The Intersection of Machine Learning, Zakat, and Transportation for a Healthy Society

Moch Yusuf Asyhari[1], Pratiwi Susanti[2], Juwari[3], Khairul Adilah Ahmad[4], Norin Rahayu Shamsuddin[5], Taniza Tajuddin[6]

[1,2,3] Informatics Engineering, Universitas PGRI Madiun, Madiun, Indonesia
[4,5,6] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Kedah Branch, Malaysia
yusuf.asyhari@unipma.ac.id

Abstract. The recent Covid-19 pandemic has shown that transportation has a significant role in creating a healthy society. Human mobility can be tracked and traced to map the spread of infectious diseases using screening, tracing, and tracking methods through policies and applications. This momentum has accelerated the massive application of technology, one of which is using Machine Learning. Zakat also has an essential role during the pandemic quarantine policy. The potential of the combination of Machine Learning, Zakat, and transportation should now be increased to improve the Quality of Life to achieve a healthy society. The intersection of these three topics was researched using data scraping from the web dimension as one of the platforms that accommodates a collection of research publications. Data is limited to abstracts of research results within the last five years. Next, the data obtained was preprocessed, text-processed, and analyzed using the FP-Growth Algorithm for Association Rule. Research published so far shows that there is an intersection between Machine Learning, Zakat, and Transportation to achieve a healthy society, even with varying intersection probability values. Machine learning and transportation have the most potent intersection with health, while zakat is still lower. The results of this research can open up collaborative research topics between these domains. The results of this research can also be used as suggestions for policy-making by stakeholders to explore the existing potentials of various journals that have been published.

Keywords: Association Rule, Data Science, Machine Learning, Transportation, Zakat

1 Introduction

The Covid-19 pandemic has taught us that transportation is essential in the health sector [1]. The relationship between transportation and other fields is interesting to know. This step is one way to measure the extent of cross-disciplinary research that has been carried out. It is hoped that each research result will have great benefits. Cross-sector collaboration is expected to produce even more significant benefits.

A. Pradipta et al. (eds.), *Proceedings of the 2nd International Conference on Railway and Transportation 2023 (ICORT 2023)*, Advances in Engineering Research 231,
https://doi.org/10.2991/978-94-6463-384-9_40

1.1 Healthy Society

The Covid-19 pandemic has proven that a healthy society significantly impacts world development in various sectors [2]. This issue often becomes a concern at every opportunity in meetings between nations. Post-Covid-19 has become a momentum to improve the health of society as an effort to achieve Sustainable Development Goals (SDGs) and Quality of Life. Sustainable Development (SDGs) is a global call to improve the lives of everyone, leaving no one behind by 2030 [3], [4]. The challenge of building a healthy society requires collaboration from many lines. One step is to explore the relationships between these sectors. Some interesting sectors in this research are Transportation, Zakat, and Artificial Intelligence represented by Machine Learning.

1.2 Machine Learning, Zakat, and Transportation

Interpretability in machine learning (ML) is critical for high-stakes decision making and problem solving. There are many technical challenges in Machine Learning [5]. Machine Learning is a variety of tools for learning from data. Machine Learning can be divided into four types, namely Supervised Learning, Unsupervised Learning, Reinforcement Learning, and Semi-Supervised Learning [6]. Machine learning can be used in many applications and has experienced significant changes in various industries, such as technology, healthcare, economics, transportation and other fields.

The importance of zakat is shown in the Al-Quran surah Al Baqarah verse 48 which reads, "Establish prayer, pay zakat and bow together with those who bow." Zakat is an obligation for a Muslim who believes and has the ability to pay zakat. The results of zakat collection have clear recipients. Al-Quran surah At Taubah verse 60 explains that there are seven groups of zakat recipients. These groups are the poor, poor people, zakat administrators, converts who are persuaded by their hearts to free slaves, people who are in debt, for the cause of Allah and for those who are on their way.

