JOIV : Int. J. Inform. Visualization, 8(1) - March 2024 395-400

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Cluster Analysis of Japanese Whiskey Product Review Using K-Means Clustering

Deden Witarsyah [a,*], Moh Adli Akbar [a], Villy Satria Praditha [a], Maria Sugiat [b]

[a] School of Industrial and System Engineering, Telkom University, Dayeuhkolot, Bandung, 40257, Indonesia
[b] Distance Learning Program of Magister Management, Telkom University, Dayeuhkolot, Bandung, 40257, Indonesia
Corresponding author: [*]dedenw@telkomuniversity.ac.id

***Abstract*— Since 2008, the Japanese whiskey business has grown steadily. Overall, the whiskey market (at factory price) is expected to reach $2.95 billion in 2019, accounting for 8.6 percent of the entire alcoholic beverage industry. The rise in popularity of Japanese whiskey is associated with the country's growing international reputation. Founded 1985 as an independent bottler, Master of Malt was the first company to service clients who ordered single malt whiskey through the mail-order system. Master of Malt's omnichannel approach encompasses all channels available to the company. Known as their 'omnichannel,' this refers to the organization's capability to provide speed and precision from any place at any time. As their brand has grown over the years, they have used various marketing strategies, including a website redesign and rebuild that involved the creation of all relevant content and designing and constructing landing pages for their website. Following a clustering technique, we discovered that the data is being divided into four distinct groups and that these clusters may serve as a recommender system based on the occurrence of terms in each of the categories. Our summarizing component combined phrases related to the exact subtopics and provided users with a concise summary and sentimental information about the group of phrases.**

***Keywords*— K-Means clustering; Japanese whiskey; omnichannel.**

Manuscript received 12 Apr. 2022; revised 23 Jan. 2023; accepted 2 Sep. 2023. Date of publication 31 Mar. 2024.

I. Introduction

According to the Ministry of Finance, the Japanese whiskey industry has consistently increased since 2008. On an annual basis, the overall whiskey market (at factory price) was predicted to reach $2.95 billion in 2019, accounting for 8.6% of the total alcoholic beverage market [1]. The increase in consumption is believed to be linked to the growing international recognition of Japanese whiskey as a premium product of superior quality. In 1985, Master of Malt was established as an independent bottler, and it was the first firm in the world to handle customers who placed orders for single malt whiskey over the mail. When a firm can deliver speed and precision at any time and from any place, it is called its omnichannel capability. To provide a smooth purchasing experience, the omnichannel strategy incorporates all accessible channels [2]. Master of Malt redesigned their website in 2010, and they've continued to invest in two things since then: people and technology. Over the years, they have used several methods to promote their brand, including a further redesign and rebuild that included producing all critical content and designing and developing landing pages for their website.

With customer reviews, there are various things we can do to enhance the overall quality of our products. To check if we could identify any topics, we might calculate the frequencies of terms or word combinations that appear in a subject and see if we could find any issues that way. Modern natural language processing (NLP) models have advanced to a high level of sophistication. They can be trained to determine that two texts are semantically identical even when the wording is significantly different. [3]. A model of this kind was developed [4]–[6] and utilized to group customer input into subjects, resulting in a more compelling customer feedback monitoring process.

II. Materials and Method

An in-depth pattern of grouped item categories was discovered by applying the k-means clustering technique in this research [4]. This investigation results in an automated classification model that can effectively categorize the items. An online shop website in Malaysia has been used to group e-

commerce items. This research gives a step-by-step cluster analysis utilizing k-means clustering to group the products. The findings reveal that e-commerce goods were divided into three clusters, each with its features. This clustering approach can act as a recommender system based on the number of times a word appears in each cluster.

