International Journal Software Engineering and Computer Science (IJSECS)
4 (2), 2024, 647-655
Published Online August 2024 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN : 2776-4869, E-ISSN : 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v4i2.2874.

RESEARCH ARTICLE Open Access

Data Mining Modeling Using the K-Means Algorithm to Analyze the Impact of New Media on Early Childhood Psychology at Bimba Rainbow Kids Sukmajaya

Sugiyono
Informatics Engineering Study Program, Faculty of Computer Technology, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: inosoguy007@gmail.com.

Haryati *
Informatics Engineering Study Program, Faculty of Computer Technology, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Corresponding Email: hharyati963@gmail.com.

Frencis Matheos Sarimole
Informatics Engineering Study Program, Faculty of Computer Technology, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: matheosfrancis.s@gmail.com.

Tundo
Informatics Engineering Study Program, Faculty of Computer Technology, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: asna8mujahid@gmail.com.

Received: June 29, 2024; Accepted: July 20, 2024; Published: August 10, 2024.

Abstract: New media, particularly the internet, has become an integral aspect of contemporary life, fundamentally altering the ways in which individuals interact, learn, play, and access information. The continuous evolution of new media, driven by technological advancements, exerts a profound influence on its users, with implications that span various dimensions of human experience. This study aims to analyze and classify the psychological impact of new media on early childhood, specifically within the context of Bimba Rainbow Kids Sukmajaya, utilizing the K-Means data mining method. This research employs a qualitative approach to uncover the underlying factors that shape the psychological effects observed in young children. The anticipated outcomes of this study are expected to contribute significantly to the academic discourse on the influence of new media on early childhood psychology. Moreover, the findings hold potential relevance for educators, parents, teachers, policymakers, and the general public who are invested in comprehending the broader implications of new media on the psychological development of early childhood.

Keywords: Early Childhood; Psychological Impact; Data Mining; K-Means; New Media.

1. Introduction

Technological advancements in recent decades have significantly transformed the ways humans interact, communicate, and access information. New media technologies, particularly the internet, have become an integral part of daily life. In this digital age, access to various forms of new media has become incredibly easy, directly influencing multiple aspects of life, including the learning processes and development of children. Children today grow up in environments saturated with new media technologies, such as the internet, smart devices, and social media platforms. Exposure to new media from an early age has raised concerns among education and psychology experts, as children are considered the most vulnerable group to the influences of new media [1]. The rapid pace of technological advancement has altered the ways children interact, learn, and process information, which can impact their cognitive, emotional, and social development [2].

High exposure to new media can have various effects on children's psychological development. Excessive or inappropriate use of new media can lead to issues such as media addiction, exposure to inappropriate content, and disruptions in sleep patterns and physical health [3]. Other negative effects include the potential for media misuse and reduced social interaction, which is crucial for children's social development. However, on the other hand, new media also brings benefits, such as enhanced technological skills, global awareness, and access to interactive education, which can enrich children's learning experiences [4]. To analyze the effects of new media on child development, one effective method is data mining, particularly using the K-Means algorithm. The K-Means algorithm is a clustering technique that can group data based on specific characteristics, in this case, children's interaction patterns with new media [5]. This method allows researchers to identify groups of children with similar new media usage patterns and then analyze the positive and negative effects that may arise [6].

This study aims to utilize the K-Means algorithm to analyze the impact of new media on the psychology of early childhood at Bimba Rainbow Kids Sukmajaya. The research will identify factors influencing the impact of new media on young children and how these impacts can be categorized using the K-Means algorithm. This method has proven effective in various prior studies, such as clustering data related to the impact of COVID-19 [7], social aid management [5], and academic data analysis [8], suggesting it may yield relevant results in this study. By gaining a better understanding of how new media affects early childhood psychology, the results of this research are expected to provide significant contributions to educators, parents, teachers, policymakers, and the general public. The findings will aid in formulating more effective strategies for managing media exposure in children, with the primary goal of supporting their psychological development and well-being.

