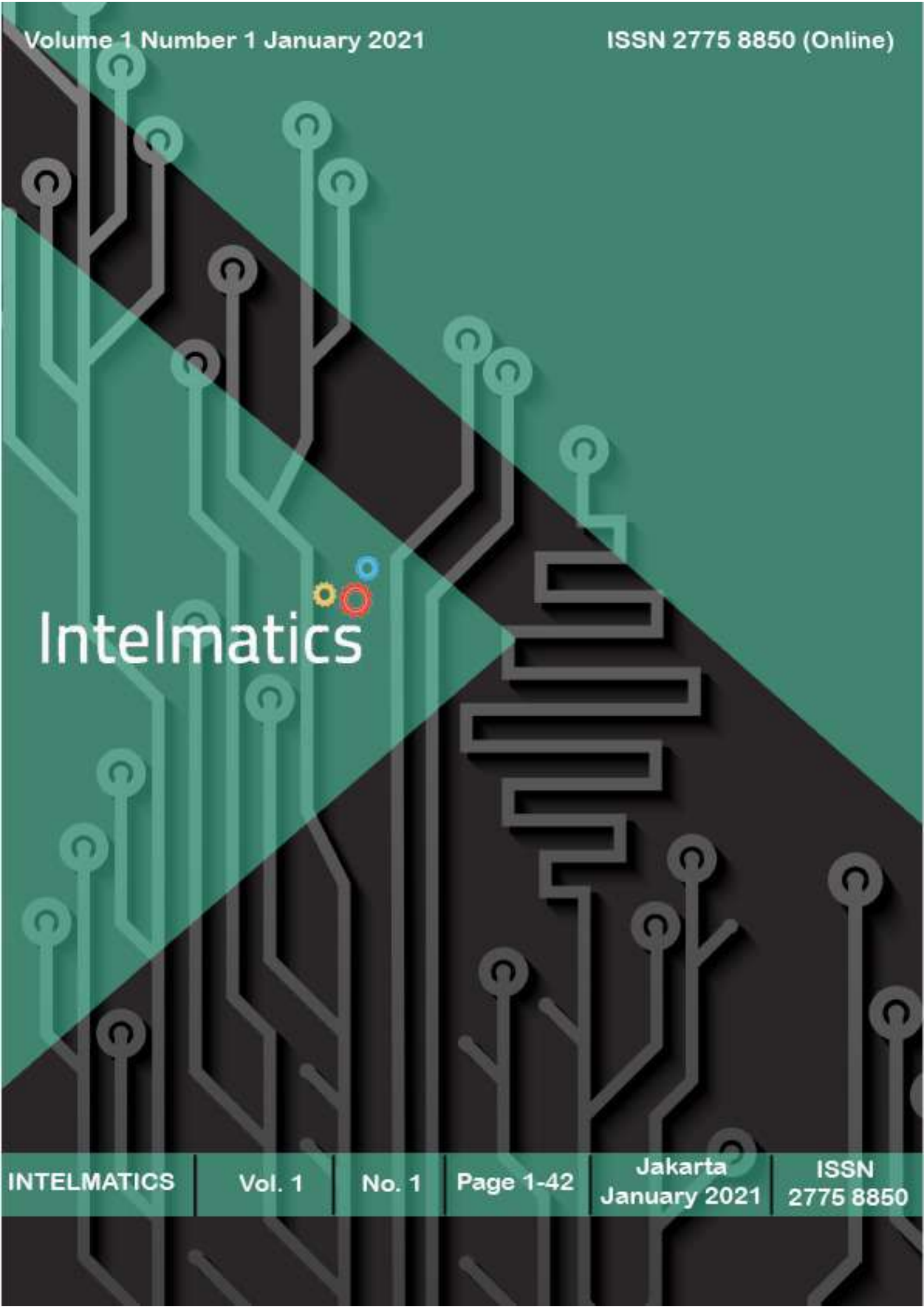
Volume 1 Number 1 January 2021 ISSN 2775 8850 (Online)



Intelmatics

INTELMATICS | Vol. 1 | No. 1 | Page 1-42 | Jakarta January 2021 | ISSN 2775 8850

Editorial Team

(https://sinta.kemdikbud.go.id/authors/profile/5974359) (https://www.scopus.com/authid/detail.uri?authorId=56703180200) (https://scholar.google.com/citations?user=ZogHjrsAAAAJ&hl=id&oi=sra)

Is Mardianto

Fakultas Teknologi Industri, Universitas Trisakti, Indonesia

(https://sinta.kemdikbud.go.id/authors/profile/5974114) (https://www.scopus.com/authid/detail.uri?authorId=57200111682) (https://scholar.google.com/citations?user=xN3I7OEAAAAJ&hl=id&oi=sra)

Syandra Sari

Fakultas Teknologi Industri, Universitas Trisakti, Indonesia

(https://sinta.kemdikbud.go.id/authors/profile/5987050) (https://www.scopus.com/authid/detail.uri?authorId=23393545800) (https://scholar.google.com/citations?user=kGTdm4wAAAAJ&hl=id&oi=sra)

