Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method

Ibrahim*[1], Sri Wahyuni[2]

[1,2]Program Studi Magister Teknologi Informasi, UniversitasPembangunan Pancabudi, Medan, Indonesia

Article Info	ABSTRACT
Corresponding Author: Ibrahim E-mail: ibrahim31@guru.smk.belajar.id	This study aims to analyze the effect of social media usage and sleep duration on students' Grade Point Average (GPA) using the clustering method in data mining. Social media has become an integral part of students' lives, but excessive use can have a negative impact on sleep time and, indirectly, on academic achievement. Using the clustering method, this study groups students based on their social media usage patterns, sleep hours, and GPA to identify groups with certain characteristics. Data collected from 520 students were analyzed using the K-Means clustering algorithm, which resulted in three main groups: a group with high social media usage and low GPA, a group with a balanced sleep pattern and moderate GPA, and a group with adequate sleep time and high GPA. The results of the analysis showed that students with high social media usage tend to have lower sleep hours and lower GPA than students who have sufficient sleep duration. This study is expected to be a basis for campuses to develop programs to improve student welfare, especially in regulating social media usage and improving sleep quality. The data used in this study uses data from the Kaggle.com platform which provides various types of data worldwide. This research is expected to provide insight for students, lecturers and people with regard to this method. **Keywords:** Data Mining, Clustering, Social Media, Sleep Hours, GPA, K-Means, Students.

INTRODUCTION

The use of social media has become an integral part of everyday life, including among college students. Social media offers convenience in communicating, sharing information, and entertainment. However, along with the increasing use of social media, concerns have arisen that the time spent on these platforms may affect various aspects of college students' lives, especially sleep patterns and academic performance. Several studies have shown that excessive use of social media can result in decreased sleep quality and, ultimately, impact college students' academic performance [1] [5]. Adequate sleep is essential for mental and physical health, and plays an important role in the learning process and memory consolidation. College students who have sufficient sleep hours tend to have better academic performance than those who lack sleep [12]. On the other hand, lack of sleep due to excessive use of social media can have a negative impact on concentration, cognitive

*Data Mining Analysis of the Influence of Social Media on Students'
Sleep Hours and GPA Using the Cluster Method- Ibrahim et.al*

155

abilities, and understanding of lecture material [2]. Cumulative Grade Point Average (GPA) is one indicator of academic achievement that shows the level of understanding and success of students in the learning process. Factors such as social media use and sleep patterns are believed to have an influence on college students' GPA. However, there has not been much research that utilizes data mining techniques, especially clustering methods, to understand the patterns and relationships between these factors and college students' GPA. Data mining is a data analysis method that allows the identification of hidden patterns in large data sets. One of the widely used data mining techniques is clustering, which aims to group data based on certain similar characteristics [4]. In this context, the K-Means Clustering method can be applied to group students based on the duration of their social media use, their sleep hours, and their GPA, so that clear patterns and relationships can be obtained between the three variables.

This study aims to analyze the effect of social media usage and sleep duration on students' GPA using the clustering method. Through this grouping, it is expected to find groups of students with certain characteristics, such as students who have high social media usage and low sleep duration, and how this relates to their academic achievement. By understanding this pattern, educational institutions are expected to develop programs and strategies that support students' lifestyle balance, which can ultimately contribute to improving academic achievement. Based on this background, this study will discuss the following questions: 1. How do social media usage and sleep duration affect students' GPA? 2. What is the pattern of student grouping based on social media usage, sleep duration, and GPA? This study is expected to contribute to the field of education, especially in understanding the factors that influence students' academic achievement, as well as providing recommendations regarding effective time management to improve their well-being and academic performance.

To overcome these problems, one approach that can be considered is the application of data mining techniques, especially in classifying social media user behavior. Data mining allows organizations to extract valuable information from large data sets, identify patterns, and make predictions about future social media user behavior. The K Mean Clustering Algorithm method, which is one of the most popular decision tree algorithms, can be used to classify social media user data based on certain features, such as frequency of use of social media applications, types of social media, and demographics of social media users. By using this method, the world of education can more easily group social media users, balanced sleep patterns and student GPAs more optimally and effectively.