2. Research Method

In this research, various data collection methods were employed to gather the necessary information. The first method used was observation, which involved direct monitoring of the subjects at Paud Rainbow Kids Sukmajaya. The researcher observed the activities of the students during the learning process and recorded relevant and important information. This method provided an opportunity to gather firsthand data on the children’s behaviors and interactions within their learning environment, allowing for an in-depth understanding of their responses to new media. The second method utilized was interviews. This method involved structured conversations between the researcher and the respondents to obtain detailed information related to the research topic. In this study, the researcher conducted interviews with teachers and staff at the school to gain insights into their perspectives on the learning practices at Bimba Rainbow Kids Sukmajaya. These interviews were crucial for understanding the pedagogical approaches used and the observed effects of new media on the children’s psychological and behavioral development. Questionnaires were the third method used in this study. This method involved the distribution of systematically arranged questions designed to measure stress levels and emotional responses in children. The questionnaires were administered to the parents or guardians of the children who were the subjects of the research, seeking their observations and opinions on the topic under study. The responses from the questionnaires provided quantitative data that could be analyzed to identify patterns and correlations between new media usage and psychological outcomes in the children.

The fourth method employed was a literature review. This method involved the analysis of previously published sources, such as books, journals, articles, and other relevant documents. The purpose of the literature review was to build a foundation of existing knowledge on the subject matter, identifying key themes and findings that have been established by other researchers. By examining these sources, the researcher could contextualize the current study within the broader field of research and ensure that the study’s objectives were aligned with existing scholarship. By using these methods, the researcher was able to collect comprehensive data necessary to support the analysis and findings of this study. The research was conducted at Bimba Rainbow Kids, located in Abadijaya, Sukmajaya, Kota Depok, Jawa Barat 16417.

The first step in the research methodology was data collection. During this phase, the researcher selected and gathered data relevant to the research objectives. The data for this study were collected directly from Bimba Rainbow Kids Sukmajaya using three primary methods: surveys administered to parents or guardians, student assessments, and direct interviews with teachers, as well as observations of children interacting with new media. The surveys provided by the parents or guardians included questions related to the children's media usage habits, the duration of media use, the children's preferences for certain types of media, and the observed effects noted by the parents or guardians. The student assessments and interviews with teachers were conducted in person, with the researcher asking specific questions and reaching a consensus on the evaluation criteria used by the teachers and the researcher. Direct observation of the children in a controlled environment was also conducted, allowing the researcher to note the children's immediate responses.

The second step involved data preprocessing, which included a series of steps to clean and prepare the data for analysis using the K-Means method. This process involved removing incomplete or irrelevant data, handling missing values, dealing with outliers, and transforming data as necessary. Data preprocessing also included selecting the most relevant features for analysis using the K-Means method. Ensuring that the data was clean and well-prepared was crucial for the accuracy and reliability of the subsequent analysis. The third step was data normalization, a process in which the data was adjusted to a more standardized or normalized scale. Data normalization helps eliminate scale differences between variables, ensuring that the analysis is not biased by differences in the magnitude of the data. In this study, normalization was performed using the StandardScaler method from the sklearn library, which adjusts the data distribution to have a mean of zero and a standard deviation of one. This step was essential because the K-Means algorithm calculates distances between data points based on feature scale. The fourth step was data modeling. In this phase, the researcher developed the K-Means model, representing the preprocessed data. The K-Means method is a clustering technique that groups data into clusters based on similarities determined by the distance between data points. The final step was data analysis. At this stage, the researcher analyzed the collected and modeled data. After the clustering process was completed, the data was evaluated using metrics such as inertia and Silhouette Score to assess the quality and interpretation of the generated clusters. The results of the clustering were visualized in plots to facilitate the extraction of information that could be used for decision-making or further understanding of the data.

3. Result and Discussion

3.1 Results

In the initial phase of this study, the researcher undertook a comprehensive data collection process directly from Bimba Rainbow Kids Sukmajaya, using multiple methods such as surveys directed at parents or guardians, student assessments, direct interviews with teachers, and observations of students. These varied approaches ensured a robust dataset, reflecting a wide range of variables and perspectives. The primary aim of this phase was to gather accurate and relevant information concerning the psychological and behavioral impact of new media on children, specifically within the context of early childhood education. The researcher collaborated closely with the involved teachers to ensure the accuracy of the evaluations, which were conducted on 106 students who were the subjects of the study. The variables considered in this evaluation included parental control, social activity, emotional response, cognitive ability, and screen time. The collected data were regularly organized and processed using Microsoft Excel, which served as a foundational tool for the initial data handling and preparation (see Figure 1 Student Data). Once the data were collected, they were transferred to Google Colab for further processing. This step involves a meticulous preprocessing procedure aimed at ensuring the cleanliness and integrity of the data, which is crucial for subsequent analysis. Preprocessing included the removal of incomplete or invalid entries, addressing missing values through imputation or exclusion, and the identification and elimination of outliers that could potentially distort the analysis results. In this phase, unnecessary data such as IDs, names, gender, and age were excluded from the analysis to focus on the most relevant variables (Figure 2 Preprocessing).

