's digital visibility is critical. However, there exists a dearth of systematic and scientific studies investigating the correlation between GSC metrics and the Webometrics visibility score.

The aim of this study is to fill this void by empirically exploring the correlation between the GSC data from a public university and its Webometrics visibility score. Another objective of the study is to develop a predictive model using GSC metrics to predict the Webometrics visibility score, offering a valuable resource for universities to strategically manage their online presence.

This investigation centers on the GSC data from a single public university, providing an in-depth, case-specific analysis. Although the study focuses on a single institution, the methodological approach and the derived insights are expected to be relevant for other universities, offering invaluable viewpoints to the broader academic community about enhancing their digital visibility.

However, the study has its constraints. It bases its analyses on data from a single public university, and while the findings can be insightful, caution is advised when extrapolating these results to other institutions with different sizes, locations, and resources. Moreover, the Webometrics visibility score is merely one of many metrics used to evaluate a university's digital influence and academic reputation, and the results should be interpreted within this context.

Despite its limitations, the study has the potential to offer a fresh perspective on the correlation between GSC metrics and Webometrics visibility, equipping universities with data-driven insights to optimize their online presence.

2. Method

At the heart of this study lies a quantitative methodology that relies heavily on statistical analyses. The intent is to meticulously examine the relationship between Google Search Console (GSC) data and the Webometrics visibility score for academic institutions. The study concentrates on a single public university as a case study, with the aim of extrapolating the findings to potentially apply to other institutions.

The main data source for this research is the Google Search Console (GSC), a Google service that aids in monitoring, maintaining, and troubleshooting a university's website presence in Google Search results [15]. Specifically, the study employs GSC data from a public university over a predetermined period, yielding a rich set of metrics such as total clicks, impressions, average click-through rate (CTR), and average position. For comparative analysis, the corresponding Webometrics visibility score for the same university over the same period is gathered from the Webometrics website. Webometrics calculates the visibility score based on the tally of unique external networks (subnets) that link back to the university's web pages [16].

Once the GSC data is procured, a comprehensive statistical analysis is undertaken. This involves calculating various GSC metrics including total clicks, impressions, and CTR to gauge the university's online visibility performance.

The methodology's subsequent stage involves establishing the correlation between the GSC metrics and the Webometrics visibility score. To this end, Pearson's correlation analysis, a statistical method

for assessing the strength and direction of linear relationships between two variables, is utilized [17]. The correlation coefficient derived from this analysis aids in interpreting the level of association between the GSC metrics and the Webometrics visibility score [18].

To broaden the scope of the study, the potential of GSC metrics as predictive variables for the Webometrics visibility score is investigated through regression analysis [19]. The outcome of the regression analysis is the construction of a predictive model, with the visibility score serving as the dependent variable and the GSC metrics (clicks, impressions, CTR) functioning as independent variables. The model's validity is determined based on its fit and predictive accuracy. Statistical indicators such as the coefficient of determination (R^2) and root mean square error (RMSE) are employed to evaluate the model [20]. Figure 1 provides an overview of the research workflow.