This research [5] focuses on social network analysis to understand the behavior and structure of a social network. The research and study of data mining communities and social network analysis have multiplied. Interactions on such social media sites generate massive data since billions of active users keep their accounts current. Academics and industry must understand and anticipate the behavior of such online social networks. They are focusing on Twitter data in this research. In this paper, tweets are retrieved, processed, evaluated, and represented geographically. The study's primary purpose is to show users' tweets in a particular region and to visualize grouped tweets based on geographical information. To cluster the tweets, utilize the k-means clustering technique. The number of clusters in the k-means algorithm is specified and chosen at random or derived from the data set using the Elbow approach. The elbow technique runs the k-means clustering algorithm on a dataset for a range of values and calculates the sum of squared errors. Using the Elbow approach, they utilize the k-means clustering technique based on the geotagged information after determining the number of clusters necessary for the dataset.

This study [6] focuses on patent trends. They look at patent trends via the lens of text clustering. They create a vector space for patent documents based on the TF-IDF (Inverse Document Frequency) model and apply the k-means method to text clustering. This study develops a PTA software system that analyzes and mines patent data using artificial intelligence and big data technologies. They concentrate on patent text processing in Chinese in this work. As a result, the CNKI patent database is the primary system data source. After acquiring the patent information data, they extract the abstract and sovereign text content. However, the extracted text cannot be utilized since it is a whole paragraph. Word segmentation aims to divide patent text phrases into Chinese vocabulary segments. In the Chinese context, various terms known as "stop words" have no effect or link to the original meaning of the patent. While segmenting the patent text, stop words must be filtered. As a result, they transform unstructured text data types into numerical values that may be processed using the Vector Space Model (VSM). They employ the conventional k-means approach to find the initial center. The silhouette coefficient is used in this study to identify the initial cluster center and the number of clusters. They produce many cluster centers randomly and then compute the silhouette coefficient of each cluster center's clustering findings. The system has four primary modules: the patent search module, the patent categorization module, the patent trend prediction module, and the patent cluster analysis model. The text clustering result may be represented in various ways, including E-Charts diagrams.

CRISP-DM (Cross Industry Standard Process for Data Mining) is a standardized paradigm. The standard specifies six distinct stages that must be completed one or more times [7]. It has a global presence is one of the most often utilized analytics models. It looks for datasets' patterns, trends, and correlations [8]. As shown in Figure 1, those are the steps followed in CRISP-DM.

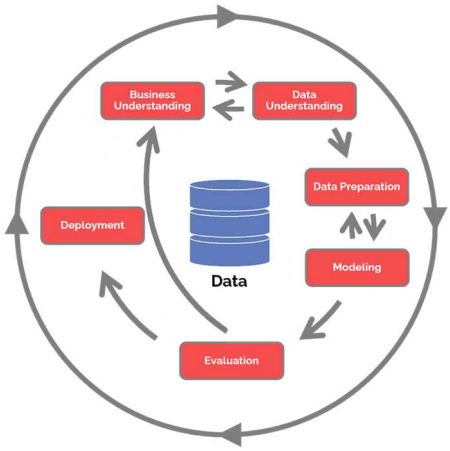

Fig. 1 Cross-Industry Standard Process for Data Mining

A. Business Understanding

The business comprehension phase is concerned with determining the precise objectives and needs for data mining. In this section, you'll spell out the business success criteria you'll use to assess if the project was a financial success. This phase creates a primary method for achieving business and data mining objectives [8].

B. Data Understanding

As part of the data comprehension process, an effort is made to get a preliminary overview of the available data and its quality. This entails determining if all of the needed data (to achieve the data mining objectives) is accessible and devising a strategy for deciding which data is required. First, you explain the collected data, including the format, amount, and any additional surface characteristics detected. Choose if the data obtained meets your needs. During this trial, you may encounter unexpected program (or hardware) constraints, such as limits on the number of cases or fields or an inability to read the data formats of your sources.

C. Data Preparation

One of the most critical and time-consuming data mining components is data preparation. Data preparation is expected to require 50–70 percent of a project's time and effort. The data preparation step generates a final data collection that will be the foundation for the following modeling phase. This includes adding derived attributes, which are new attributes that combine existing characteristics in the same record [9].