Gatot Budi Santoso

Fakultas Teknologi Industri, Universitas Trisakti, Indonesia

(https://sinta.kemdikbud.go.id/authors/profile/6107942) (https://www.scopus.com/authid/detail.uri?authorId=7801436887) (https://scholar.google.com/citations?user=FKmQ8ykAAAAJ&hl=id&oi=sra)

Laser Narindro

Fakultas Teknologi Industri, Universitas Trisakti, Indonesia

(https://www.scopus.com/authid/detail.uri?authorId=57207937646) (https://scholar.google.com/citations?user=3IYDcrkAAAAJ&hl=id&oi=ao)

Abdul Rochman

(https://sinta.kemdikbud.go.id/authors/profile/6647458) (https://scholar.google.com/citations?user=2VaX60EAAAAJ&hl=id&oi=ao)

Fakultas Teknologi Industri, Universitas Trisakti, Indonesia

Yunia Ningsih

Fakultas Teknologi Industri, Universitas Trisakti, Indonesia

TEMPLATE

(https://docs.google.com/document/d/1UuCljmUhwWP4RFazWiJ9M6gjKB9z28wf/edit?usp=sharing&ouid=104493949290278026636&rtpof=true&sd=true)

MAIN MENU

Click here to **Submit** (/index.php/intelmatics/about/submissions#onlineSubmissions)

1. Author Guideline (https://e-journal.trisakti.ac.id/index.php/intelmatics/about/submissions#authorGuidelines)

2. Focus and Scope (https://e-journal.trisakti.ac.id/index.php/intelmatics/FocusandScope)

3. Publication Ethics (https://e-journal.trisakti.ac.id/index.php/intelmatics/PublicationEthics)

4. Peer Review Process (https://e-journal.trisakti.ac.id/index.php/intelmatics/PeerReviewProcess)

4. Editorial Team (https://e-journal.trisakti.ac.id/index.php/intelmatics/EditorialTeam)

5. Reviewer Team (https://e-journal.trisakti.ac.id/index.php/intelmatics/ReviewerTeam)

6. Plagiarism Check (https://e-journal.trisakti.ac.id/index.php/intelmatics/PlagiarismPolicy)

7. Unique Visits (https://statcounter.com/p12931691/summary/?account_id=5116944&login_id=5&code=08a88f090f134bf5faa7bf1c42f8238c&guest_login=1)

8. Journal History (https://e-journal.trisakti.ac.id/index.php/intelmatics/JournalHistory)

10. Journal Business Model (https://e-journal.trisakti.ac.id/index.php/intelmatics/JournalBusinessModel)

11. Author Index (https://e-journal.trisakti.ac.id/index.php/intelmatics/search/authors)

12. Publication Frequency (https://e-journal.trisakti.ac.id/index.php/intelmatics/PublicationFrequency)

13. Reviewer/EditorJoin (https://e-journal.trisakti.ac.id/index.php/intelmatics/ReviewerEditorJoin)

14. Archiving (https://e-journal.trisakti.ac.id/index.php/intelmatics/Archiving)

15. . Mailing Address (https://e-journal.trisakti.ac.id/index.php/intelmatics/MailingAddress)

16. Retraction Policy (https://e-journal.trisakti.ac.id/index.php/intelmatics/RetractionPolicy)

17. Copyright Notice (https://e-journal.trisakti.ac.id/index.php/intelmatics/CopyrightNotice)

18. Open Access Policy (https://e-journal.trisakti.ac.id/index.php/intelmatics/OpenAccessPolicy)

19. Indexing and Abstracting (https://e-journal.trisakti.ac.id/index.php/intelmatics/IndexingandAbstracting)

(https://sinta.kemdikbud.go.id/journals/profile/14451)

(https://scholar.google.com/citations?user=Swy3MjgAAAAJ&hl=id&authuser=6)

(https://garuda.kemdikbud.go.id/journal/view/32717)

(https://search.crossref.org/search/works?q=2775-8850&from_ui=yes)

(https://portal.issn.org/resource/ISSN/2775-8850)

(https://app.dimensions.ai/discover/publication?search_mode=content&and_facet_source_title=jour.1408685)

(https://www.citefactor.org/search/keywords/journals/Intelmatics)

(https://scite.ai/journals/intelmatics-R5Y34)

TOOLS

REFERENCE MANAGEMENT TOOLS

APPLICATION PLAGIARISM CHECKER

LANGUAGE

English (https://e-journal.trisakti.ac.id/index.php/intelmatics/user/setLocale/en?source=%2Findex.php%2Fintelmatics%2FEditorialTeam)

INFORMATION

For Readers (https://e-journal.trisakti.ac.id/index.php/intelmatics/information/readers)

For Authors (https://e-journal.trisakti.ac.id/index.php/intelmatics/information/authors)