Research conducted by Ginting,.& Hidayat,DR (2021) with the title Analysis of Social Media Usage on Student Academic Achievement Using the K-Means Algorithm Method [8]. Research conducted by. Nasir, M, & Abdurrahaman, A. (2020) with the title Research on the Influence of Social Media on Student Academic Productivity [15]. Research conducted by Murniati, E. (2022) with the title Data Mining in Education: Understanding Student Performance and Behavior Using K-Means Clustering.[14].

Thus In this study, the author chose to use the C4.5 method as the main solution for Thus In this study, the author chose to use the Cluster method as the main solution to classify social media users towards academic achievement. This method was chosen because of its ability to handle incomplete data and produce easy-to-understand classification rules. In addition, K-Means Cluster also has advantages in terms of efficiency

Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- **Ibrahim et.al**

156

and accuracy in data classification. By applying this method, it is expected to obtain deeper insights into social media usage patterns and social media user behavior towards academic achievement.,

RESEARCH METHODOLOGY

The research stage flow is a series of systematic steps that must be taken by a researcher in conducting research, starting from data collection, data preprocessing, clustering modeling, and analysis of results. This flow ensures that the research is carried out in a structured and directed manner, so that the data and findings obtained are reliable and meet the research objectives. The following is an overview of the research stage flow:
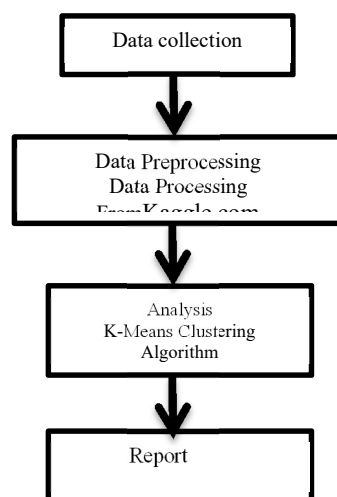

Figure 1. Research Stages

This image is a research flow diagram consisting of several main stages.

1. Data collection

   Data were collected through surveys and questionnaires distributed to students at several universities. The questionnaires included information on: Duration of social media use (per day), Daily sleep duration (hours of sleep per day), GPA as an indicator of academic achievement The sample taken was 200 students who were randomly selected to obtain sufficient data representation.

2. Data Preprocessing

   The preprocessing stage is carried out to ensure that the data is clean and ready for analysis. The steps in data preprocessing include: Data Cleaning: Eliminating incomplete or invalid data. Data Normalization: Converting data into a uniform scale so that the clustering algorithm can work more optimally. Data Encoding: If there is categorical data, then coding is carried out to facilitate analysis.

3. Clustering Modeling

   The K-Means Clustering algorithm is used to group the data. This algorithm was chosen because of its ability to form groups that are easy to interpret. The steps in clustering modeling include: Determining the optimal number of clusters (k) using the elbow or silhouette score method. Grouping data into k groups based on similar characteristics (duration of social media use, hours of sleep, and GPA). Generating centroids for each cluster as the center of each group.

4. Algorithm Result Analysis

*Data Mining Analysis of the Influence of Social Media on Students'
Sleep Hours and GPA Using the Cluster Method-* **Ibrahim et.al**

157

After cluster formation, interpretation and analysis of the results are carried out to understand the characteristics of each group of students. This analysis includes: Identifying patterns in each cluster, such as groups of students with high social media duration but low GPA. Looking at the relationship between variables of social media use and sleep hours to GPA. Providing recommendations based on the patterns found, especially related to the balance of social media use and sleep time.

5. Report

The final stage is the preparation of a report that summarizes the results of the research, including the findings, analysis, and conclusions of the research process. This report serves as the final documentation that can be shared or published.

This flowchart shows the research steps systematically from problem identification to the preparation of the final report.