D. Modelling

The analytical heart of the data mining process is modeling. This is where modeling approaches are chosen and used. Many modeling approaches require certain data assumptions, such as that all characteristics have uniform distributions or that no missing values are permitted. There are often many factors that may be modified with every modeling program [9].

E. Deployment

After data preparation, model development, and model verification, the chosen model is employed in the deployment

step. The acquired information must now be processed and given to the client to be used without difficulty. A specific monitoring process plan is required for the project to monitor the deployment of the data mining result(s). This strategy considers the unique kinds of deployment. The appropriate planning of a maintenance strategy aids in avoiding unduly extended periods of inaccurate data mining result consumption [10].

F. Clustering

Clustering is categorizing a population or collecting data points and allocating them to cluster groups based on their commonalities [10]. Clustering is the technique of arranging data items into disconnected groups such that data within the same cluster are similar, but data from other sets differs [11]. Clustering is one of the essential tasks in exploratory data analysis [12]. It involves the unsupervised categorization of patterns into groups. Clustering's primary purpose is to get insight into data.

G. K-Means Algorithm

The K-means clustering technique is a partitioning, clustering, and unsupervised learning approach based on centroid values [13]. Because it is simple and easy to implement, versatile in almost every aspect of the algorithm (initialization, distance function, termination criterion, etc.), and has a time complexity that is linear in N, D, and K (in general, D N and K N), the K-Means algorithm is the most widely used partitional clustering algorithm. K-means, on the other hand, have several drawbacks [14]. For starters, it requires the number of clusters, K, to be specified; it can only detect compact, hyper spherical clusters that are well separated; and it is sensitive to noise and outlier points, as even a few such points can significantly influence the means of their respective clusters [14].

$$E = \sum_{i=1}^{k} \quad \sum_{p \in C_i} \quad dist(p, c_i) \qquad (1)$$

E is the sum of the squared error for all objects in the data set; $dist(p, c_i)$ r esents the cluster; p is the point in space representing a given object; c_i is the centroid of cluster C_i [9].

H. TF-IDF Algorithm

TF-IDF (term frequency-inverse document frequency) is a typical term weighting system for representing textual documents as vectors. The TF/IDF is used for classification, visualization, retrieval, clustering, etc. The weight xij refers to the word t in document di and is often a three-part product.; one which depends on the presence or frequency of t_j in d_i [15]. TF-IDF's most common weighting is

$$x_{ij} = TF_i \,.\, IDF_j \,.\, (\sum_i (TF_{ij}\, IDF_j)^2)\text{-}½ \qquad (2)$$

III. Result and Discussion

Data exploration is the first step in the data mining process; during this phase, the dataset will be accessed and processed following the algorithm's specifications. In this study, we used RapidMiner to process the data, and we will analyze it using TF-IDF and K-Means Clustering to find patterns.

A. Data Collection

The data for this study was derived from MasterofMalt's database of Japanese Whiskey Products, which is available online. There are 1121 entries in this collection for Japanese whiskey reviews. RapidMiner can use the data since it is already in excellent condition.

Bottle_name	Brand	Title	Review_Content
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Overpriced di...	Dull taste. High pr...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Delicious	Delicious! sugare...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Good for begi...	I am not a whisky ...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Yamazaki Tut...	This is a terrible Y...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Very Nice	First time and I lik...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Unworthy of a...	I canâ€™t believe...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Japanese Ya...	My friend brought ...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Delicious.	Nice. I like some ...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Yamazaki Dis...	Stands against Gl...
The Yamazaki Single Malt Whisky - Distillerâ€™s Reserve	Yamazaki	Can't Wait	Recommended b...

Fig. 2 Review of Japanese Whiskey Products

As seen in Figure 2, the data has four attributes: Bottle name, Brand, Title, and Review content. Each property is described in detail in Table 1 below.