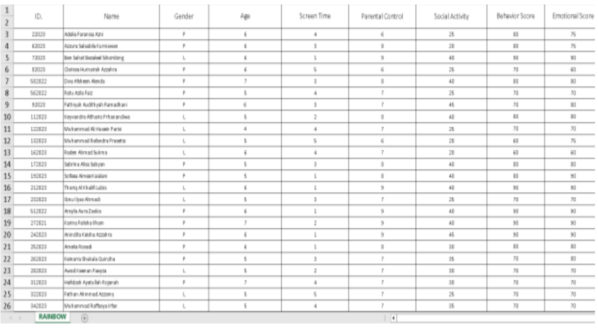
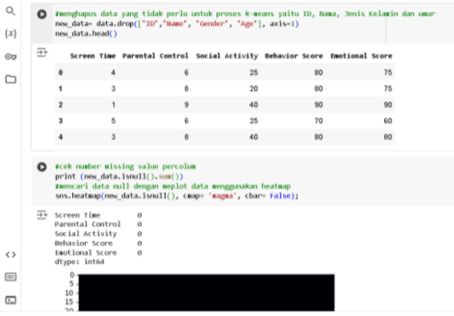


Figure 1. Student Data



```
#menghapus data yang tidak perlu untuk proses k-means yaitu ID, Nama, Jenis Kelamin dan umur
new_data= data.drop(["ID","Name", "Gender", "Age"], axis=1)
new_data.head()
```

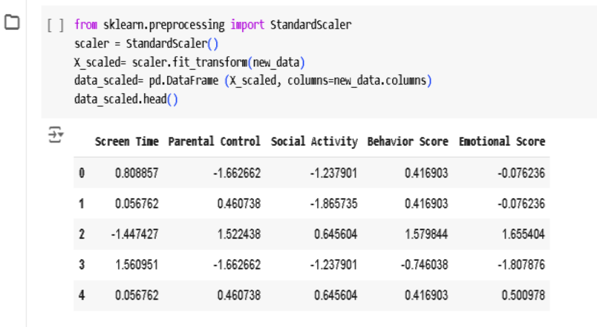
	Screen Time	Parental Control	Social Activity	Behavior Score	Emotional Score
0	4	6	25	80	75
1	3	8	20	80	75
2	1	9	40	90	90
3	5	6	25	70	60
4	3	8	40	80	80

```
#cek number missing value percolum
print (new_data.isnull().sum())
#mencari data null dengan meplot data menggunakan heatmap
sns.heatmap(new_data.isnull(), cmap= 'magma', cbar= False);
```

```
Screen Time         0
Parental Control    0
Social Activity     0
Behavior Score      0
Emotional Score     0
dtype: int64
```

Figure 2. Preprocessing

Following the preprocessing stage, the data were subjected to normalization, a critical process to ensure that all features are scaled uniformly and do not disproportionately influence the clustering results. Normalization was performed using the StandardScaler method from the sklearn library, which adjusts the data distribution to have a mean of zero and a standard deviation of one. This step is particularly important in the context of the K-Means algorithm, as it relies on calculating distances between data points based on feature scales. Without normalization, the algorithm might produce biased or inaccurate clusters, which would undermine the validity of the study's findings (Figure 3 Data Normalization). After normalization, the K-Means algorithm was applied to perform data clustering. The initial step in this process involved determining the optimal number of clusters, which was accomplished using the Elbow Method. This method involves plotting the within-cluster sum of squares (inertia) against the number of clusters and identifying the point where the rate of decrease sharply changes, forming an "elbow" in the graph. In this study, the Elbow Method indicated that the most appropriate number of clusters was two, as this was where the elbow was most pronounced (Figure 4 Elbow Method). Following this determination, the researcher proceeded with the implementation of the K-Means algorithm using Python, supported by the sklearn library.



```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_scaled= scaler.fit_transform(new_data)
data_scaled= pd.DataFrame (X_scaled, columns=new_data.columns)
data_scaled.head()
```