Data Mining

Data mining is a procedure used for processing information to create hidden patterns from the information being processed. Information processed using this information mining method then creates new knowledge that comes from old information and the results of processing the information can be used to determine a decision in the future. Information mining is also a series of activities to create interesting patterns from large amounts of information, after which the information can be placed in a database, information warehouse or data storage [6]-[13]-[14]-[18]-[20]

Classification

Classification is a way of grouping objects based on the characteristics possessed by the classification object. In the process, classification can be done in many ways, either manually or with the help of technology. Classification that is done manually is a classification that is done by humans without the help of intelligent computer algorithms. While classification that is done with the help of technology has several algorithms, including K-Means, Support Vector Machine, [4], [8], [14][20]

K-Means Algorithm

The K-Means algorithm is a clustering method that groups data into a number of groups (clusters) based on similar characteristics. In K-Means, the desired number of groups (k) is determined in advance. The process includes several main steps: Determine k initial center points (centroids) randomly. Group each data to the nearest centroid based on distance (usually using Euclidean distance). Update the centroid position by calculating the average position of the data in each cluster. Repeat the process of clustering and updating centroids until the centroid position is stable or there is no significant change.[6],[8],[10] Sample data after pre-processing is as follows:

Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- **Ibrahim et.al**

158

Table 1.sample data taken from kagel.com

NO	Name	Gender	Semester Level)	Study Program/Major)	How long do you use Social Media in a day?	How long do you sleep a day?	Do you feel that your academic achievement (GPA) is good?	Can the use of Social Media affect your academic performance?	Which Social Media Platform do you use the most?	Does Social Media usage affect your sleep hours?
1	Siti Rahmah Basri	Woman	3	Information Technology	More than 2 hours	Less than 6 hours	No	Yes	TikTok	Yes
2	Nadya Elfareta Azarin	Woman	3	Information Technology	More than 2 hours	Less than 6 hours	No	Yes	TikTok	Yes
3	This is Wayan Erdiani	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	TikTok	Yes
4	English	Woman	3	Information technology	More than 6 hours	More than 8 hours	Yes	Yes	Tiktok	Yes
5	Dinda	Woman	3	Information Technology	Less than 6 hours	Less than 8 hours	Yes	Yes	Tiktok	Yes
6	Nadya Elfareta Azarin	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	Tiktok	Yes
7	Augustine	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	Tiktok	Yes
8	Rangga	Man	3	Information Technology	Less than 6 hours	Less than 8 hours	No	Yes	WhatsApp	Yes
9	Arya Dwinata Mustapa	Man	3	Information technology	More than 6 hours	Less than 8 hours	Yes	Yes	Tiktok and WhatsApp	Yes
10	Alfi Zahrah Muharramah	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	Tiktok	Yes
11	Jalaludin Muflih	Man	3	Information Technology	More than 6 hours	Less than 8 hours	No	Yes	Tiktok	Yes
	Wiranago	Man	3	Civil Engineering	More than 6	More than	Yes	No	Facebook and	Yes

*Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- **Ibrahim et.al***

159

No	Name	Gender		Major	Sleep 1	Sleep 2			Social Media	
12					hours	8 hours			Twitter	
13	Anawula Raise the Heights	Woman	5	Information Technology	More than 6 hours	More than 8 hours	No	No	Instagram	No
14	The story of the greatness of Helmi	Woman	3	Information Technology	More than 6 hours	More than 8 hours	No	Yes	Tiktok	No
15	Yoga	Man	5	Civil Engineering	More than 6 hours	Less than 8 hours	No	No	Wow, tiktok	No
16	Alfredo Rivaldo	Man	3	Civil Engineering	More than 6 hours	Less than 8 hours	No	Yes	WhatsApp	Yes
17	Fikry Haykhal	Man	3	Information Technology	More than 6 hours	Less than 8 hours	No	No	Whatsapp	Yes
18	Ahmad the Great	Man	3	Information technology	More than 6 hours	Less than 8 hours	No	Yes	Facebook	Yes
19	Endro	Man	3	Information Technology	Less than 6 hours	Less than 8 hours	Yes	Yes	Tiktok	No
20	Saudi Arabian Bayu	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	ig, tiktok and wa	Yes
21	what is ode safira al zahra A	Woman	3	D3 Civil Engineering	More than 6 hours	Less than 8 hours	Yes	Yes	WA	Yes
22	Angga Wahyu Prasetyo	Man	5	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	Tiktok	Yes
23	afternoon	Man	3	Civil Engineering	More than 6 hours	Less than 8 hours	Yes	No	tiktok	Yes
24	Siti Rahmah Basri	Woman	3	Information Technology	Less than 6 hours	Less than 8 hours	Yes	Yes	Whatsapp	Yes
25	faith	Man	3	Information Technology	More than 6 hours	Less than 8 hours	No	Yes	whatsapp and instagram	Yes
26	Trias	Man	3	D3 Electronic Engineering	More than 6 hours	Less than 8 hours	Yes	Yes	Tiktok	Yes

Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- Ibrahim et.al

160

27	English: Annisa Auliya Ramadhani	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	Whatsapp	No
28	Irfan	Man	7	Information Technology	Less than 6 hours	Less than 8 hours	Yes	No	Instagram	Yes
29	The Light of Allah	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	WhatsApp	Yes
30	Wild	Woman	3	Information Technology	Less than 6 hours	More than 8 hours	Yes	Yes	WhatsApp	No
31	Diamond	Woman	5	Architectural Engineering	Less than 6 hours	Less than 8 hours	Yes	Yes	Instagram	Yes
32	Harnelia	Woman	3	Information Technology	Less than 6 hours	Less than 8 hours	No	No	Whatsapp	No
33	Mahdi Markun	Man	3	Information Technology	Less than 6 hours	More than 8 hours	Yes	Yes	WhatsApp	No
34	Al Gustaf	Man	3	D3 Electronic Engineering	Less than 6 hours	Less than 8 hours	Yes	No	Instagram	No
35	Nur Chaya	Woman	1	Information Technology	Less than 6 hours	Less than 8 hours	Yes	No	Instagram	Yes
36	Muhammad Asrain	Man	3	Architectural Engineering	More than 6 hours	More than 8 hours	Yes	Yes	Whatsapp	No
37	Annisa Ramadani	Woman	3	Public Health	More than 6 hours	More than 8 hours	Yes	No	Instagram	No
38	Rahma Damayanti	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	No	Twitter	Yes
39	The Light of Allah	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	Yes	WhatsApp	Yes
40	Wild	Woman	3	Information Technology	Less than 6 hours	More than 8 hours	Yes	Yes	WhatsApp	No
41	Diamond	Woman	5	Architectural Engineering	Less than 6 hours	Less than 8 hours	Yes	Yes	Instagram	Yes

*Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- **Ibrahim et.al***

42	Mursawal	Man	3	Information Technology	Less than 6 hours	Less than 8 hours	No	Yes	Instagram	Yes
43	Muh Rusdiansyah	Man	6	Law	More than 6 hours	More than 8 hours	Yes	Yes	Tiktok ig wa	No
44	God's revelation	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	No	Yes	Instagram	Yes
45	Natalia Devi	Woman	3	Information Technology	More than 6 hours	Less than 8 hours	No	Yes	Instagram	Yes
46	Reva	Woman	3	informatics engineering	More than 6 hours	Less than 8 hours	No	Yes	tiktok	Yes
47	Amar Adi	Man	3	Information Technology	More than 6 hours	Less than 8 hours	Yes	No	Facebook, Instagram, YouTube	Yes
48	Wisdom	Woman	3	Information technology	More than 6 hours	Less than 8 hours	No	Yes	Tik Tok	Yes
.....	.....	.....	.....	.....	.....	.....	.....	...	.....	.....
520	Vira Febriyanti	Woman	3	Counseling guidance	More than 6 hours	Less than 8 hours	Yes	No	instagram	Yes

Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- Ibrahim et.al

162

RESULTS AND DISCUSSION

To obtain decision tree rules in the K-Means algorithm, the steps taken are:

a. Selecting an attribute as the root. A root is obtained from the highest gain value of the existing attributes.

b. Create a branch for each value,

c. Divide the cases into branches, and

d. Repeat the process for each branch until all cases on the branch have the same class.

In order to be processed in Rapidminer Software, the previously processed data is first saved in Microsoft Excel file format with the extension *.csv.

a. Cell Selection Sessionto be imported (Stage 2) and variable selection as shown in Figures 3 and 4 below.