An interesting part of the relationship between transportation and zakat is shown in one of the groups entitled to receive zakat. Travelers or people who are on a journey and no longer have provisions to get to their destination are one of the groups receiving zakat. Zakat recipients are very closely related to human mobility and transportation. This research tries to measure the extent of research involving zakat and transportation, as well as machine learning and health.

1.3 Previous Condition

The relationship between transportation and health is closely related in various publications, such as research on the role of transportation on health [3], transportation on quality of life in urban areas [7], transportation challenges to reach health facilities during the pandemic [8], and more research related to these two fields. A similar relationship also applies between machine learning and health, such as machine learning for dermatology [9], the intersection of machine learning with epidemiology [10], and so on. Although not as much as transportation and machine learning, relationships between zakat and health have also been found, such as pandemic mitigation with zakat [11], economic recovery after a pandemic with zakat [12]. In stark

contrast to these relationships, the intersection between transportation, machine learning, and zakat is more difficult to trace. These three domains have their respective roles to play in achieving a healthy society. At least domain intersection between the three domains needs to be proven.

Methods for proving relationships include Cosine Similarity, Jaccard Similarity, Dice's Similarity, Clustering, Classification, Collaborative Filtering, and Association Rule. The Cosine Similarity, Jaccard Similarity, and Dice's Similarity methods compete with the same goal of understanding the relationships of words with different understandings [13], [14]. Text Clustering is used to group words based on patterns found [15]. Text Classification technique for studying data sets by providing labels [16]. Text Correlation is done by extracting words to compare their correlation [17]. Collaborative Filtering is used to select data according to the desired preferences [18]. The Association Rule was used in this research with the researcher's consideration of more targeted measurements with a top-down approach to finding data patterns compared to other methods.

2 Material and Methods

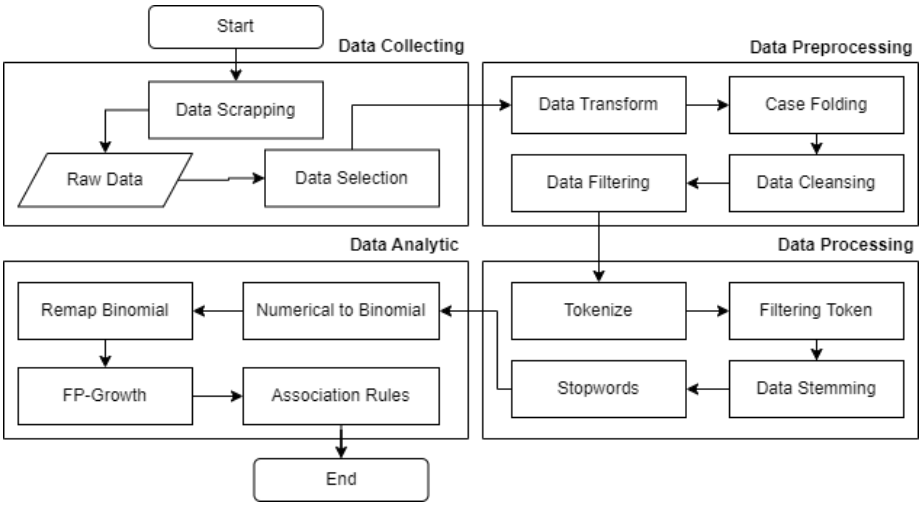

Fig. 1. The research method used in this study is divided into four stages. Data Collecting Stage, Data Preprocessing Stage, Processing Stage, and Data Analytic Stage.

2.1 Data Collecting

The first step taken in this research was collecting data. The desired data target is international journals within the last five years. The journal used is a free journal that can be accessed by anyone via the internet. There are various methods for collecting journal data on open internet sources, one of which is the Data Scrapping technique.

Data Scraping. Data scraping is a process of extracting data from other sources, such as websites [19]. The source of the journal data was carried out from the dimensions index, which is located at https://www.dimensions.ai/.

Data Selection. The data that will be used in the mining process first undergoes an attribute selection process because raw data has a variety of information and metadata that are not all used [20]. Efforts to select information that is relevant and appropriate to the research are carried out at this stage.