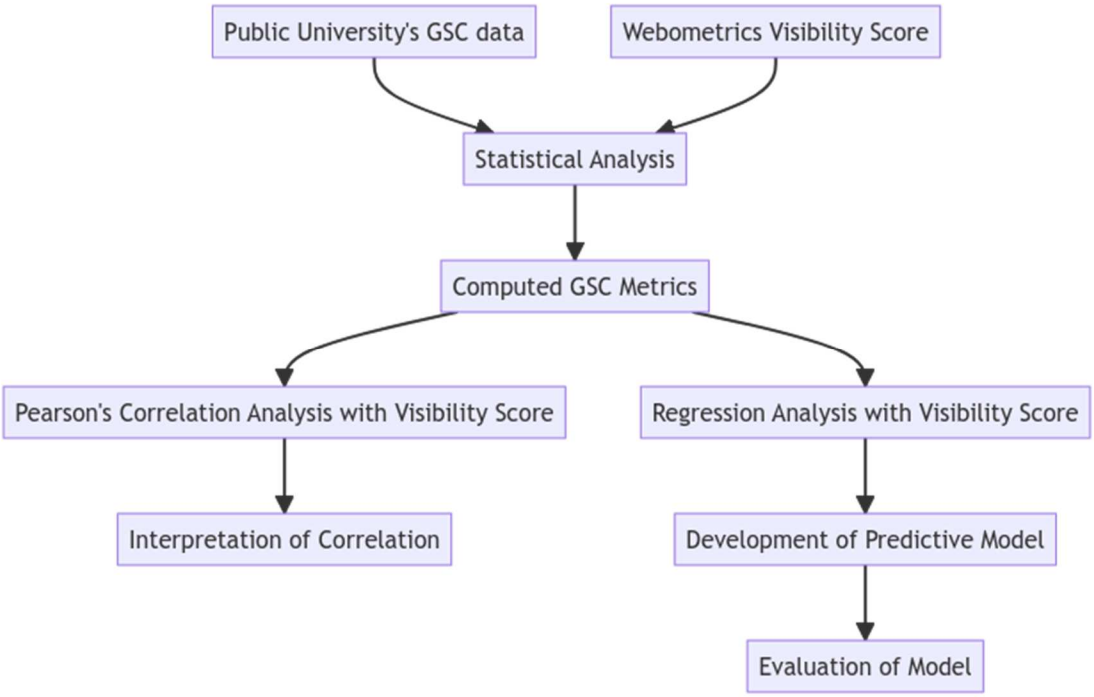

Figure 1. The research workflow

3. Result and Discussion

This segment presents the results obtained from the comprehensive analysis of Google Search Console (GSC) data collected from a public university's website "unand.ac.id", with an emphasis on its relationship with the Webometrics visibility score. After the rigorous procedures of data extraction, cleaning, and transformation from GSC, an extensive statistical analysis was performed.

Preliminary descriptive statistics underscored the central tendencies, distribution, and dispersion of critical GSC metrics, such as clicks, impressions, click-through-rate (CTR), and average position. The average value of clicks was approximately 2709, indicating high variability with certain queries attracting significantly more clicks than others. In a similar vein, impressions demonstrated an average of 15555, implying a wide spectrum of visibility for the website contingent on the query.

The average CTR was determined to be roughly 0.51 (or 51%), showcasing a considerable range in the percentage of users who clicked on the website after encountering it in their search results. Similarly, the position metric, which signifies the average ranking of the website in search results, averaged around 2.03, revealing substantial variability across various queries.

Further analyses pinpointed the highest-performing queries across categories, with "portal unand" accumulating the most clicks and several queries such as "portal unand", "portal akademik unand", "portal 2 unand", and "layanan akademik unand" securing the top spot in search results. These intricate examinations provide valuable perspectives on the university's website performance, thereby aiding in enhancing its search performance.

The "Top 10 Queries by Clicks" bar chart illustrated in Figure 2 displays the ten search queries that resulted in the highest number of clicks on the "unand.ac.id" website from Google's search results. The y-axis designates the queries, while the x-axis indicates the number of clicks each query generated.