Figure 2. Selecting Cells to Import

Figure 3 Data Statistics

*Data Mining Analysis of the Influence of Social Media on Students'
Sleep Hours and GPA Using the Cluster Method-* **Ibrahim et.al**

163

The operator validation page appears, which is divided into training and testing areas, as shown in Figure 4.
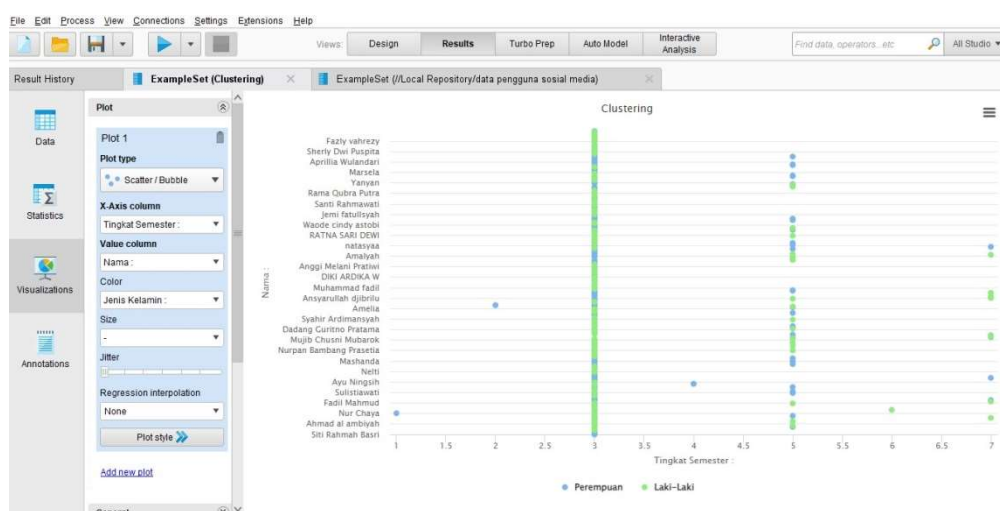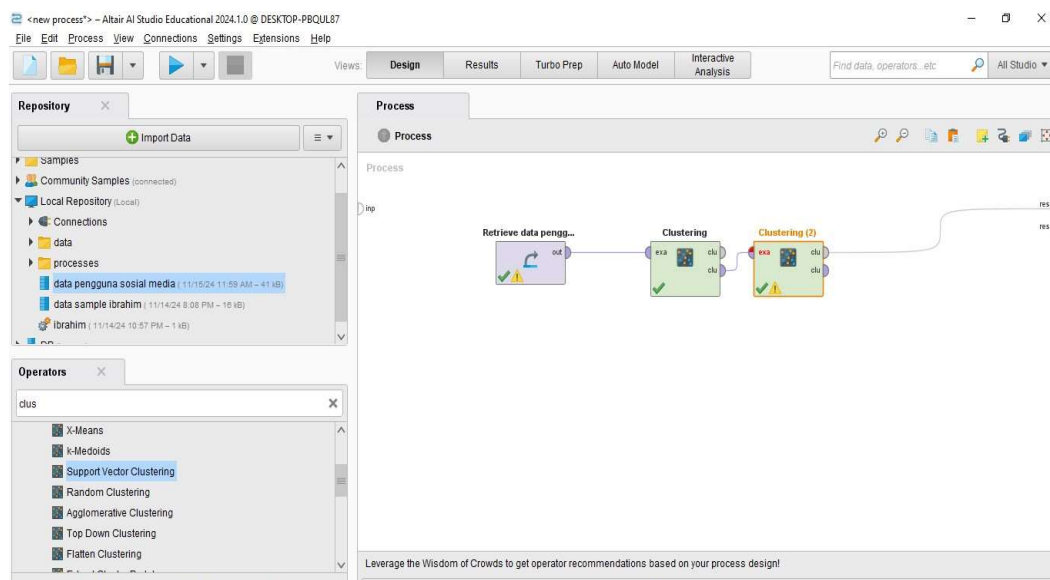

Figure 5 Cluster Decision Results in the Result view

CONCLUSION

From the graph obtained in the test with RapidminerStudio software, this study successfully identified the pattern of influence of social media usage and sleep duration on students' Cumulative Achievement Index (GPA) using the K-Means Clustering method. Based on the results of the analysis, three groups of students with different characteristics were obtained: Students with high social media usage and low sleep duration tend to have lower GPA. Students with balanced sleep patterns and moderate social media usage tend to have moderate GPA. Students with sufficient sleep duration and low social media usage show higher GPA.

Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- **Ibrahim et.al**

164

REFERENCES

[1]   Aji, H. M., & Widyastuti, S. (2020). The Influence of Social Media Usage on Student's Academic Performance. *Journal of Educational Research*, 15(2), 123-135.

[2]   Arora, S., & Agarwal, A. (2018). Impact of Sleep Deprivation on Academic Performance of College Students. *International Journal of Health Sciences*, 9(1), 56-65.

[3]   Badan Pusat Statistik (BPS). (2021). Statistik Media Sosial Indonesia. Jakarta: BPS.

[4]   Chen, L., Liu, W., & Zeng, Y. (2020). Application of K-Means Clustering in Educational Data Mining. *Procedia Computer Science*, 176, 134-142.

[5]   Chukwuere, J. E., & Chukwuere, P. C. (2017). The Impact of Social Media on Social Lifestyle: A Case Study of University Female Students. *Gender and Behavior*, 15(4), 9966-9981.

[6]   Dewi, R. N., & Kurniawan, S. (2019). Data Mining for Clustering University Students' Performance Using K-Means Algorithm. *Journal of Informatics and Data Mining*, 10(2), 67-76.

[7]   Ellis, R. A., & Goodyear, P. (2019). *Student Experiences of E-learning in Higher Education: The Ecology of Sustainable Innovation*. London: Routledge.

[8]   Ginting, L., & Hidayat, D. R. (2021). Analisis Penggunaan Media Sosial terhadap Prestasi Akademik Mahasiswa Menggunakan Algoritma K-Means. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(3), 197-206.

[9]   Hedges, D. V., & Tindall, J. R. (2019). Social Media, Sleep, and Academic Achievement: A Study among University Students. *American Journal of Social Science Research*, 6(1), 45-53.

[10] Huang, Y., & Zhang, Y. (2018). K-Means Clustering for Predicting Student Performance. *International Journal of Educational Research*, 89, 117-125.

[11] Indriani, L., & Firmansyah, E. (2020). Pengaruh Penggunaan Media Sosial terhadap Tingkat Prestasi Belajar Mahasiswa di Era Digital. *Jurnal Ilmu Pendidikan*, 27(2), 84-92.

[12] Li, Y., & Qi, F. (2019). The Role of Sleep in Learning and Memory Consolidation. *Journal of Neuroscience Education*, 12(3), 145-157.

[13] Maulana, R., & Putra, F. (2021). Educational Data Mining for Identifying Academic Performance Factors among Students Using Clustering Methods. *Journal of Data Science*, 19(2), 102-114.

[14] Murniati, E. (2022). Data Mining in Education: Understanding Student Performance and Behavior Using K-Means Clustering. *Journal of Information Technology and Education*, 17(1), 23-31.

[15] Nasir, M., & Abdurrahman, A. (2020). Pengaruh Media Sosial terhadap Produktivitas Akademik Mahasiswa. *Jurnal Sains Komunikasi*, 3(1), 78-89.

[16] Pasaribu, S. A., & Simatupang, P. D. (2021). The Effect of Sleep Patterns on Academic Success: A Clustering Approach. *Journal of Health and Well-being Studies*, 13(4), 78-89.

[17] Salim, A., & Purwanto, E. (2018). Factors Influencing Academic Performance among University Students in Indonesia. *Indonesian Journal of Higher Education*, 14(2), 231-246.

*Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- **Ibrahim et.al***

165

[18]   Santoso, Y., & Kurniawan, R. (2019). Clustering Analysis on Student Learning Behavior Using Data Mining Techniques. *Journal of Education and Information Technologies*, 16(2), 167-178.

[19]   Setiawan, M. A., & Kusumaningrum, D. (2020). The Impact of Social Media on Student Academic Achievement: A Case Study in Indonesia. *Journal of Educational Technology*, 18(3), 134-142.

[20]   Zhang, X., & Liu, M. (2017). Clustering Student Behavior in Educational Data Mining: A Review and Case Study. *Computers & Education*, 113, 47-62.

Data Mining Analysis of the Influence of Social Media on Students' Sleep Hours and GPA Using the Cluster Method- **Ibrahim et.al**

166