TABLE I
ATTRIBUTE DESCRIPTION

Attributes	Descriptions
Bottle Name	Each brand has its bottle name. Additionally, this property provides the age of each whiskey.
Brand	The name of the Japanese distillery that produces the whiskey.
Title	The reviews' titles
Review Content	The Product Review's Description

B. Data Preprocessing

This section contains the data preparation stage. This section will explain the data preparation step, which includes converting a nominal attribute to text, choosing the attribute that will be utilized in the clustering process, and converting the data into tokens to be used in the clustering process [16]. This process in RapidMiner is shown in Figure 3.

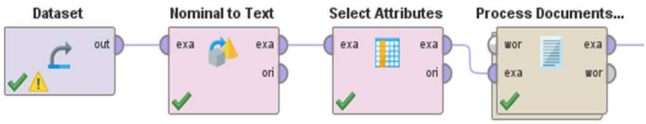

Fig. 3 Data Preprocessing Steps

1) Nominal to Text: This step converts the nominal characteristics type from numeric to textual. This step will convert all nominal properties to string attributes, which will be helpful later. Because it can only be provided with strings and not nominal characteristics, changing the nature of these attributes will be beneficial later in the tokenization process.

2) Select Attributes: Following converting the attribute type to string, the following step is to pick the attributes to use in the clustering process. In this study, we shall discuss two characteristics: The title and Review_content attributes include data about the product reviews. As shown in Figure 4, the dataset that has been applied Nominal to text and selects the attributes process has only two attributes: Title and review content.

Title	Review_Content
Overpriced dissapointment	Dull taste. High price. No finish. Over-hyped and disapp...
Delicious	Delicious! sugared red fruits and sweet with a morish, c...
Good for beginners. i know c...	I am not a whisky expert but i really love the taste. The e...
Yamazaki Tutorial	This is a terrible Yamazaki. Very young, unsherried, no f...
Very Nice	First time and I like it - fresh but not thin and a sweet de...
Unworthy of al the hype over ...	I canâ€™t believe all these commenters who are fawni...

Fig. 4 Dataset after selecting the attributes.

3) Tokenization: This section will divide the text of a document into tokens. Tokenization is an NLP (Natural Language Processing) approach that splits a text into tokens, words, or phrases. These tokens are utilized in further processing. Tokenization aids in the identification of significant terms. To separate tokens, white space, line breaks, or punctuation marks are used [17]. White spaces and punctuation are not included in the tokens generated. It is challenging to tokenize documents that lack white spaces. Tokenizers ensure the document's dependability. This process will produce tokens of a single word, the best choice, before eventually constructing the word vector [18].

4) Filtering Stopwords: Stop words often have no special meaning, such as a, an, the, of, and so on. A list of stop words is created and compared to the original content. If a term is on this list found, it is removed from the original text. When matching queries to documents, stop words have no value [19]. Consequently, removing the stop words did not impact the page's semantics. We will utilize the RapidMiner stop words list in this study since it already has all the stop words in English.

5) Transform Cases: This section will transform all the characters to lower cases.

Word	Attribute Name	Total Occurences	Document Occurences
absolutely	absolutely	45	45
alcohol	alcohol	45	39
amazing	amazing	73	63
auction	auction	57	40
bad	bad	37	34
balanced	balanced	41	37
beautiful	beautiful	37	35
bit	bit	86	70
blend	blend	75	59
blended	blended	53	44
bottle	bottle	297	225
bottles	bottles	75	62
bought	bought	69	63

Fig. 5 Clustering result

Bottle, Japanese, whiskey, smooth, and friendly are some of the most frequently used terms, each having over 200 occurrences. A new property displays each term and its total occurrences following the tokenization process. As shown in Figure 5, Cluster 1 has 137 items, whereas Cluster 2 contains 124. Cluster 3 has 193 items, whereas Cluster 4 contains 667.