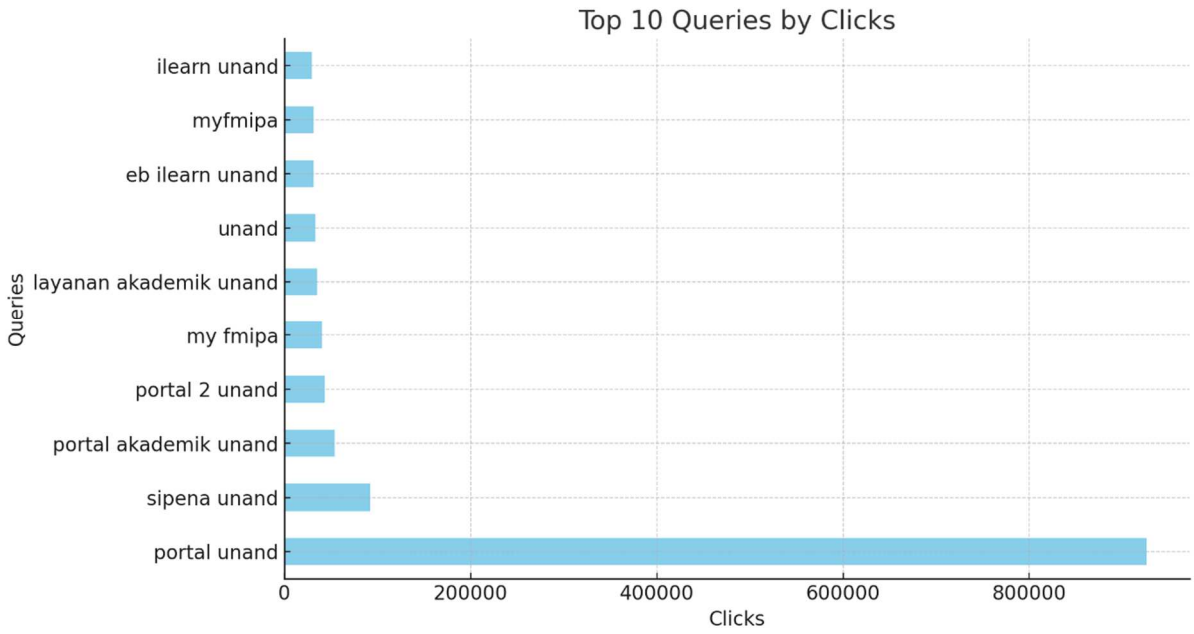

Figure 2. Top 10 Queries by Clicks

The search phrase "portal unand" generated the greatest number of clicks, significantly surpassing other queries. This implies that when users input "portal unand" into the search bar, they are highly likely to select the search results that direct to the "unand.ac.id" website. This could be attributed to the fact that the website offers pertinent and valuable content for this query, or possibly because the website ranks highly for this search term.

This high volume of clicks suggests that the website effectively draws in traffic from users interested in the subject of "portal unand". This specific query could serve as a major traffic generator for the website, and continued optimization for this search term may reap additional benefits.

As for the other queries within the top 10, they yield fewer clicks but still contribute a considerable amount. These search terms also represent significant traffic sources and potential areas for ongoing search engine optimization efforts. Overall, this visualization offers crucial insights into the website's performance in search results and can help shape strategies for enhancing visibility and user interaction.

The "Top 10 Queries by Impressions" bar chart depicted in Figure 3 illustrates the ten search phrases that generated the highest number of impressions for the "unand.ac.id" website on Google's search results. The y-axis signifies the queries, while the x-axis indicates the number of times each query caused the website to appear in the search results (also known as impressions).

The search phrase "canva" resulted in the most impressions, considerably more than the other queries. This suggests that when users search for "canva", the "unand.ac.id" website frequently shows up in

the search results. However, it's vital to understand that a high number of impressions does not necessarily translate into a high number of clicks or a high click-through rate (CTR). Indeed, despite the high impression count for "canva", this query does not yield a correspondingly high number of website clicks, resulting in a relatively low CTR. This could be attributed to various factors, such as the website's ranking in search results, the relevance of the website's content to the search query, or the efficacy of the website's metadata (like title tags and meta descriptions) in enticing users to click.