	Screen Time	Parental Control	Social Activity	Behavior Score	Emotional Score
0	0.808857	-1.662662	-1.237901	0.416903	-0.076236
1	0.056762	0.460738	-1.865735	0.416903	-0.076236
2	-1.447427	1.522438	0.645604	1.579844	1.655404
3	1.560951	-1.662662	-1.237901	-0.746038	-1.807876
4	0.056762	0.460738	0.645604	0.416903	0.500978

Figure 3. Data Normalization

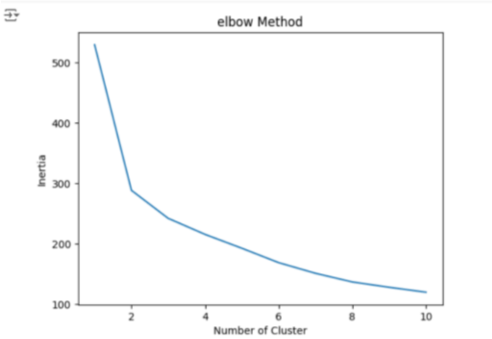
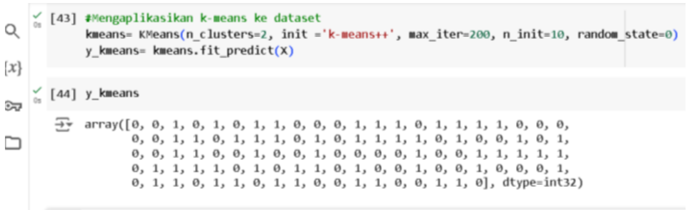


Figure 4. Elbow Method

Once the clusters were established, the researcher added cluster labels to the original dataset, enabling further analysis (Figure 5 Application of the K-Means Algorithm). The addition of these labels allowed for a detailed examination of how different groups of students, identified based on their interaction with new media, differed in terms of the selected variables. Visualization of the clustering results was then conducted, which provided a clear representation of how the data were divided into two distinct clusters, as well as the positioning of the centroids for each cluster (Figure 6 Adding Labels). The visualization was crucial for interpreting the results and understanding the relationships between different variables within each cluster.

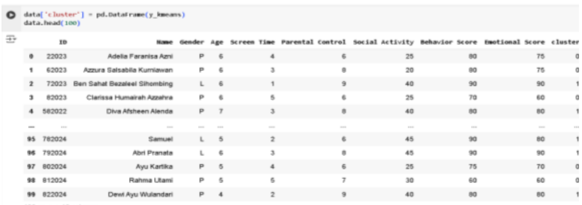


```
[43] #Mengaplikasikan k-means ke dataset
     kmeans= KMeans(n_clusters=2, init ='k-means++', max_iter=200, n_init=10, random_state=0)
     y_kmeans= kmeans.fit_predict(X)

[44] y_kmeans

array([0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0,
       0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1,
       0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1,
       0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1,
       0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0], dtype=int32)
```

Figure 5. Application of the K-Means Algorithm



```
data['cluster'] = pd.DataFrame(y_kmeans)
data.head(100)
```

	ID	Name	Gender	Age	Screen Time	Parental Control	Social Activity	Behavior Score	Emotional Score	cluster
0	22023	Adelia Faranisa Azni	P	6	4	6	25	80	75	0
1	62023	Azzura Salsabila Kurniawan	P	6	3	8	20	80	75	0
2	72023	Ben Sahat Bezaleel Sihombing	L	6	1	9	40	90	90	1
3	82023	Clarissa Humairah Azzahra	P	6	5	6	25	70	60	0
4	582022	Diva Afsheen Alenda	P	7	3	8	40	80	80	1
...	...	...	...	...	...	...	...	...	...	...
95	782024	Samuel	L	5	2	6	45	90	80	1
96	792024	Abri Pranata	L	6	3	8	45	90	90	1
97	802024	Ayu Kartika	P	5	4	6	25	75	70	0
98	812024	Rahma Utami	P	5	5	7	30	60	60	0
99	822024	Dewi Ayu Wulandari	P	4	2	9	40	80	80	1

100 rows × 10 columns

Figure 6. Adding Labels

The visualization revealed two primary types of data points: black and blue dots representing Cluster 0 and Cluster 1, respectively (Figure 7 Cluster Visualization). Black dots indicated data belonging to Cluster 0, while blue dots represented data in Cluster 1. These points illustrate how the data were grouped based on similarities in features. Additionally, yellow dots were used to mark the centroids or central points of each cluster. These centroids are crucial in determining the proximity of data points to the center of their respective clusters, effectively summarizing the average location of data within each cluster.