2.2 Data Preprocessing

Data attributes that have been pre-selected may have misunderstanding information when processed. This is because the existing data is not equivalent or not well identified. Preparing data before processing is one of the determinants for the results of data processing later. The data preprocessing stages can be divided into data transformation, case folding, data cleansing, and data filtering.

Data Transform. The process of converting data into another form according to the processing method that will be applied. The data used in the data mining method is in the form of text so that all data that still contains nominal amounts is completely converted into text. This method is used to avoid differences in perception between nominal and text [21].

Case Folding. The process of changing all data items to lowercase so that each capitalized word has the same position as the other words. The final goal at this stage is to maintain data consistency and uniformity [22].

Data Cleansing. The process of selecting inaccurate data, inconsistent data, and irrelevant data. This process is carried out by deleting various characters or symbols that have no meaning in the research [21].

Data Filtering. The process of selecting data according to what is needed. This research limits the data used in journals that contain the words Transportation, Machine Learning, and Zakat. Data that was previously very large is now reduced with the advantage that the processing process is more focused and faster.

2.3 Data Processing

Data processing begins with deciphering each word in each existing sentence. This stage takes the longest time compared to the other stages. The more rows of data that are processed, the more words that must be parsed one by one. This stage can be divided into Tokenization, Filtering, and Stemming.

Tokenization. The process of changing sentences in a document. The process involves cutting each word into a token so that it can be analyzed [23]. The results obtained are the distribution of words in each row of data.

Filtering. Procedure for translating and sorting a word that still contains foreign words into a standard word [23]. This stage is also often known as Stopword Removal. This section requires a wordlist or stoplist tool to identify words that are important and words that are less important for analytical data. The words to be removed are in the stop list, such as the words "and", "or", "in", "from", and so on.

Stemming. The procedure used to remove affixes in the form of prefixes and suffixes in each word [23]. This process is to equalize the basic words used without paying attention to the affixes attached. For example, changing words with affixes to the words "healthy", "healthful", "healthier", "healthily", and "healthiest" to the basic word "health".

2.4 **Data Analytic**

The data is ready for analysis after the data processing stage is complete. This stage aims to find meaning in the word patterns that have been successfully described. The technique used at this stage is the Association Rule. This technique is a procedure-based method approach that aims to find relationships between objects in a collection of data sets [24]. The general form of using the Association Rule is to use LHS => RHS where LHS and RHS are a set; during a transaction, the presence of RHS items also appears in the transaction [25]. The rule for each item in the LHS to appear as an association is usually expressed in equation (1).

$$\{X,Y\} => \{Z\}\ (support = 10\%,\ confidence = 50\%) \quad (1)$$

The reference used by [25] shows that support is the possibility of consumers buying several products together. This calculation technique determines whether an item or itemset is worthy of finding its confidence value. For example, from the entire transaction process, what is the probability that items X and Y were purchased together. This explanation can be shown in equations (2) and (3).

$$Support(A,B) = P(A \cap B) \quad (2)$$

$$Support(A,B) = (\sum \text{transaction contains A and B}) / (\sum \text{transaction}) \quad (3)$$

Confidence in the Association Rule is the number of possibilities that arise when several products are purchased when one product is chosen before, for example, how often product Y is chosen when the user chooses product X. This rule can be expressed by equation (4).

$$Confidence = P(B \mid A) \quad (4)$$

The Association Rule technique is often used to find product pairs to buy. This research applies this approach to words that will appear together. The words specified are Transportation, Machine Learning, Zakat, and Health. This approach is expected to find intersections or non-intersections between these domains.

3 Result and Discussion

In line with the research methods that have been prepared, the research results can be divided into Data Collecting, Data Preprocessing, Data Processing, and Data Analytics. Each stage is carried out sequentially to produce an intersection conclusion between the specified words.

3.1 **Data Collecting**

The process of data collection uses data scraping techniques for dimension-indexed data sets. The built data set contains a collection of journal contents from various sources in the last five years. The total Raw Data obtained was 44,360 rows of data.