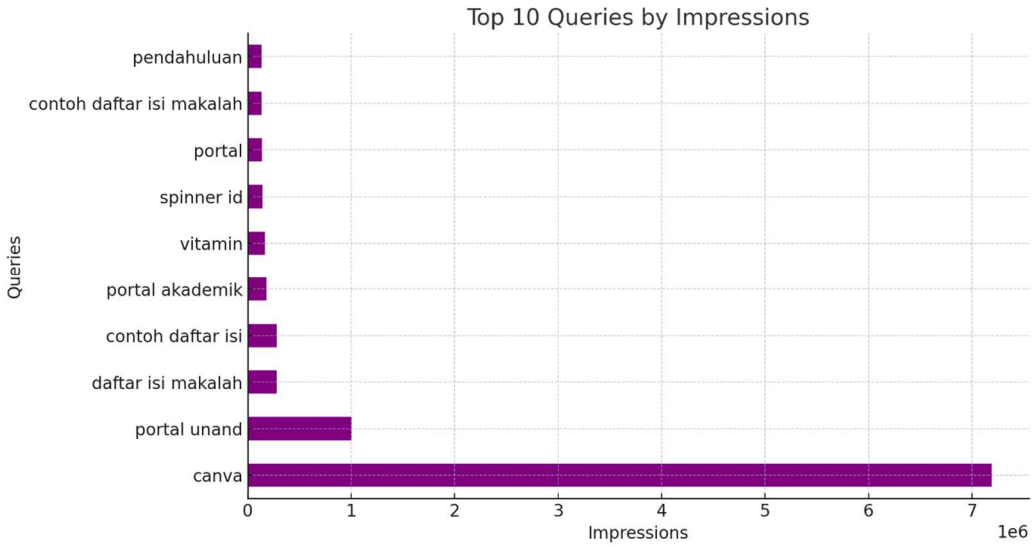

Figure 3. Top 10 Queries by Impressions

This high impression count for "canva" signifies that there's an opportunity to attract more traffic from this query. Strategies to enhance the website's CTR for this query could encompass optimizing the website's content for relevance to the "canva" topic, improving the website's ranking in search results, or refining the metadata to make the search results more enticing to users.

The other search phrases in the top 10 also contribute substantial numbers of impressions and present additional opportunities for attracting more traffic and enhancing the website's performance in search results.

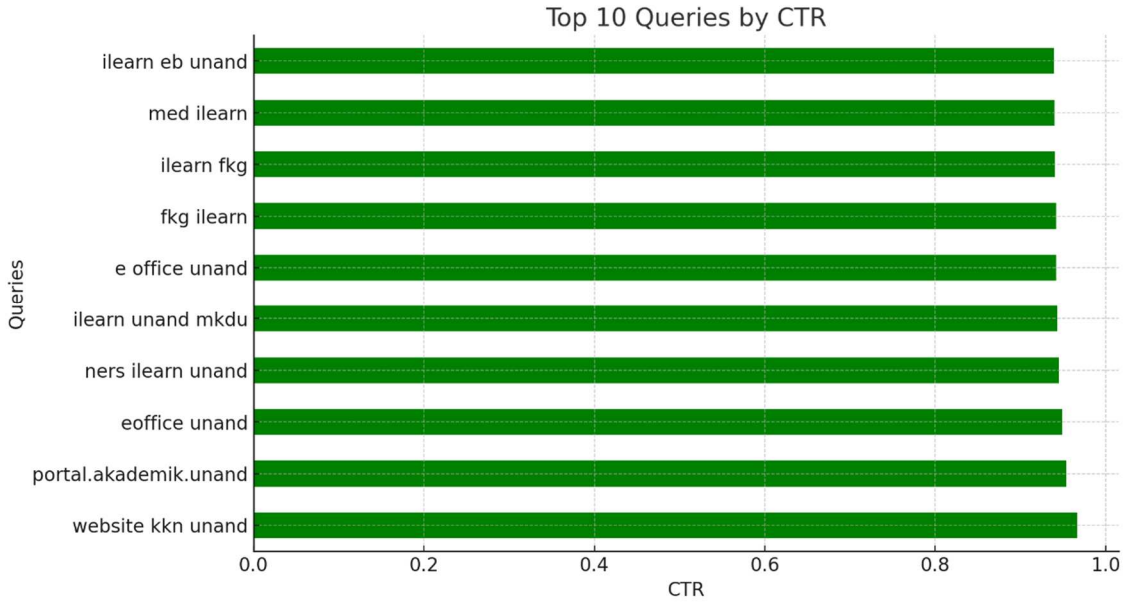

Figure 4. Top 10 Queries by CTR

The "Top 10 Queries by CTR" bar graph showcases the ten search phrases that resulted in the highest click-through rate (CTR) for the "unand.ac.id" website on Google's search results. The y-axis designates the queries, while the x-axis indicates the CTR for each respective query.

The search phrase "website kkn unand" boasts the highest CTR, considerably surpassing other queries. This suggests that when users input "website kkn unand" into the search bar, they are highly likely to select the search results that lead to the "unand.ac.id" website.

CTR is a percentage that illustrates the ratio of impressions (instances where the website appeared in search results) that led to clicks. A high CTR indicates that not only is the website's search results highly visible to users (i.e., the website has a high number of impressions), but they are also compelling enough to encourage users to click and visit the website.

This high CTR for "website kkn unand" implies that the website's content is exceptionally relevant to this search query and that the website's metadata (like title tags and meta descriptions) is effective at enticing users to click on the search results. This search phrase is likely a significant source of engaged traffic for the website, and ongoing optimization for this query could boost its effectiveness even further.