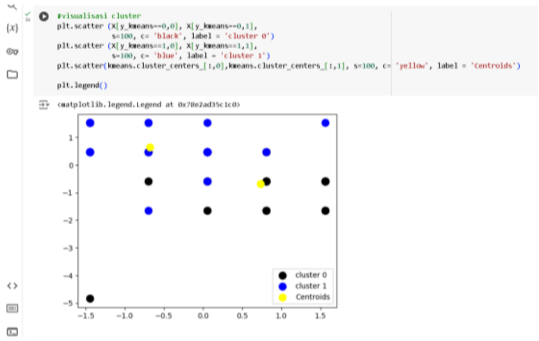


Figure 7. Cluster Visualization

	cluster	size	Percentage
0	0	51	0.481132
1	1	55	0.518868

Figure 8. Cluster Percentage

The clustering results were further analyzed to identify the characteristics of each cluster. Cluster 0 consisted of students with high screen time, whereas Cluster 1 consisted of students with low screen time. The final analysis showed that students in Cluster 1, who had lower screen time, exhibited higher levels of parental control, greater social activity, higher behavior scores, and higher emotional scores. Conversely, students in Cluster 0, who had higher screen time, were associated with lower levels of parental control, reduced social activity, lower behavior scores, and lower emotional scores (Figure 8 Cluster Percentage). These findings suggest a correlation between the duration of media exposure and various psychological and behavioral variables. Specifically, children with higher screen time tend to experience less parental control, engage in fewer social activities, and demonstrate lower behavioral and emotional well-being compared to those with lower screen time. The data analysis indicates that high screen time is associated with negative outcomes in several key areas of child development, while lower screen time correlates with more positive outcomes. This analysis provides valuable insights into the impact of media usage on early childhood development and underscores the importance of managing screen time to promote better psychological and behavioral health in children (Figure 9 Cluster 0 and Figure 10 Cluster 1).

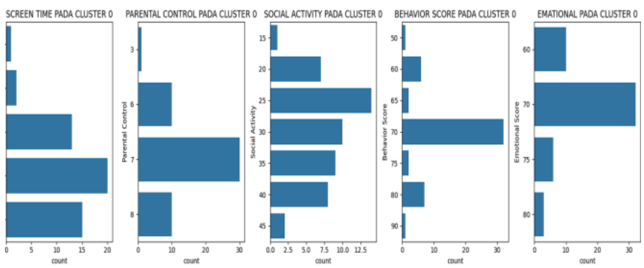


Figure 9. Cluster 0

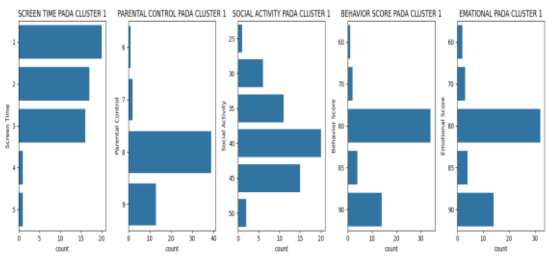


Figure 10. Cluster 1

3.2 Discussion

The findings from this study clearly demonstrate the relationship between the duration of new media use and various aspects of psychological and behavioral development in early childhood. By utilizing the K-Means algorithm to cluster data based on children's interaction patterns with new media, the results offer a detailed understanding of how new media affects different groups of children. high screen time and lower parental control. Children in Cluster 0, characterized by high screen time, tend to have lower levels of parental control. This may be due to several factors, including the difficulty parents face in limiting their children's access to new media devices or a lack of stringent supervision. On the other hand, children in Cluster 1, who have lower screen time, exhibit higher levels of parental control. This finding aligns with existing literature that suggests increased parental supervision can mitigate the negative impacts of excessive media use on children [3].