For Librarians (https://e-journal.trisakti.ac.id/index.php/intelmatics/information/librarians)

UNIQUE_VISITS

00015832 (https://statcounter.com/)

(https://e-journal.trisakti.ac.id/index.php/intelmatics/about/aboutThisPublishingSystem)

Published: 2024-08-30

Articles

Analysis Of Topic Movement & Conversation Membership On Twitter Using K-Means Clustering (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/21002)

Agung Sediyono, Josua Valentino Hutagalung, Binti Solihah

Pdf (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/21002/11976)

Application of IOT Technology in The Control of Organic Waste Processing Machines with PT100 Sensors and Heaters for Fertilizer Healing and Animal Feeding (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/20913)

Ricardo Dharma, Gatot Budi Santoso, Is Mardianto

Pdf (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/20913/11977)

Implementation Enterprise Resource Planning (ERP) ODOO Version 15.0 Manufacturing Module at CV. Razzaq Berkah Mulia (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/21073)

Ibnu Fajar Setiawan, Teddy Siswanto, Dedy Sugiarto

PDF (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/21073/11995)

Design of Study Program Performance Dashboard using Streamlit (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/20643)

Moch Ilham A Ilham, Teddy Siswanto, Dedy Sugiarto

96-100

PDF (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/20643/11996)

Abstract: 310 | PDF downloads:149

COMPARATIVE SENTIMENT ANALYSIS OF VISITOR REVIEWS FOR WATERBOM BALI TOURIST ATTRACTION ON TRIPADVISOR SOCIAL MEDIA USING RANDOM FOREST AND NAÏVE BAYES CLASSIFICATION (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/19278)

Hilmi Abdul Gani Hilmi, Binti Solihah, Syandra Sari
101-109

PDF (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/19278/12696)

Abstract: 81 | PDF downloads:38

PERFORMANCE COMPARISON OF TWITTER SENTIMENT ANALYSIS USING FASTTEXT SVM AND TF-IDF SVM: A CASE STUDY ON ELECTRIC MOTORCYCLES (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/18145)

Wishnu Abhinaya Sulaba, Binti Solihah, Syandra Sari
110-116

PDF (https://e-journal.trisakti.ac.id/index.php/intelmatics/article/view/18145/12697)

Abstract: 53 | PDF downloads:32

4. Peer Review Process (https://e-journal.trisakti.ac.id/index.php/intelmatics/PeerReviewProcess)

4. Editorial Team (https://e-journal.trisakti.ac.id/index.php/intelmatics/EditorialTeam)

5. Reviewer Team (https://e-journal.trisakti.ac.id/index.php/intelmatics/ReviewerTeam)

6. Plagiarism Check (https://e-journal.trisakti.ac.id/index.php/intelmatics/PlagiarismPolicy)

7. Unique Visits (https://statcounter.com/p12931691/summary/?account_id=5116944&login_id=5&code=08a88f090f134bf5faa7bf1c42f8238c&guest_login=1)

8. Journal History (https://e-journal.trisakti.ac.id/index.php/intelmatics/JournalHistory)

10. Journal Business Model (https://e-journal.trisakti.ac.id/index.php/intelmatics/JournalBusinessModel)

11. Author Index (https://e-journal.trisakti.ac.id/index.php/intelmatics/search/authors)

12. Publication Frequency (https://e-journal.trisakti.ac.id/index.php/intelmatics/PublicationFrequency)

13. Reviewer/EditorJoin (https://e-journal.trisakti.ac.id/index.php/intelmatics/ReviewerEditorJoin)

14. Archiving (https://e-journal.trisakti.ac.id/index.php/intelmatics/Archiving)

15. . Mailing Address (https://e-journal.trisakti.ac.id/index.php/intelmatics/MailingAddress)

16. Retraction Policy (https://e-journal.trisakti.ac.id/index.php/intelmatics/RetractionPolicy)

17. Copyright Notice (https://e-journal.trisakti.ac.id/index.php/intelmatics/CopyrightNotice)

18. Open Access Policy (https://e-journal.trisakti.ac.id/index.php/intelmatics/OpenAccessPolicy)

19. Indexing and Abstracting (https://e-journal.trisakti.ac.id/index.php/intelmatics/IndexingandAbstracting)

(https://sinta.kemdikbud.go.id/journals/profile/14451)

(https://scholar.google.com/citations?
user=Swy3MjgAAAAJ&hl=id&authuser=6)

(https://garuda.kemdikbud.go.id/journal/view/32717)

(https://search.crossref.org/search/works?q=2775-
8850&from_ui=yes)

(https://portal.issn.org/resource/ISSN/2775-8850)

(https://app.dimensions.ai/discover/publication?
search_mode=content&and_facet_source_title=jour.1408685)

(https://www.citefactor.org/search/keywords/journals/Intelmatics)

(https://scite.ai/journals/intelmatics-R5Y34)