The authors focus more on the modules and techniques and perform statistical calculations while processing and exploring data sets using RapidMiner. The authors' first step was to implement the TF-IDF algorithm. The Term Frequency-Inverse Document Frequency (TF-IDF) technique effectively estimates the weight of each frequently used word. This approach computes the term frequency (TF) and the Inverse Document Frequency (IDF) values for each token (word) in the corpus [18]. As a result, the initial step for authors is to check for the TF value to identify how often a word occurs in a manuscript. The more frequently a term appears, the more valuable it becomes. Figure 6 depicts the outcomes of the five stages outlined above, including the IDF scoring. After the frequency term was determined, the TF findings of this research consisted of 86 columns, indicating that 86 terms satisfied the minimal support of 20%.

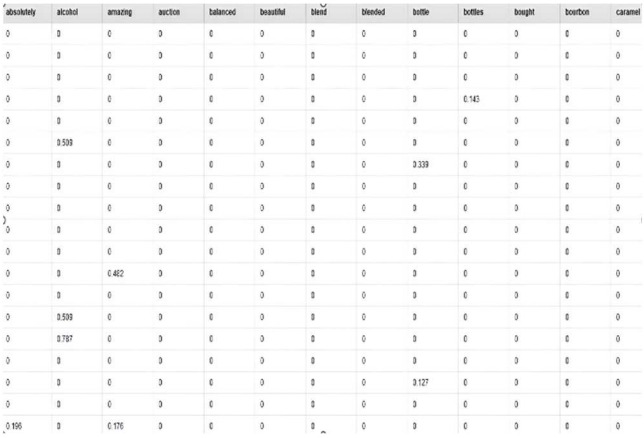

Fig. 6 Dataset after Tokenize and Filtering Stopwords.

C. Clustering Process

After the data preparation and preprocessing are done, the next is to do k-means clustering. In this research, the author is using the k-means method from RapidMiner. The k-means algorithm determines a set of k clusters and assigns each example to precisely one cluster [20]. A cluster is defined by the position of the center in the n-dimensional space of the Attributes of the ExampleSet [21-25]. Some researchers use process clustering in their research to obtain valid results [26-30]. The similarity between Examples is based on a distance measure between them. Figure 7 shows 10 sample keywords and which cluster they belong to.

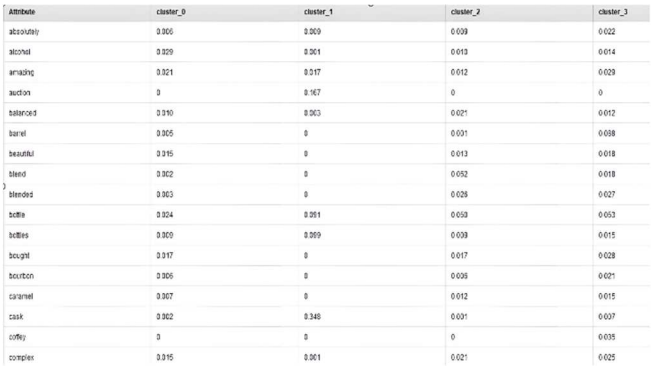

Fig. 7 Clustering result

This cluster has some intriguing data. Bottle, price, Japanese, Yamazaki, year, nose, time, value, bottles, Japan, and so on are all found in Cluster 1. Malts, review, lottery, year, blend, fantastic, discover, and other terms appear in Cluster 2. Smooth, sweet, drink, flavor, Japanese, fantastic, vanilla, bottle, alcohol, love, beautiful, and other words

appear in Cluster 3. Whiskey, excellent, smooth, incredible, good, ultimately, love, flavors, and other words appear in Cluster 4—some words, such as Japanese, years, bottles, and fruity, fit into several clusters.

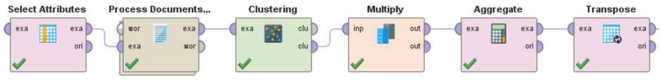

Fig. 8 RapidMiner Modules

Figure 8 shows each RapidMiner module, from preprocessing through clustering. Following the clustering procedure, three further phases are added.