The study also reveals that new media usage affects children's social activities. Children in Cluster 0, who have high screen time, show lower social activity compared to those in Cluster 1. This reduction in social interaction may be attributed to the time children spend in front of screens, which might decrease their opportunities for direct interaction with peers. This finding is consistent with previous studies that indicate excessive media use can isolate children from healthy social interactions that are crucial for their development [9]. The data analysis further shows that children with high screen time in Cluster 0 tend to have lower behavioral and emotional scores. This suggests that excessive exposure to new media may negatively impact children's emotional well-being and behavior. New media, particularly inappropriate digital content, can trigger negative behaviors or lead to emotional issues such as anxiety or depression in children. Conversely, children in Cluster 1, who have lower screen time, demonstrate higher behavioral and emotional scores, indicating that better management of screen time supports healthier emotional and behavioral development [4]. Using the K-Means algorithm, this study successfully grouped children into two main clusters based on their media usage patterns and related psychological effects. Cluster 0, consisting of children with high screen time, showed several concerns related to their development, while Cluster 1, with low screen time, exhibited more positive outcomes. This grouping allows researchers to not only understand the relationships between different variables but also to formulate more specific recommendations for parents and educators on managing children's media use.

Based on these findings, several recommendations can be made to parents, educators, and policymakers. First, it is crucial to enhance parental control over children’s media use by setting clear screen time limits and ensuring that the content accessed by children is age-appropriate. Second, encouraging social activities away from screens, such as playing with friends or participating in group activities, is important for developing children's social and emotional skills. Third, educators and schools can play a key role in providing parents with education about the effects of new media and how to manage its use effectively. Overall, this study underscores the importance of managing screen time in early childhood to support psychological and behavioral development. The results from the clustering using the K-Means algorithm provide a clear picture of how media usage patterns can influence various aspects of child development. These findings can serve as a foundation for developing more effective strategies for managing media exposure in children, ultimately enhancing their overall well-being.

4. Related Work

The K-Means clustering algorithm has been extensively researched and applied across various disciplines, underscoring its utility and adaptability in analyzing complex and heterogeneous data sets. In the domain of public health, Rizal and Khotimah (2022) employed K-Means to cluster populations impacted by COVID-19 and to categorize regions based on the severity of the pandemic. This methodological approach facilitated more targeted public health interventions and optimized resource allocation, demonstrating the algorithm's effectiveness in addressing large-scale health crises [1]. A similar approach was adopted by Azhari *et al.* (2023), who utilized K-Means to analyze COVID-19 data, focusing specifically on categorizing areas according to infection frequency and severity. While both studies leveraged K-Means in public health, Rizal and Khotimah's research focused on broader regional analysis, whereas Azhari *et al.* provided a more granular categorization, thus enabling more precise public health strategies [7].

In socio-economic research, K-Means has been employed to analyze complex social data, with studies like that of Sudibyo *et al.* (2020), which clustered regions in Indonesia based on poverty indicators. This clustering provided valuable insights that could inform policymakers in the design of effective poverty alleviation programs [2]. Similarly, Ikhwan and Aslami (2020) applied K-Means to manage social assistance programs, clustering beneficiaries according to their socio-economic characteristics. This application of K-Means facilitated more efficient resource distribution by targeting assistance to those most in need. While both studies used K-Means in socio-economic, Sudibyo *et al.* focused on regional-level analysis, whereas Ikhwan and Aslami applied the algorithm at the individual level to optimize social assistance delivery [5].

The pharmaceutical sector has also benefited from the application of K-Means. For instance, Gustientiedina, Adiya, and Desnelita (2019) utilized the algorithm to classify drugs based on usage patterns, which helped optimize inventory management and improve distribution strategies within healthcare systems. Their research highlighted the algorithm's versatility in processing medical data, particularly in operational [3]. A parallel study by Tambunan (2021) also applied K-Means in the pharmaceutical domain, focusing on drug classification to enhance broader operational efficiencies in healthcare settings. While both studies demonstrated the utility of K-Means in healthcare, Gustientiedina *et al.* concentrated on usage patterns, whereas Tambunan emphasized broader operational improvements [12].