TOOLS

REFERENCE MANAGEMENT TOOLS

APPLICATION PLAGIARISM CHECKER

LANGUAGE

English (https://e-journal.trisakti.ac.id/index.php/intelmatics/user/setLocale/en?source=%2Findex.php%2Fintelmatics%2Fissue%2Fview%2F1266)

INFORMATION

For Readers (https://e-journal.trisakti.ac.id/index.php/intelmatics/information/readers)

For Authors (https://e-journal.trisakti.ac.id/index.php/intelmatics/information/authors)

For Librarians (https://e-journal.trisakti.ac.id/index.php/intelmatics/information/librarians)

UNIQUE_VISITS

00015840 (https://statcounter.com/)

(https://e-journal.trisakti.ac.id/index.php/intelmatics/about/aboutThisPublishingSystem)

Analysis of Topic Movement & Conversation Membership on Twitter using K-Means Clustering

Josua Valentino Hutagalung[1], Agung Sediyono[2*], Binti Solihah[3]

[123]Informatics Study Program, Faculty of Industrial Technology, University of Trisakti

Abstract: **Humans are born to socialize with each other. Social media is one of the media to be able to socialize with each other. Twitter is one of the social media that contains hundreds of millions of tweets where the tweet contains news, products that are currently popular, even about the daily life of users who can change. Social Context Analysis is a tool to analyze social changes and individual needs in society from time to time. In this study, the author uses the K-means Clustering method to group topics on Twitter. Based on the result, it shows that topic movement can be detected in moving from one cluster to another cluster. It also is followed by membership movement. Therefore, it can be concluded that this study can be used to show topics and its membership movement.**

Keywords: ***K-mean, Social Context Analysis, Clustering***

Article History:
Submission date: - revision date: - accepted date:

I. INTRODUCTION

Humans are living beings who always want to socialize because humans are social creatures. In general, socialization is the process of a person to live, recognize social norms and values that are useful for shaping one's attitude in accordance with the behavior or demands of society. Socializing can be done by means of communication between a person / individual that can produce information. Twitter is a social media in the form of microblogging (small blog) where users or users can send tweets that contain information, news, or can even contain expressions of heart & thoughts. Tweet is a content that is an element of Social Context To do Social Context Analysis, you can use the Clustering method and several studies have used the Clustering method [1].

Clustering method can be used to perform Social Context Analysis. This is because clustering aims to group the partition of an object into one group that has the same characteristics and different characteristics will be grouped in other groups [2]. So the clustering method can apply Social Context theory where Social Context is used as an analytical tool to identify, understand, and make predictions related to social problems and community needs [3].

In this study, we will analyze the topic movement (context of conversation) and also the movement of membership (users) on Twitter with the stage of clustering the topic of conversation within one week

II. LITERATURE REVIEW

The research conducted by Muhammad Ali with his colleague Farindika Metandi with the title "Determination of Types of Comments on PSSI (Indonesian Football Association) Tweets" is an example of using the K-means Clustering method. In this study, the researchers used 50 commentary data taken from Twitter users who discussed PSSI (Indonesian Football Association) which was used as research material. The researcher also determined 3 types of comments, namely irony, cynicism and sarcasm. Initially, the researcher analyzed the data based on the needs of the researcher and counted the number of positive and negative words in each comment manually. After that, do the clustering process with K-means clustering. In the first iteration, determine the number of clusters and in this study use 3 clusters. After determining the number of clusters, determine the centroid or cluster center provided that the first centroid is the average result for each criterion, the second centroid is the maximum value of each criterion and the third centroid is the minimum value of each criterion. Especially in the first iteration the value of the starting point of the cluster center (centroid) is chosen randomly. After determining the centroid, calculate the distance of the object to the nearest cluster center using Euclidean Distance. After that group objects based on the minimum distance to determine the members of each cluster. After all data has been placed in the nearest cluster, the study calculates the new cluster center (centroid) based on the average members in the cluster and produces the first iteration cluster center (centroid). The iteration process is carried out in the same way until the new cluster center does not change in the previous cluster center (centroid). The study obtained clustering results up to the 5th iteration. The results showed that there were 3 clusters, the first cluster was 20% cynical comments, the second cluster 6% sarcasm comments and the third cluster 54% irony comments. The study also conducted comparisons and tests using the Rapid Miner application. The results of the study using the Rapid Miner application produced 3 clusters, namely the first cluster of 70%, the second cluster of 10% and the third cluster of 20%. The calculation using K-means Clustering is different from the calculation using the Rapid Miner application because the study uses the average value, maximum value and minimum value to determine the initial centroid [4].