- Multiply: This method will duplicate the data obtained from the clustering process, resulting in two copies.
- Aggregate: This method will perform the average function on each of the columns in the dataset.
- Transpose: The data from the previous step will be transposed in this process so that the current rows will become columns.

To see the result on each cluster more clearly, the authors will show each cluster in Word Cloud. A word cloud is a collection, or cluster, of words in various sizes on a computer screen. The larger and stronger the word appears, the more often it is referenced inside a specific text and the greater the importance of the term is deemed to be. Figure 9 displays a word cloud for Cluster 1, Figure 10 shows a word cloud for Cluster 2, Figure 11 displays a word cloud for Cluster 3, and Figure 12 shows a word cloud for Cluster 4.

Fig. 9 Words Cloud of Cluster 1

Cluster 1 concentrates on each company and the excellent feedback from their clients. This cluster explains the characteristics of Japanese whiskey from the viewpoint of consumers who do not reside in Japan. MasterofMalt might utilize this kind of information to improve the quality of their whiskey.

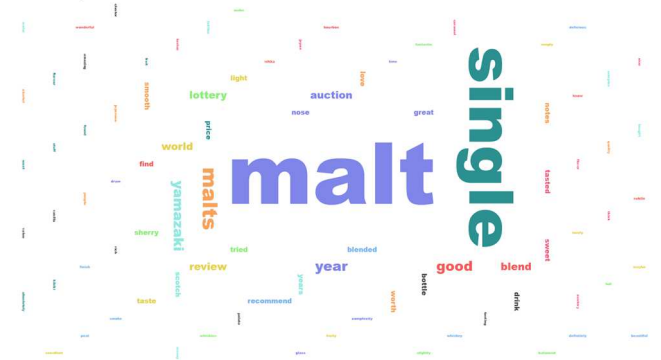

Fig. 10 Words Cloud of Cluster 2

Cluster 2 contains information regarding the bottle name, The Yamazaki Single Malt Whiskey. This cluster displays what foreign consumers have to say about this specific product. MasterofMalt might utilize Cluster 2 to provide the brand with further in-depth analysis.

Fig. 11 Words Cloud of Cluster 3

Cluster 3 offers information regarding the good and bad comments that were left on the website. This cluster indicates that the majority of customers are pleased with the product. MasterofMalt might use the information from Cluster 3 to promote one of the favorable reviews and showcases to potential customers by using the information from Cluster 3.

Fig. 12 Words Cloud of Cluster 4

Cluster 4 provides information regarding whiskey and general reviews of other products. This cluster displays the overall ratings of Japanese whiskey, which has received favorable feedback from international buyers. As a result of this information, MasterofMalt may be able to encourage more manufacturers to market their goods internationally via the use of MasterofMalt.

IV. Conclusion

This research offers a strategy for cluster analysis of Japanese whiskey goods based on consumer feedback. Our summarizing component worked by grouping phrases about the exact subtopics and providing users with a concise summary and sentiment information. To overcome the subtopic issue, we used a clustering technique. The experimental findings show that the data are classified into four unique categories. The result of that is word clouds representing 4 clusters. These clusters show unique reviews about whiskey. This clustering approach may function as a recommender system based on words in each cluster.

REFERENCES

[1] M. Otsuka, "Market Overview - Whiskey is Up in Japan," 2020. [Online]. Available: https://www.fas.usda.gov/data/japan-market-overview-whiskey-japan

[2] A. Mosquera, C. Olarte Pascual, and E. Juaneda Ayensa, "Understanding the customer experience in the age of omni-channel shopping," Revista ICONO14 Revista científica de Comunicación y Tecnologías emergentes, vol. 15, no. 2, pp. 92–114, Jul. 2017, doi:10.7195/ri14.v15i2.1070.