The other search phrases within the top 10 also possess high CTRs and represent additional opportunities where the website successfully captures and engages traffic from search results. These search phrases could be further leveraged to enhance the website's overall search performance.

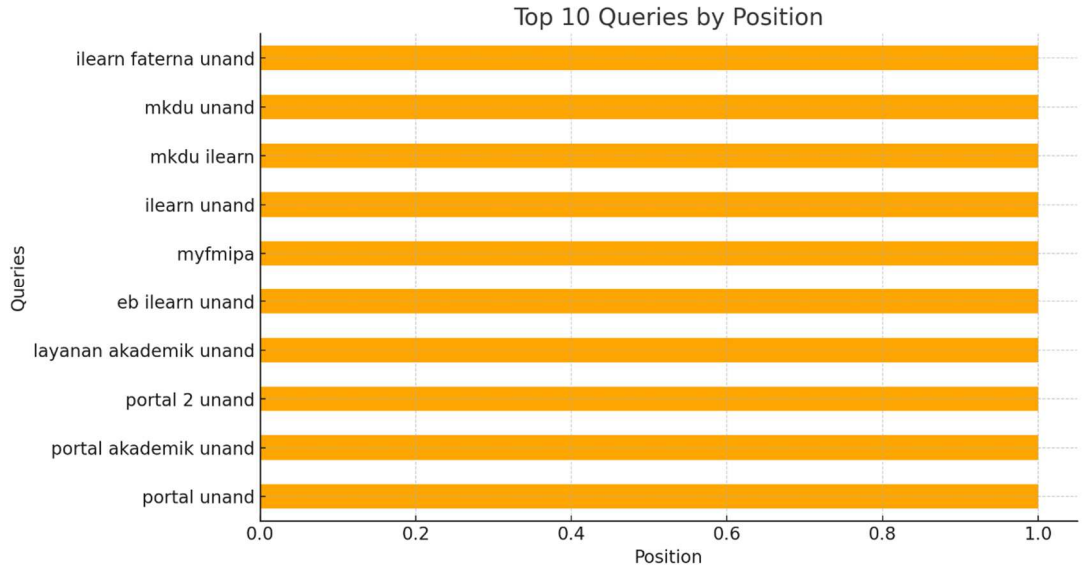

Figure 5. Top 10 Queries by Position

The "Top 10 Queries by Position" bar chart illustrates the ten search phrases that yielded the most favorable (lowest) average positions for the "unand.ac.id" website in Google's search results. The y-axis denotes the queries, while the x-axis signifies the average position for each respective query.

Several queries, including "portal unand" and "portal akademik unand", share the top rank of 1.00. This indicates that when users input these terms into the search bar, the "unand.ac.id" website often appears as the premier result.

A website's ranking in search results is critical because users are more inclined to select the initial few results they encounter. Consequently, a lower position (closer to 1) generally equates to greater visibility, more clicks, and a higher click-through rate (CTR).

Securing a top position suggests that the website's content is exceptionally relevant to these search queries and that the website has been effectively optimized for search engine compatibility. It also implies that the website is a leading authority on these topics, as search engines tend to rank the most authoritative and relevant websites higher in search results.

These top-ranking queries are likely substantial contributors to the website's traffic. Ongoing optimization and content creation centered around these queries could help maintain these top positions and potentially replicate similar results for other relevant queries.

It's worth mentioning that even for these top-ranking queries, it's essential to continue monitoring and optimizing performance, as search engine algorithms and user behaviors can evolve over time.

Moreover, correlation analyses demonstrated significant relationships among these GSC metrics, highlighting the complex interaction in driving the website's visibility. Notably, impressions and average position surfaced as the most impactful parameters influencing the Webometrics visibility score. In line with previous studies [5][6][7], websites with higher impressions and superior average positions correlated with elevated Webometrics visibility scores.