Some of the attributes found in the data scraping process include Publication ID, DOI, Title, Abstract, Anthology title, PubYear, Volume, Issue, Pagination, Authors, Authors Affiliations for Name of Research organization, Authors Affiliations for Country of Research organization, Dimensions URL, Times cited, and Cited references. This research chose the Abstract attribute to find the relationship between Machine Learning, Transportation, and Zakat on Health. The results of taking a data set whose attributes have been selected are shown in Table 1.

Table 1. Example of Data Selection for Abstract Attributes from the Raw Data obtained

Raw	Abstract
1	Transportation disadvantage may have important implications for the health, well-being, and quality of life of older adults. This study used the 2015 National Health Aging Trends Study, a nationally representative study of Medicare beneficiaries aged 65 and over (N = 7,498) ……
2	Although the Chinese government emphasizes the significance of public transportation development and encourages green travel, no empirical study has examined whether the expansion of public transportation facilitates the mitigation of carbon emissions.….

3.2 **Data Preprocessing**

This section carries out the data cleaning process from scrapping data that has been taken previously. The stages in the data cleaning process include carrying out the append all process for all data sets that have been collected, then carrying out the data transformation process by changing the value from nominal to text value, and the final step of the preprocessing process is carrying out the missing value process. The process of missing values in the data set can be overcome by replacing or deleting words that are not needed. Examples of preprocessing results are shown in Table 2.

Table 2. Example of Preprocessing Result

Raw	Abstract
1	transportation disadvantage may have important implications for the health wellbeing quality of life of older adults this study used the national health aging trends study a nationally representative study of medicare beneficiaries aged over ……..
2	although the chinese government emphasizes the significance of public transportation development encourages green travel no empirical study has examined whether the expansion of public transportation facilitates the migration of carbon emissions…….

3.3 **Data Processing**

Table 3 shows that the word health leads to the ten words that appear most frequently in this study. Machine learning is in sixth and seventh place, while transportation is in eighth place. The word zakat is far away at number 179. This means that zakat still needs to be widely published and has a vast space for research exploration.

Table 3. WordList Result

No	Word	Total Occurence	Document Occurences	No	Word	Total Occurence	Document Occurences
1	health	16135	7075	6	Learning	8122	3093
2	Study	9031	5048	7	Machine	7285	2939
3	Results	6251	4672	8	transportation	5173	2736
4	Data	10209	4543	9	Analysis	3703	2587
5	Methods	4937	3558	10	Care	6907	2474

The Machine Learning and Transportation journal has the closest relationship to the health domain. Zakat journals tend to be further away from the health domain. Machine Learning shows a more even distribution, which means it has a broader range of research collaboration (see Fig. 2.)

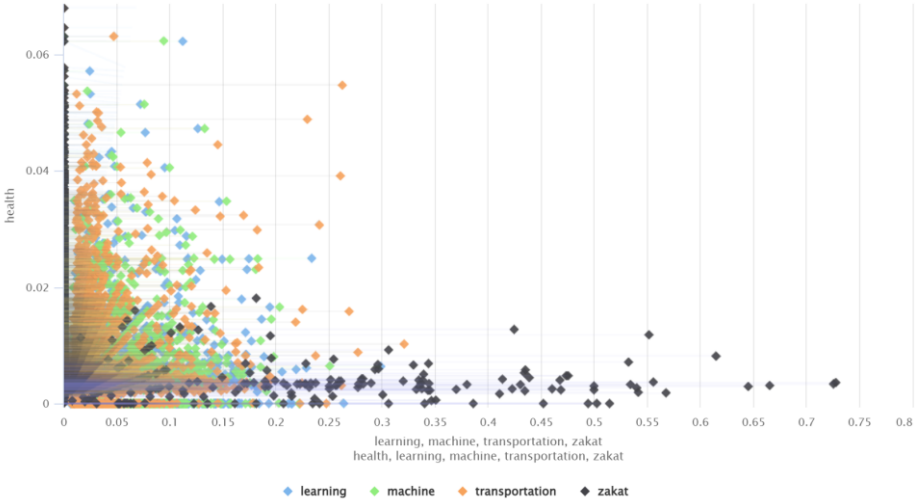

Fig. 2. The word distribution diagram on WordList for machine, learning, transportation, and zakat to the journal that contains health.