Clustering by the Fuzzy C-Means method which is part of the Fuzzy Clustering [5]. The study used secondary data with 24 districts / cities in Central Sulawesi. The research grouped the regions with the variables of the number of cases of malnutrition, the number of health centers, the number of health workers, families who behaved in a clean and healthy life, the number of poor people and the average household member. In contrast to using the k-means method, the FCM (Fuzzy C-Means) method first determines the value of the U partition matrix with a size of n x p where n is the number of data samples, which is 24 and p is the attribute or parameter, which is 7. After that, determine the initial parameter value. such as number of clusters, weighting power, maximum iteration, smallest expected error, initial objective function and initial iteration. After that generate random numbers as matrix elements in the initial partition U24x2 where 24 is the number of data and 2 is the number of desired clusters. After that determine the center of the V2x7 cluster which is the center of the first iteration cluster and calculate the objectivity function. Next, calculate changes to the U matrix. After that, check the stop condition with $|P1 - P0| >$ the smallest error value (in this study 10-5) then proceed to the next iteration until $|Pt - Pt\text{-}1| <$. The research carried out up to 55 iterations to produce 2 clusters with the help of the MATLAB R2013 application. In the first cluster, 18 districts/cities have high levels of malnutrition. In the second cluster, 6 districts/cities have low levels of malnutrition [6].

Research of Clustering and geovisualization of twitter text data with the k-means algorithm for cases of forest fires and natural disasters. The study used the K-means method with the aim of knowing the distribution of tweet locations regarding cases of forest fires and natural disasters. The research uses R shiny and connects to Twitter with an API. The study used 12 hashtags to retrieve tweet data and produced 33689 tweets and took 2160 that had geolocation. After getting 2160 tweet data, the study did a manual check for duplicates with Excel

*Corresponding author
E-mail addresss : trisakti_agung06@trisakti.ac.id

so as to produce 1681 tweets. The tweet data is exported into a CSV form which has the header username, date, time, text, tweetID, placeID, city center, location, longitude, and latitude. After getting the tweet data, the researcher carried out the tokenizing stage where this process changed capital letters to lowercase letters, removed mentions, URLs, punctuation marks and numbers with the gsub function in the tm package. After the researcher tokenizing, the researcher carried out the data normalization stage where at this stage changed the non-standard words into standard words according to the dictionary. In addition, it normalizes abbreviations such as " hrs " " tdk " to " harus " " tidak ". After that, the researcher carried out the stopwords removal stage where this stage removed words that often appeared such as conjunctions, prepositions, exclamations, and adverbs. So words like " dan " ," tidak " are lost at this stage. After that, the researcher carried out the stemming stage where at this stage changed the word into a basic word. Like removing the prefix me – or the suffix "kan". In this process the researcher uses the program code from Wahyunintyas (2016) for Indonesian. Furthermore, the researchers carried out the process of making the Term Document Matrix (TDM) which aims to obtain a matrix of the number of occurrences of terms in the document. In this process, the researcher obtained 1681 documents with 4037 terms. It is necessary to reduce the number of terms because it makes the dimensions of the matrix large. Term reduction is done with the help of the removeSparseTerm function in the R application. After the researcher determines the sparse value of 99%, the researcher gets a new term value of 80 terms. After that the results of the TDM process are converted into a matrix and transposed on R to perform clustering with K-Means. Each document is searched for similarities by using the Euclidean distance function. After that, the researcher selected the value of k to get the smallest SSE (Sum Squared Error) value. The study conducted tests with a value of k = 2 to k = 7 and the lowest SSE (Sum Squared Error) result was at a value of k = 7 of 5693.169. The results of the study produced 7 clusters and with 7 categories, namely forest fires, smoke, tornadoes, floods, landslides, earthquakes and eruptions which were then mapped according to the location of the tweet using the Shiny framework [7].

Previous researches shows that K-mean method can be used to cluster the topics in social media conversition. By weekly analyasis on twitter conversation using K-mean method, it can be presented the weekly movement of topics conversition. Futhermore, by analysing the membership of each topic, it can be detected the movement of membership as well.

III. RESEARCH METHODOLOGY

The research stages consist of Twitter Data Acquisition, Text Preprocessing, TFIDF Weighting, Dimensional Reduction with NMF (Non-Negative Matrix Factorization), determining the best K value with Silhoutte Coefficient, Clustering using K-means, and Membership Transfer Identification. Important to note that K-means is one most powerful clusteting method [8]. The following is a flowchart of the research stages:

Fig 1. Research Phase

3.1. Data Acquisition

Data Acquisition on Twitter aims to get tweet data from Twitter using the Twitter API. Twitter Data Acquisition in this research is using Twitter data crawling method which is connected via Twitter API. To get the Twitter API, you can access it via https://developer.twitter.com/ then click on "Create App" after that fill in all the required forms. After successfully creating the App, click on the name of the previously registered App, then click the "Keys and Tokens" section so that the API Key, API Secret Key, Access Token and Access Secret Token can be obtained. The previously obtained API Key, API Secret Key, Access Token and Access Secret Token are used for crawling Twitter data using scripts in Python. Researchers use the Twitter API Search feature by inputting the required keywords. The twitter server will provide tweet data according to the keywords you are looking for and can be saved in csv format. Researchers conduct data acquisition for 7 days to be compared at the next stage.