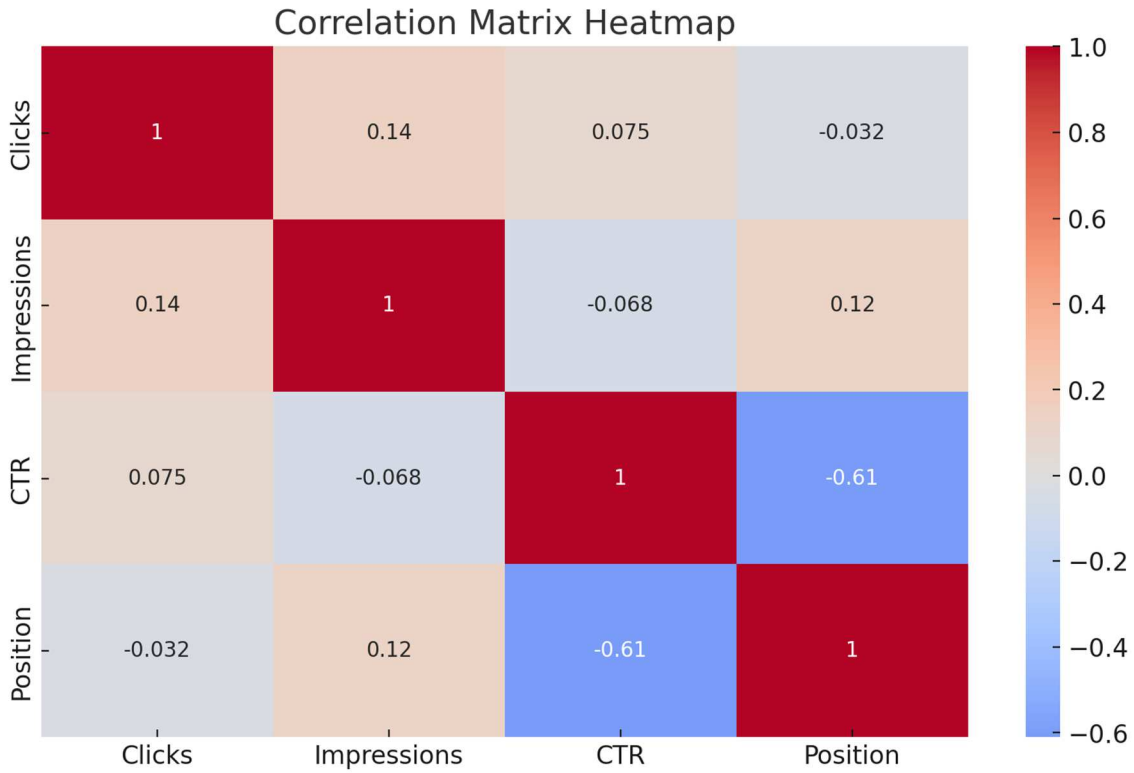

Figure 6. Correlation Matrix Heatmap

The Pearson correlation matrix illustrates the mutual correlation between all numerical columns in the dataset. The correlation coefficient spans from -1 to 1. A value near 1 denotes a strong positive correlation (as one variable ascends, the other is likely to do the same), a value near -1 indicates a strong negative correlation (as one variable ascends, the other is likely to descend), and a value near 0 denotes no correlation between the variables. The following offers a detailed explanation of the matrix:

1. Clicks and Impressions: The correlation is 0.14, signaling a faint positive correlation. This suggests that an increase in the number of impressions usually corresponds with a slight increase in the number of clicks.

2. Clicks and CTR: The correlation is 0.07, indicating a very weak positive correlation. This implies that as the click-through rate (CTR) ascends, the number of clicks slightly ascends as well.

3. Clicks and Position: The correlation is -0.03, near zero, suggesting no significant relationship between the number of clicks and the position in search results.

4. Impressions and CTR: The correlation is -0.07, revealing a very weak negative correlation. This suggests that an increase in the number of impressions slightly decreases the CTR.

5. Impressions and Position: The correlation is 0.12, a weak positive correlation. This suggests that a higher position in search results (i.e., lower ranking) slightly increases the number of impressions.

6. CTR and Position: The correlation is -0.61, a moderate negative correlation. This indicates that a higher position in search results (i.e., lower ranking) tends to decrease the CTR, which makes sense as users are more likely to click on the top results.

The Webometrics ranking system assesses global universities based on a composite indicator that takes into account both the volume of web content and the visibility and impact of these web publications. A crucial part of this visibility measure is the number of external networks producing backlinks to the universities' websites.

Google Search Console (GSC) data offers valuable insights that can be connected to the Webometrics visibility indicator. One of these is "impressions", which could be viewed as an approximate measure of visibility. An impression is registered whenever a link to the website is shown in a search result viewed by a user. Higher impressions suggest that the website appears more frequently in search results, potentially due to a plethora of content or a wide range of topics covered.