3.4 **Data Analytic**

The word descriptions resulting from data processing in the previous stage are processed again to have more significant meaning. The data is reprocessed using the FP-Growth algorithm from the Association Rule method. Association Rule results are shown in Table 4.

Table 4. Association Rule Result

Row	Premises	Conclusion	Support	Confidence
1	Machine	Learning	0,3213	0,9687
2	Transportation	Health	0,2684	0,8692
3	learning, machine	Health	0,2218	0,6902
4	Zakat	Health	0,0205	0,8750
5	Transportation	Learning	0,0062	0,0201
6	Transportation	health, learning	0,0049	0,0157
7	Transportation	Machine	0,0035	0,0113
8	learning, machine	transportation	0,0029	0,0091
9	learning, machine, transportation	Health	0,0019	0,6538
11	health, zakat	transportation	0,0007	0,0330
12	learning, machine	Zakat	0,0001	0,0035
14	Zakat	learning, machine	0,0001	0,0048
15	Zakat	health, learning, machine	0,0001	0,0048
16	health, zakat	learning, machine	0,0001	0,0055

Each premise and its partners have at least a support and confidence value. When we set a minimum support value limit in Table 4 with 10% and a minimum confidence value limit of 50%, only the top three premise and conclusion pairs meet the criteria. Machine and Learning are the words that appear most often together at 32% and are most often used when the word Machine appears first at 96%.

The higher the support value, the higher the chance that the pair of words will be used together in a study. The higher confidence value means that the words in the conclusion attribute also exist in the journal with the premise attribute words. The fundamental difference between support and confidence is the intersection probability, where the premises appear simultaneously, while the conditional probability of confidence appears if the condition of the first premise has occurred.

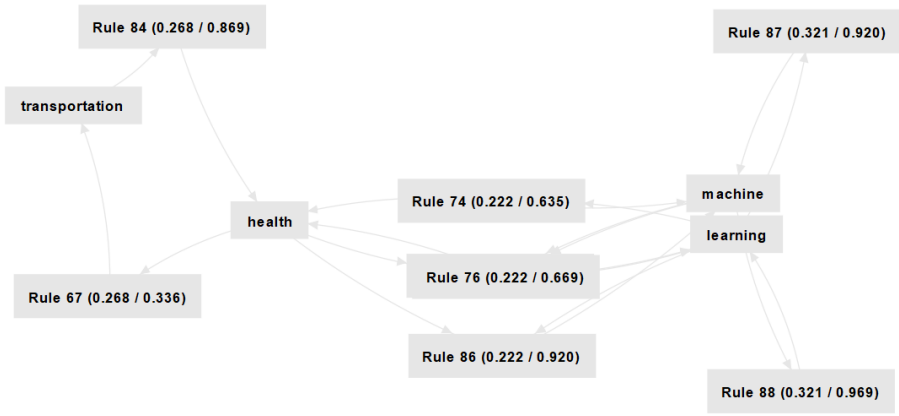

Fig. 3. The intersection of Machine Learning, Transportation, and Zakat on Health using the Association Rule technique.

The results of the support and confidence measurements produce many rules. After determining the minimum limit of 20%, several rules that are still available are shown in Fig. 3. Seven rules indicate the intersections that appear.

Rule 87 and Rule 88. Machine and learning often appear together at 32%, and the word machine appears followed by learning at 92%. The intersection between the words machine and word learning has been very high in scientific publications in the last five years.

Rule 84 and Rule 67. Rule 84 is the intersection between transportation and health at 26.8%, and the appearance of the word transportation in health journals is 86.9%. Rule 67 is the opposite of Rule 84, where the co-occurrence is 26.8%, and the appearance of the word health in transportation journals is 33.6%.