3.2. Text Preprocessing

At this stage, the tweet data that has been obtained previously is prepared before entering the next stage. The stage of text preprocessing data is divided into 6 stages, namely:

3.2.1. Cleaning

In the cleaning stage, it is the stage to clean each tweet of unnecessary words/symbols.

Table 1. Example Cleaning

Before	After
@Smart_Bisnis pagi min.indihome gangguan udah lebih dr 24 jam ini min. mana solusinya??? jgn cuma nunggu2 aja	pagi min...indihome gangguan udah lebih dr jam ini min. mana solusinya??? jgn cuma nunggu aja

@Telkomsel hi admin, saya pengguna telkomsel dan indihome, apakah akhir ini ada gangguan ya di daerah jakarta utara https://t.co/8KmKMdiqfA	hi admin, saya pengguna telkomsel dan indihome, apakah akhir ini ada gangguan ya di daerah jakarta utara
@liuniverseyy HP aktifnya melalui DM akun baru kami di @IndiHomeCare ya Kak. Terima Kasih - Auli(2/2)	HP aktifnya melalui DM akun baru kami di ya Kak. Terima Kasih -Auli(/)'

3.2.2. Casefolding

The Casefolding stage is the stage of changing the letters in the data into lowercase letters and also deleting other than letters a to z.

Table 2. Example Casefolding

Before	After
pagi min...indihome gangguan udah lebih dr jam ini min. mana solusinya??? jgn cuma nunggu aja	pagi min indihome gangguan udah lebih dr jam ini min mana solusinya jgn cuma nunggu aja
hi admin, saya pengguna telkomsel dan indihome, apakah akhir ini ada gangguan ya di daerah jakarta utara	hi admin saya pengguna telkomsel dan indihome apakah akhir ini ada gangguan ya di daerah jakarta utara
HP aktifnya melalui DM akun baru kami di ya Kak. Terima Kasih - Auli(/)'	hp aktifnya melalui dm akun baru kami di ya kak terima kasih auli

3.2.3. Tokenizing

In the Tokenizing stage, tweets in the form of sentences will be broken down into words.

Table 3. Example Tokenizing

Before	After
pagi min indihome gangguan udah lebih dr jam ini min mana solusinya jgn cuma nunggu aja	['pagi', 'min', 'indihome', 'gangguan', 'udah', 'lebih', 'dr', 'jam', 'ini', 'min', 'mana', 'solusinya', 'jgn', 'cuma', 'nunggu', 'aja']
hi admin saya pengguna telkomsel dan indihome apakah akhir ini ada gangguan ya di daerah jakarta utara	['hi', 'admin', 'saya', 'pengguna', 'telkomsel', 'dan','indihome', 'apakah','akhir','ini','ada', 'gangguan','ya','di', 'daerah', 'jakarta', 'utara']
hp aktifnya melalui dm akun baru kami di ya kak terima kasih auli	['hp', 'aktifnya', 'melalui', 'dm', 'akun', 'baru', 'kami','di','ya','kak','terima', 'kasih', 'auli']

3.2.4. Word Normalization

In word normalization, tweet data will be changed from non-standard words to standard ones. In addition, at this stage abbreviations such as "yg" will be changed to "yang".

Table 4. Example Word Normalization

Before	After
['pagi', 'min', 'indihome','gangguan', 'udah', 'lebih','dr', 'jam','ini','min','mana','solusinya', 'jgn', 'cuma','nunggu', 'aja']	['pagi','admin', 'indihome', 'gangguan','sudah', 'lebih', 'dari', 'jam', 'ini', 'admin', 'dimana', 'solusinya','jangan','hanya','menunggu', 'saja']
['hi', 'admin', 'saya', 'pengguna','telkomsel', 'dan','indihome','apakah', 'akhir','ini', 'ada', 'gangguan','ya', 'di','daerah', 'jakarta', 'utara']	['hi', 'admin', 'saya', 'pengguna', 'telkomsel', 'dan','indihome', 'apakah','akhir', 'ini','ada', 'gangguan', 'ya','di','daerah', 'jakarta', 'utara']
['hp', 'aktifnya', 'melalui','dm', 'akun', 'baru','kami','di','ya','kak', 'terima', 'kasih', 'auli']	['hp', 'aktifnya', 'melalui', 'dm', 'akun', 'baru', 'kami','di','ya','kak','terima', 'kasih', 'auli']

3.2.5. Stopwords Removal

In the Stopwords stage, tweet data that has conjunctions will be deleted because it is not important in research.