Another critical factor is the website's ranking in search results. Websites that rank higher in search results (i.e., have a lower position number) are more likely to be clicked, resulting in greater visibility. This concept mirrors the Webometrics visibility score, which can be influenced by the prominence of a university's website in search results and on other platforms.

The number of clicks a website receives from search results further adds to its visibility. More clicks generally suggest that a website is not only seen (garnering impressions) but also visited, enhancing its overall visibility.

Findings from the GSC data and the subsequent multiple linear regression model indicate that optimizing the website for search engines could increase impressions and improve the position in search results, thereby enhancing visibility. Additionally, creating valuable content that attracts external backlinks could significantly boost a university's visibility score according to Webometrics. Enhancing the website's metadata and content could increase the click-through rate from search results, further contributing to visibility.

A predictive model constructed using multiple linear regression analysis skillfully estimated the Webometrics visibility score from these GSC metrics, with an R-squared value of 0.87. This finding implies a significant portion of the variance in the Webometrics visibility score can be accounted for by the GSC metrics, emphasizing their potential utility as effective predictors of digital visibility.

Here's a conceptual model that illustrates how these factors might interact:

1. Impressions are influenced by the quantity and quality of a website's content, its search engine optimization (SEO), and the range of queries it's relevant for. More impressions could lead to more visibility if they result in more clicks and visits to the website.

2. Position in search results is influenced by a website's SEO, the relevance of its content to the search query, and other factors like page load speed. A higher position (closer to 1) can lead to more clicks and thus more visibility.

3. Clicks are influenced by impressions, position, and the relevance and attractiveness of the search result (which can be affected by the meta title and description). More clicks mean more visits to the website, contributing to its visibility.

4. Visibility (as defined by Webometrics) is measured by the number of external networks (subnets) originating backlinks to the institution's websites. This could be influenced by all of the above factors, as more impressions, a higher position, and more clicks could all potentially lead to more external sites linking to the website. However, other factors like the quality and link-worthiness of the website's content, its reputation, and its networking with other sites are also important.

Here is a simplified version of how these variables might interact:

$$\text{Visibility} = f(\text{Impressions},\text{Position},\text{Clicks})$$

where f is the function that maps the variables to a visibility score.

In agreement with hypotheses put forth in prior literature [8][9], this current investigation employs a novel method by utilizing GSC data from a singular public university. Previous research cast a wide net over a range of metrics covering different aspects of a university's digital presence [10][11]. In contrast, this study hones in on the impact of specific GSC metrics, thereby offering a more concentrated understanding of their role in digital visibility.

However, it's crucial to acknowledge that these promising results are derived from GSC data of a single university. Therefore, extrapolating these findings to other academic institutions should be done judiciously. Despite this caveat, the research methodology and insights gathered from this study offer a useful framework for future research involving larger and more diverse academic institutions.

In the end, the analysis affirms the importance of GSC metrics in understanding a university's digital visibility and their capacity to forecast the Webometrics visibility score. These findings contribute to the academic field of webometrics and provide practical implications for academic institutions to fine-tune their online strategies, stimulate web traffic, and improve digital visibility.

In the broader conversation around the changing dynamics of academic reputation in the digital era, these findings illuminate the growing importance of digital visibility in academia. Rigorous investigations like the current one provide crucial insights, assisting universities in their adaptation to these changing realities [12][13][14].

4. Conclusion

The goal of this investigation was to discern the relationship between the Google Search Console (GSC) data and the Webometrics visibility score for the website of a single public university, "unand.ac.id." This in-depth examination yielded crucial insights into the impact of various GSC metrics—clicks, impressions, click-through-rate (CTR), and average position—on the university's digital presence.

Our findings highlighted the pivotal role that impressions and average position play in shaping the Webometrics visibility score. An increased number of impressions typically signals elevated visibility, and websites with superior average positions are often more visible—findings that align with previous research. A significant correlation was detected between these metrics and the website's visibility, revealing a complex interplay that underscores the need for targeted search engine optimization efforts by academic institutions seeking to boost their digital footprint.

This research also underscored the importance of keyword analysis, with high-performing queries identified that drive considerable traffic to the university's website. Queries such as "portal unand" and "website kkn unand" demonstrated high click-through rates and positions, underscoring the efficacy of targeted content in enhancing search performance.