Rule 74, Rule 76, and Rule 86. Rule 86 is the probability of the intersection of health and machine, which is always followed by the word learning at 92%. Rules 74 and 76 are other variations between health, machine, and learning at 63.5% and 66.9%. The probability of these three words appearing together is 22.2%.

The word zakat is not found at the intersection because the support and confidence values are below 20%. This means that the word zakat only sometimes appears together with the words Health, Transportation, or Machine Learning. Collaborative research on zakat with transportation, health, and machine learning still has a broad scope for exploration. Another meaning is that few publications are still related to the intersection of zakat research with health, transportation, and machine learning.

4 Conclusion

Research collaboration between domains is needed to solve complex problems. This research uses transportation, zakat, and machine learning domains as examples to measure the probability of their intersection with health. The method used is the Association Rule with the FP-Growth algorithm. The results of the Association Rule

consist of support and confidence. Support measures the probability of intersection, while confidence measures the conditional probability if one exists.

The three selected domains are found to have intersections with each other. The words Machine and Learning most often appear together. The words transportation and health have the next most significant intersection. The words Health and Machine Learning are in third place. The word zakat does not indicate an intersection with a minimum intersection limit of 20%. An intersection involving the word Zakat was only found with a minimum intersection probability of 0.6%. This means collaborative research with the zakat domain still has more excellent discovery space than other domains, even though it does not have as many publication references as other domains.

Acknowledgements

This research was conducted in 2023 under an international research grant scheme funded by the Universitas PGRI Madiun from Indonesia and Universiti Teknologi Mara from Malaysia. The research team would like to thank the LPPM and PIKI of Universitas PGRI Madiun, Universiti Teknologi Mara, and Baznas of Madiun City for allowing the research team to add insight and knowledge through research in this scheme. Hopefully, this research can provide benefits for the progress of both nations.

References

[1] L. Budd and S. Ison, "Responsible Transport: A post-COVID agenda for transport policy and practice," *Transp. Res. Interdiscip. Perspect.*, 2020, doi: 10.1016/j.trip.2020.100151.

[2] M. Alfarizi and R. Arifian, "G20 Health Vision in Achieving SDGs 2030: Arranging the Global Health Management Architecture," *Insights Public Heal. J.*, 2022, doi: 10.20884/1.iphj.2022.3.1.5565.

[3] R. Surinach, "Sustainable Development Goals and Health," in *International Encyclopedia of Transportation: Volume 1-7*, 2021.

[4] K. Hartley, "Public Perceptions About Smart Cities: Governance and Quality-of-Life in Hong Kong," *Soc. Indic. Res.*, 2023, doi: 10.1007/s11205-023-03087-9.

[5] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, "Interpretable machine learning: Fundamental principles and 10 grand challenges," *Stat. Surv.*, 2022, doi: 10.1214/21-SS133.

[6] M. O. K. Mendonça, S. L. Netto, P. S. R. Diniz, and S. Theodoridis, "Machine learning," in *Signal Processing and Machine Learning Theory*, 2023.

[7] A. G. Othman and K. H. Ali, "Transportation and quality of life," *Plan. Malaysia*, 2020, doi: 10.21837/PM.V18I13.774.

[8] L. Oluyede, A. L. Cochran, M. Wolfe, L. Prunkl, and N. McDonald, "Addressing transportation barriers to health care during the COVID-19 pandemic: Perspectives of care coordinators," *Transp. Res. Part A Policy Pract.*, 2022, doi: 10.1016/j.tra.2022.03.010.

[9] T. Willem *et al.*, "Risks and benefits of dermatological machine learning health care applications—an overview and ethical analysis," *J. Eur. Acad. Dermatology Venereol.*, 2022, doi: 10.1111/jdv.18192.

[10] S. Rose, "Intersections of machine learning and epidemiological methods for health services research," *International Journal of Epidemiology*. 2020, doi:

10.1093/ije/dyaa035.