[3] A. Yadav, A. Patel, and M. Shah, "A comprehensive review on resolving ambiguities in natural language processing," AI Open, vol. 2, pp. 85–92, 2021, doi: 10.1016/j.aiopen.2021.05.001.

[4] N. M. N. Mathivanan, N. A. Md. Ghani, and R. M. Janor, "Analysis of K-Means Clustering Algorithm: A Case Study Using Large Scale E-Commerce Products," 2019 IEEE Conference on Big Data and Analytics (ICBDA), Nov. 2019, doi:10.1109/icbda47563.2019.8987140.

[5] N. Garg and R. Rani, "Analysis and visualization of Twitter data using k-means clustering," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Jun. 2017, doi:10.1109/iccons.2017.8250547.

[6] R. Liang, Y. Li, X. Chen, and J. Chen, "Patent Trend Analysis through Text Clustering based on K-Means Algorithm," 2020 International Symposium on Computer Engineering and Intelligent Communications (ISCEIC), Aug. 2020, doi:10.1109/isceic51027.2020.00032.

[7] H. Sanjurjo González, I. Pastor López, P. García Bringas, H. Quintián, and E. Corchado, Eds., 16th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2021). Springer International Publishing, 2022. doi:10.1007/978-3-030-87869-6.

[8] J. S. Saltz, "CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps," 2021 IEEE International Conference on Big Data (Big Data), Dec. 2021, doi:10.1109/bigdata52589.2021.9671634.

[9] J. Han, M. Kamber, and J. Pei, "Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)," 2011.

[10] Vishal, A. Kumar, and Dr. P. Saini, "Use of K-Means Clustering Method for Books Data in Acharya Raghuveer Library, Central University of Himachal Pradesh, Dharamshala, India," *Library Philosophy and Practice (e-journal)*, Nov. 2021, Accessed: Jan. 26, 2022. [Online]. Available: https://digitalcommons.unl.edu/libphilprac/6655

[11] S. P. Tamba, M. D. Batubara, W. Purba, M. Sihombing, V. M. Mulia Siregar, and J. Banjarnahor, "Book data grouping in libraries using the k-means clustering method," Journal of Physics: Conference Series, vol. 1230, no. 1, p. 012074, Jul. 2019, doi: 10.1088/1742-6596/1230/1/012074.

[12] M. E. Celebi and H. A. Kingravi, "Deterministic Initialization of The K-Means Algorithm Using Hierarchical Clustering," International Journal of Pattern Recognition and Artificial Intelligence, vol. 26, no. 07, p. 1250018, Nov. 2012, doi: 10.1142/s0218001412500188.

[13] J. Gola et al., "Advanced microstructure classification by data mining methods," Computational Materials Science, vol. 148, pp. 324–335, Jun. 2018, doi: 10.1016/j.commatsci.2018.03.004.

[14] J. Han, M. Kamber, and J. Pei, "Introduction," Data Mining, pp. 1–38, 2012, doi: 10.1016/b978-0-12-381479-1.00001-0.

[15] "TF–IDF," Encyclopedia of Machine Learning, pp. 986–987, 2011, doi: 10.1007/978-0-387-30164-8_832.

[16] A. Larasati, M. Farhan, P. Rahmawati, J. Sayono, A. Purnomo, and E. Mohamad, "Exploring the relationship among polarity, subjectivity, and clusters characteristics of visitor review on tourist destination in Malang, Indonesia," Community Empowerment through Research, Innovation and Open Access, pp. 22–27, Jul. 2021, doi:10.1201/9781003189206-5.

[17] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), Oct. 2019, doi: 10.1109/icic47613.2019.8985884.

[18] M. Darwis, G. Tri Pranoto, and Y. Eka Wicaksana, "Implementation of TF-IDF Algorithm and K-mean Clustering Method to Predict Words or Topics on Twitter," 2020.

[19] R. Nainggolan and Eviyanti Purba, "The Cluster Analysis of Online Shop Product Reviews Using K-Means Clustering," Data Science: Journal of Computing and Applied Informatics, vol. 4, no. 2, pp. 111–121, Jul. 2020, doi: 10.32734/jocai.v4.i2-2855.