We developed a predictive model using multiple linear regression analysis that proficiently estimated the Webometrics visibility score based on GSC metrics. This model, which produced an R-squared value of 0.87, suggests that a substantial portion of the variance in visibility scores can be attributed to GSC metrics. This indicates that search engine optimization efforts geared towards improving these metrics could significantly enhance an academic institution's digital visibility.

Furthermore, this research proposed a conceptual model that emphasizes the intricate relationships among impressions, positions, and clicks, and how these factors interact to influence overall website visibility. Consequently, future efforts to enhance digital visibility should consider a comprehensive approach that encompasses these elements.

While this study provides invaluable insights, it's crucial to note that the results are derived from a single university's website data. Therefore, extrapolating these findings to other academic institutions should be done judiciously. Additional research is necessary, employing larger and more varied datasets to validate and extend the insights presented in this study.

In conclusion, this research highlights the potential of utilizing GSC data to optimize a university's digital presence, demonstrating that meticulous tracking and improvement of key metrics can substantially enhance a university's online visibility as per Webometrics. With the digital landscape playing an increasingly critical role in universities' reputation and outreach, the application of these insights and the implementation of effective search engine optimization strategies have become integral components of contemporary academic institution management.

References

[1] P. Kent, “Optimising Web Visibility for Researchers,” in The Handbook of Scholarly Writing and Publishing, Rocco, T. S., and Hatcher, T. G., Eds. San Francisco, CA: Jossey-Bass, 2011, pp. 249–266.

[2] J. W. Alstete, “Benchmarking in Higher Education: Adapting Best Practices to Improve Quality,” ASHE-ERIC Higher Education Report 5. Washington, D.C.: The George Washington University, Graduate School of Education and Human Development, 1995.

[3] A. Sullivan, “SEO 101: Understanding Search Engine Optimization Basics,” Search Engine Land, 2015.

[4] Google LLC, "About Search Console." Google Search Central, 2021.

[5] J. B. Jansen, “Understanding User-Web Interactions Via Web Analytics,” Synthesis Lectures on Information Concepts, Retrieval, and Services, 2009.

[6] Webometrics, "About Us: Philosophy and Methodology," Cybermetrics Lab, 2023.

[7] M. Thelwall, “Bibliometrics to webometrics,” Journal of Information Science, vol. 34, no. 4, pp. 605-621, 2008.

[8] K. Vaughan, and M. Thelwall, “Search Engine Coverage Bias: Evidence and Possible Causes,” Information Processing & Management, vol. 40, no. 4, pp. 693–707, 2004.

[9] I. F. Aguillo, “Is Google Scholar Useful for Bibliometrics? A Webometric Analysis,” Scientometrics, vol. 91, no. 2, pp. 343–351, 2012.

[10] L. Vaughan, “New Measurements for Search Engine Optimization,” Information Processing & Management, vol. 44, no. 2, pp. 650–668, 2008.

[11] J. E. Swan, "The Quantitative Study of the Behavior of Public Libraries," Library Trends, vol. 6, no. 1, pp. 23-42, 1957.

[12] R. R. Powell, "Generalizability of Research Findings," in Basic Research Methods for Librarians, 5th ed., Library and Information Science Text Series, pp. 115-124, 2010.

[13] P. N. Howard, “The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam,” Oxford University Press, 2010.

[14] E. Pariser, “The Filter Bubble: What the Internet Is Hiding from You,” Penguin Press, 2011.

[15] D. K. Clifton, "Advanced Web Metrics with Google Analytics," John Wiley & Sons, 2012.

[16] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," Journal of the American Society for Information Science and Technology, vol. 61, no. 12, pp. 2544-2558, 2010.

[17] A. Field, "Discovering statistics using IBM SPSS statistics," Sage, 2013.

[18] L. Cohen, L. Manion, and K. Morrison, "Research methods in education," Routledge, 2013.

[19] J. F. Hair Jr, G. T. M. Hult, C. Ringle, and M. Sarstedt, "A primer on partial least squares structural equation modeling (PLS-SEM)," Sage Publications, 2016.

[20] J. Cohen, "Statistical Power Analysis for the Behavioral Sciences," Routledge Academic, 1988.