K-Means has been employed in the educational sector to analyze student data and tailor educational strategies better. Gustian and Al-Farits (2023) used the algorithm to cluster students according to their learning preferences, which can assist educators in designing personalized learning experiences [4]. Similarly, Saputra and Nataliani (2021) applied K-Means to analyze academic performance data, identifying high-achieving students who may benefit from additional educational opportunities or support. Although both studies utilized K-Means within education, they differed in their focus: Gustian and Al-Farits concentrated on learning preferences, while Saputra and Nataliani emphasized academic performance and potential interventions [8].

Demographic studies have further demonstrated the applicability of K-Means. Rahmawati and Bahtiar (2023) clustered adolescents in Desa Sindangsari based on age segmentation, providing insights into the distinct developmental needs and challenges different age groups face. This study's findings can inform more effective youth programs and policies [6]. Similarly, Simarmata and Samuel (2021) used K-Means to investigate the influence of gadget use on high school students' academic performance. Their study clustered students based on gadget usage patterns, revealing significant correlations between excessive gadget use and diminished academic outcomes. Both studies focus on youth, with Rahmawati and Bahtiar addressing developmental stages and Simarmata and Samuel exploring the impact of technology on education [9].

K-Means has been applied to analyze environmental and geographic data in disaster management. Al Halik and Septiana (2022) used the algorithm to predict regions prone to natural disasters in West Java by clustering areas based on historical disaster data. This application provided crucial insights for disaster preparedness and resource allocation [14]. Similarly, Ramadhani *et al.* (2022) utilized K-Means to cluster villages based on their susceptibility to natural disasters, using data on past events to develop more effective disaster management plans. Both studies underscore the algorithm's capacity to process large-scale environmental data for risk management [17].

In the digital realm, Nugroho, Ma'arif, and Arif (2022) conducted a systematic review on the use of K-Means in analyzing social media misuse, highlighting its effectiveness in detecting and categorizing patterns of misuse, which is critical for developing strategies to mitigate the negative impacts of social media [13]. Likewise, Afrilia *et al.* (2024) applied K-Means to optimize clustering techniques for assessing social media user activities and responses, providing insights to improve user engagement and content delivery strategies. While Nugroho *et al.* focused on misuse detection, Afrilia *et al.* concentrated on enhancing user experience, illustrating the algorithm's adaptability in various digital challenges 0.

In conclusion, the studies referenced underscore the broad applicability and effectiveness of the K-Means clustering algorithm across multiple disciplines. While these studies share a common methodological approach, they differ in their specific applications, data sets, and the challenges they aim to address. The consistent success of K-Means in these diverse fields highlights its critical role in advancing research and practical applications, offering robust solutions for complex analytical tasks.

5. Conclusion

This study utilized the K-Means algorithm to analyze the impact of new media usage on the psychology of early childhood at Bimba Rainbow Kids Sukmajaya. The analysis revealed that children with high screen time exposure tend to exhibit lower levels of parental control, limited social activities, and reduced behavioral and emotional scores compared to children with lower screen time exposure. These findings underscore the importance of regulating screen time in early childhood, given the significant effects it can have on various psychological aspects. The study highlights the necessity of a holistic approach in managing media exposure among children, one that not only considers the direct effects of technology but also its broader impact on psychological and social development. The implications of this research are particularly relevant for the development of more effective educational policies aimed at managing digital technology in early childhood education settings. Furthermore, the findings offer valuable guidance for parents in making more informed decisions regarding their children's media use, with the primary goal of enhancing the quality of social interactions and emotional well-being. These results provide a foundation for policy interventions designed to support healthy psychological and social development in young children in the digital age.

References

[1] Rizal, S., & Khotimah, R. Q. (2022). Penerapan data mining untuk clustering data penduduk yang terdampak Covid-19 menggunakan algoritma K-Means. *Jurnal Pendidikan dan Konseling, 4*(4), 2781–2792.