Table 5. Example Stopwords Removal

Before	After
'indihome', 'gangguan', 'sudah', 'lebih', 'dari', 'jam', 'ini', 'admin', 'dimana', 'solusinya', 'jangan', 'hanya', 'menunggu', 'saja']	'indihome', 'gangguan', 'jam', 'admin', 'dimana', 'solusinya', 'menunggu']
['hi', 'admin', 'saya', 'pengguna', 'telkomsel', 'dan', 'indihome', 'apakah', 'akhir', 'ini', 'ada', 'gangguan', 'ya', 'di', 'daerah', 'jakarta', 'utara']	['hi', 'admin', 'pengguna', 'telkomsel', 'indihome', 'gangguan', 'daerah', 'jakarta', 'utara']
['hp', 'aktifnya', 'melalui', 'dm', 'akun', 'baru', 'kami', 'di', 'ya', 'kak', 'terima', 'kasih', 'auli']	['hp', 'aktifnya', 'dm', 'akun', 'kak', 'terima', 'kasih', 'auli']

3.2.6. Stemming

In the Stopwords stage, tweet data that has conjunctions will be deleted because it is not important in research.

Table 6. Example Stemming

Before	After
['pagi', 'admin', 'indihome', 'gangguan', 'jam', 'admin', 'dimana', 'solusinya', 'menunggu']	pagi admin indihome ganggu jam admin mana solusi tunggu

*Corresponding author
E-mail address : trisakti_agung06@trisakti.ac.id

['hi', 'admin', 'saya', 'pengguna', 'telkomsel', 'dan', 'indihome', 'apakah', 'akhir', 'ini', 'ada', 'gangguan', 'ya', 'di', 'daerah', 'jakarta', 'utara']	hi admin guna telkomsel indihome ganggu daerah jakarta utara
['hp', 'aktifnya', 'dm', 'akun', 'kak', 'terima', 'kasih', 'auli']	hp aktif dm akun kak terima kasih auli

3.3. TF-IDF (Term Frequency Inverse Document Frequency)

TF-IDF is a method for interpreting a word (term) into numbers in a document. TF (Term Frequency) is the total appearance of a word (term) in a document. The term that occurs most often will have a large value in a document. Next is to find the value of the IDF (Inverse Document Frequency). The IDF (Inverse Document Frequency) value is obtained from the large number of documents that have a word (term). If a word (term) is owned by a few documents, the greater the influence of the word (term) on a document. However, if a word (term) is owned by many documents, the smaller the influence of a word (term) on the document. Here is the formula for finding TF-IDF:

$$W_d = tf_{dt} * IDF_t$$

Description:
D : Document to – d t : Word (term) to – t
W : The weight of the document – d over the word – t
Tf : The number of words (terms) to be searched in a document
IDF : Inverse Document Frequency

3.4. NMF (Non-Negative Matrix Factorization)

NMF is a word representation matrix with topics that have non-negative values so that this matrix is easier to interpret.

NMF is a method for factoring a V matrix of size m x n into non-negative matrices. A general NMF decomposition method which can be expressed by the equation:

$$V = WH$$

3.5. Silhouette Coefficient

Silhouette Coefficient is used to see the quality and strength of the group, how well or poorly an object is placed in the group. This method is a combination of the method of separation and cohesion [9]. To calculate the silhouette coefficient value, it is necessary to calculate the silhouette index value from the i-th data. The silhouette coefficient value is obtained by finding the maximum value of the Global Silhouette Index from the number of clusters 2 to the number of clusters n-1.

Subjective criteria for grouping measurements based on the Silhouette Coefficient (SC) [10], can be seen in the following table:

Table 7. Silhouette Value Criteria

SC Value	Criteria
≤ 0,25	Bad
0,26 – 0,50	Weak
0,51 – 0,70	Good
0,71 – 1,00	Strong

3.6. Social Context Model Development with K-means Clustering

At this stage, tweet data that has passed the Text Preprocessing stage will enter the data analysis stage using clustering techniques. In this study, data collection was carried out periodically, so data analysis was carried out periodically. The following is a flowchart in the development of a social context model with K-means clustering:

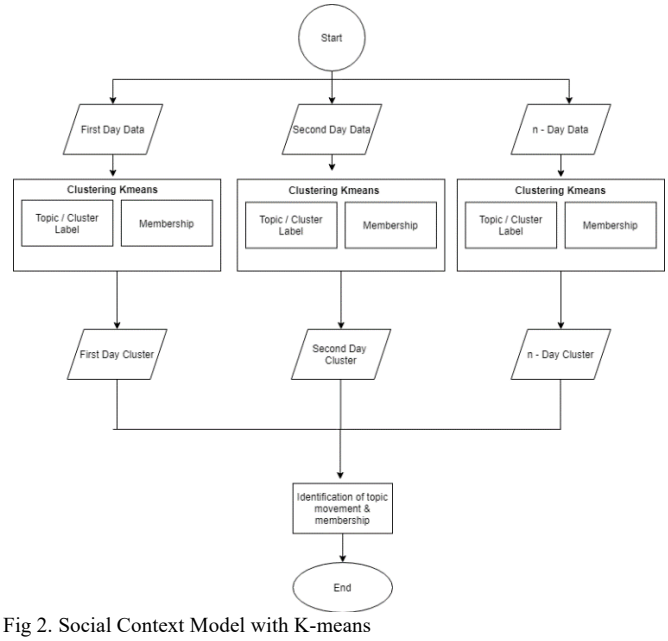


Fig 2. Social Context Model with K-means

IV. RESULT AND DISCUSSION

4.1. Data Acquisition

The data retrieval (acquisition) process uses the Twitter data crawling method with the python programming language supported by the tweepy library. The Data Crawling process utilizes the API features of Twitter and the resulting data is a file with the csv extension. The twitter API feature takes the key from the twitter developer portal that has been provided. The Twitter Data Crawling process is carried out during August 09, 2021 to August 15, 2021. The keywords used adjust to 5 Trending Topics every day from August 09, 2021 to August 15, 2021.