[11] I. A. AbdulKareem, M. S. bin Mahmud, M. Elaigwu, and A. F. Abdul Ganiyy, "Mitigating the Effect of Covid-19 on the Society Through the Islamic Social Finance," *J. Manag. Theory Pract.*, 2021, doi: 10.37231/jmtp.2021.2.1.83.

[12] M. Chotib, "Zakat management concept to accelerate health and economic recovery during the covid-19 pandemic," *Open Access Maced. J. Med. Sci.*, 2021, doi: 10.3889/oamjms.2021.7394.

[13] T. Wahyuningsih, Henderi, and Winarno, "Text Mining an Automatic Short Answer Grading (ASAG), Comparison of Three Methods of Cosine Similarity, Jaccard Similarity and Dice's Coefficient," *J. Appl. Data Sci.*, 2021, doi: 10.47738/jads.v2i2.31.

[14] T. Bin Sarwar, N. M. Noor, M. S. U. Miah, M. Rashid, F. Al Farid, and M. N. Husen, "Recommending Research Articles: A Multi-Level Chronological Learning-Based Approach Using Unsupervised Keyphrase Extraction and Lexical Similarity Calculation," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3131470.

[15] M. H. Ahmed, S. Tiun, N. Omar, and N. S. Sani, "Short Text Clustering Algorithms, Application and Challenges: A Survey," *Applied Sciences (Switzerland)*. 2023, doi: 10.3390/app13010342.

[16] H. Wu, Y. Liu, and J. Wang, "Review of text classification methods on deep learning," *Computers, Materials and Continua*. 2020, doi: 10.32604/CMC.2020.010172.

[17] M. Liu, Z. Zheng, B. Qin, Y. Liu, and Y. Li, "Text correlation calculation based on passage-level event representation," *Sci. Sin. Informationis*, 2020, doi: 10.1360/SSI-2019-0272.

[18] A. Ghabayen and B. Ahmed, "Enhancing Collaborative Filtering Recommendation Using Review Text Clustering," *Jordanian J. Comput. Inf. Technol.*, 2021, doi: 10.5455/jjcit.71-1609969782.

[19] A. Himawan, A. Priadana, and A. Murdiyanto, "Implementation of Web Scraping to Build a Web-Based Instagram Account Data Downloader Application," *IJID (International J. Informatics Dev.*, vol. 9, no. 2, pp. 59–65, 2020, doi: 10.14421/ijid.2020.09201.

[20] M. A. Jassim and S. N. Abdulwahid, "Data Mining preparation: Process, Techniques and Major Issues in Data Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1090, no. 1, p. 012053, 2021, doi: 10.1088/1757-899x/1090/1/012053.

[21] V. Agarwal, "Research on Data Preprocessing and Categorization Technique for Smartphone Review Analysis," *Int. J. Comput. Appl.*, vol. 131, no. 4, pp. 30–36, 2015, doi: 10.5120/ijca2015907309.

[22] I. Y. R. Pratiwi, "Hoax news identification using machine learning model from online media in Bahasa Indonesia," *MATRIX J. Manaj. Teknol. dan Inform.*, vol. 12, no. 2, pp. 58–67, 2022, doi: 10.31940/matrix.v12i2.58-67.

[23] A. Ariansyah and M. Kusmira, "Analisis Sentimen Pengaruh Pembelajaran Daring Terhadap Motivasi Belajar Di Masa Pandemi Menggunakan Naive Bayes Dan Svm," *Fakt. Exacta*, vol. 14, no. 3, p. 100, 2021, doi: 10.30998/faktorexacta.v14i3.10325.

[24] Y. Apridonal M, W. Choiriah, and A. Akmal, "Penerapan Data Mining Menggunakan Metode Assiciation Rule Dengan Algoritma Apriori Untuk Analisa Pola Penjualan Barang," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 193–198, 2019, doi: 10.33330/jurteksi.v5i2.362.

[25] M. Fauzy, K. R. Saleh W, and I. Asror, "Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung," *J. Ilm. Teknol. Infomasi Terap.*, vol. 2, no. 3, 2016, doi: 10.33197/jitter.vol2.iss3.2016.111.