[2] Sudibyo, N. A., Iswardani, A., Sari, K., & Suprihatiningsih, S. (2020). Penerapan data mining pada jumlah penduduk miskin di Indonesia. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika dan Matematika, 1*(3), 199–207. https://doi.org/10.46306/lb.v1i3.42

[3] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan algoritma K-Means untuk clustering data obat-obatan. *Jurnal Nasional Teknologi dan Sistem Informasi, 5*(1), 17–24. https://doi.org/10.25077/teknosi.v5i1.2019.17-24

[4] Gustian, D., & Al-Farits, M. S. (2023). Data mining untuk melihat minat belajar siswa menerapkan metode K-Means. *Jurnal Information System Research, 4*(3), 775–784. https://doi.org/10.47065/josh.v4i3.3218

[5] Ikhwan, A., & Aslami, N. (2020). Implementasi data mining untuk manajemen bantuan sosial menggunakan algoritma K-Means. *Jurnal Teknologi Informasi, 4*(2), 208–217. https://doi.org/10.36294/jurti.v4i2.2103

[6] Rahmawati, R., & Bahtiar, A. (2023). Pengelompokan remaja berdasarkan segmentasi usia menggunakan metode K-Means clustering (Studi kasus: Desa Sindangsari). *Jurnal Riset Ilmu Akuntansi, 2*(2), 35–51. https://doi.org/10.37600/tekinkom.v2i2.115

[7] Azhari, R., Hartama, D., Lubis, M. R., Nasution, D. F., & Windarto, A. P. (2023). Analisis penerapan data mining terhadap kasus positif Covid-19 menggunakan metode K-Means clustering. *Jurnal Informatics, Electronics, and Electrical Engineering, 3*(2), 221–235. https://doi.org/10.47065/jieee.v3i2.1760

[8] Saputra, E. A., & Nataliani, Y. (2021). Analisis pengelompokan data nilai siswa untuk menentukan siswa berprestasi menggunakan metode clustering K-Means. *Jurnal Information System Informatics, 3*(3), 424–439. https://doi.org/10.51519/journalisi.v3i3.164

[9] Simarmata, R., & Samuel, Y. T. (2021). Analisa pengaruh penggunaan gadget terhadap nilai akhir siswa SMA secara umum menggunakan metode data mining (Decision Tree). *TeIKa, 11*(1), 15–28. https://doi.org/10.36342/teika.v11i1.2475

[10] Apriyani, P., Dikananda, A. R., & Ali, I. (2023). Penerapan algoritma K-Means dalam klasterisasi kasus stunting balita desa Tegalwangi. *Hello World Journal of Computer Science, 2*(1), 20–33. https://doi.org/10.56211/helloworld.v2i1.230

[11] Aristika, W., & Hartono, W. J. (2020). Penerapan clustering K-Means untuk menentukan pengaruh media sosial Facebook terhadap usaha mikro, kecil dan menengah (UMKM) di Kecamatan Pekanbaru Kota. *Jurnal Ilmu Komputer dan Bisnis, 11*(1), 2389–2395.

[12] Tambunan, M. P. (2021). Penerapan data mining dalam analisa data pemakaian obat dengan menerapkan algoritma K-Means. *Jurnal Informasi dan Teknologi Ilmiah, 8*(3), 109–113.

[13] Nugroho, B. I., Ma'arif, Z., & Arif, Z. (2022). Tinjauan pustaka sistematis: Penerapan data mining metode klasifikasi untuk menganalisa penyalahgunaan sosial media. *Jurnal Sistem Informasi dan Teknologi Informasi, 3*(2), 46–51. http://journal.peradaban.ac.id/index.php/jsitp/article/download/1265/860

[14] Al Halik, M. F., & Septiana, L. (2022). Analisa data untuk prediksi daerah rawan bencana alam di Jawa Barat menggunakan algoritma K-Means clustering. *Jurnal Information Systems Applied, Management, and Accounting Research, 6*(4), 856–870. https://doi.org/10.52362/jisamar.v6i4.939

[15] Indraputra, R. A., & Fitriana, R. (2020). K-Means clustering data COVID-19. *Jurnal Teknik Industri, 10*(3), 275–282. https://doi.org/10.25105/jti.v10i3.8428

[16] Afrilia, M. N., Rahaningsih, N., Dana, R. D., & Nuris, N. D. (2024). Optimasi analisis clustering untuk aktivitas dan respon pengguna media sosial dengan K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika), 8*(1), 148–155. https://doi.org/10.36040/jati.v8i1.8334

[17] Ramadhani, D. I., Damayanti, O., Thaushiyah, O., & Kadafi, A. R. (2022). Penerapan metode K-Means untuk clustering desa rawan bencana berdasarkan data kejadian terjadinya bencana alam. *JURIKOM (Jurnal Riset Komputer), 9*(3), 749. https://doi.org/10.30865/jurikom.v9i3.4326.