[20] K. Orkphol and W. Yang, "Sentiment Analysis on Microblogging with K-Means Clustering and Artificial Bee Colony," International Journal of Computational Intelligence and Applications, vol. 18, no. 03, Sep. 2019, doi: 10.1142/s1469026819500172.

[21] H. Rumapea, M. Zarlis, S. Efendy, and P. Sihombing, "Improving Convective Cloud Classification with Deep Learning: The CC-Unet Model", Int. J. Adv. Sci. Eng. Inf. Technol., vol. 14, no. 1, pp. 28–36, Feb. 2024.

[22] A. A. A. Al-zanam, O. H. A. E. Hussein Alhomery, and C. P. Tan, "Mental Health State Classification Using Facial Emotion Recognition and Detection", Int. J. Adv. Sci. Eng. Inf. Technol., vol. 13, no. 6, pp. 2274–2281, Dec. 2023.

[23] E. A. Anggari, "Assessing the Accuracy of Land Use Classification Using Multi-spectral Camera From LAPAN-A3, Landsat-8 and Sentinel-2 Satellite: A Case Study in Probolinggo-East Java", Int. J. Adv. Sci. Eng. Inf. Technol., vol. 13, no. 5, pp. 1622–1627, Oct. 2023.

[24] Z. Rustam, S. Hartini, F. Novkaniza, J. Pandelaki, R. Hidayat, and M. Ezziyyani, "An Efficient and Robust Ischemic Stroke Detection Using a Combination of Convolutional Neural Network (CNN) and Kernel K-Means Clustering", Int. J. Adv. Sci. Eng. Inf. Technol., vol. 13, no. 3, pp. 969–974, Jun. 2023.

[25] V. Ratnasari and A. T. R. Dani, "Mapping the Provincial Food Security Conditions in Indonesia Using Cluster Ensemble-Based Mixed Data Clustering-Robust Clustering with Links (CEBMDC-ROCK)", Int. J. Adv. Sci. Eng. Inf. Technol., vol. 13, no. 2, pp. 611–617, Mar. 2023.

[26] M V, Ananthapadmanabha, et al. "Cluster Ensemble Method and Convolution Neural Network Model for Predicting Mental Illness". International Journal on Advanced Science, Engineering and Information Technology, vol. 13, no. 1, Feb. 2023, pp. 392-8, doi:10.18517/ijaseit.13.1.17498.

[27] Banjarnahor, Evander, et al. "Analysing Kinship in Severe Acute Respiratory Syndrome Coronavirus 2 DNA Sequences Based on Hierarchical and K-Means Clustering Methods Using Multiple Encoding Vector". International Journal on Advanced Science, Engineering and Information Technology, vol. 12, no. 6, Nov. 2022, pp. 2237-4, doi:10.18517/ijaseit.12.6.15582.

[28] Ilham, Ahmad Juniar, et al. "Data Hiding Scheme Based on Quad General Difference Expansion Cluster". International Journal on Advanced Science, Engineering and Information Technology, vol. 12, no. 6, Nov. 2022, pp. 2288-96, doi:10.18517/ijaseit.12.6.16002.

[29] Yanto, Iwan Tri Riyadi, et al. "Soft Set Parametric-Based Data Clustering for Building Data Set". International Journal on Advanced Science, Engineering and Information Technology, vol. 12, no. 5, Sept. 2022, pp. 1882-6, doi:10.18517/ijaseit.12.5.16010.

[30] Joo, Yongho, and Gilsang Yoo. "Study on Automobile Culture of Developed and Emerging Countries in Asia Using Text Mining Analysis on Social Media". International Journal on Advanced Science, Engineering and Information Technology, vol. 11, no. 1, Feb. 2021, pp. 51-56, doi:10.18517/ijaseit.11.1.14078.