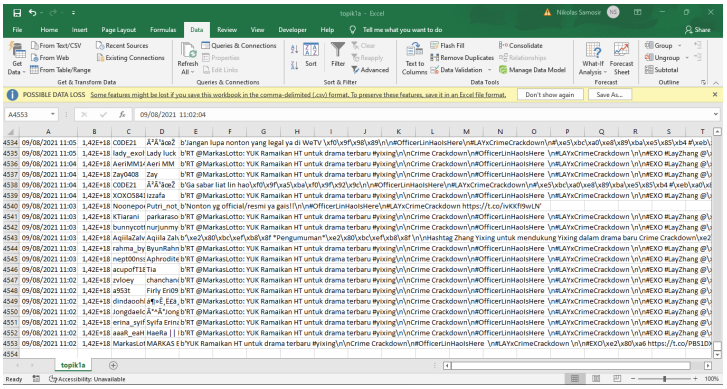


Fig 3. Output Crawling Data

4.2. Text Preprocessing

At this stage, the tweet data that has been successfully obtained through the Crawling Data process is an unstructured tweet data with CSV format. Text Preprocessing is the process of converting unstructured data into structured data before entering the next stage (Kmeans Clustering). This stage includes 6 processes, namely: Cleaning, Casefolding, Tokenizing, Normalization of Words, Elimination of Stopwords, and Stemming.

4.3. TFIDF weighting

At this stage, Tweet data that has passed the Text Preprocessing process is in the form of words or text. The data must be converted into numbers or numeric using the TF-IDF method. The following is the source code and an example of the TFIDF weighting results:

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer

vectorizer = CountVectorizer()
tf_transformer = TfidfTransformer()

datavector = vectorizer.fit_transform(text_baru['Tweet_Bersih'])
datatfidf = tf_transformer.fit_transform(datavector)
datavector.shape
```

```
(20022, 13725)
```

```
weight_val = np.asarray(datatfidf.mean(axis=0)).ravel().tolist()
```

```
weight_val_df = pd.DataFrame({'term': vectorizer.get_feature_names(),'weight':weight_val })
```

```
weight_val_df.sort_values(by='weight',ascending=False).head(5)
```

	term	weight
6744	lord	0.067029
194	adi	0.065993
711	arnoldposmamo	0.032422
1762	bts	0.029899
7254	merang	0.026104

Fig 4. Source Code and Example of TFIDF Weighting Results

4.4. TFIDF Matrix Dimension Reduction with NMF (Non-Negative Matrix Factorization)

At this stage, the TFIDF Matrix generated from tweet data has a large dimension. The TFIDF matrix is the total document size (total record tweets) X total terms (distinct). Therefore, the storage and processing of these matrices requires a long time and large resources. The NMF (Non-Negative Matrix Factorization) method is one method to reduce the dimensions of the TFIDF matrix.

4.5. Determination of K value

In K-means Clustering there is input from the user, namely the value of K. The value of K in Kmeans Clustering determines the number of clusters that will be generated by Kmeans Clustering. Silhouette Score is one method to determine whether the input K value is optimal or not.

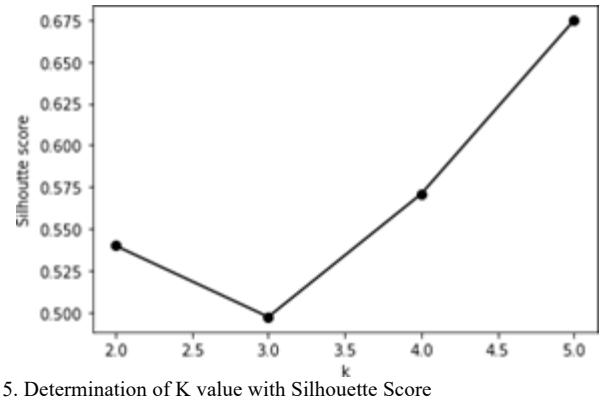


Fig 5. Determination of K value with Silhouette Score

The chart above shows that the best K value on the first day dataset (09 August 2021) is 5.

4.6. Kmeans Clustering

After the dimension reduction process using the NMF (Non-Negative Matrix Factorization) method and having found the optimal K value with the Silhoutte Score, then enter the clustering stage with Kmeans Clustering. At this stage using the Kmeans library which is already available in Python.

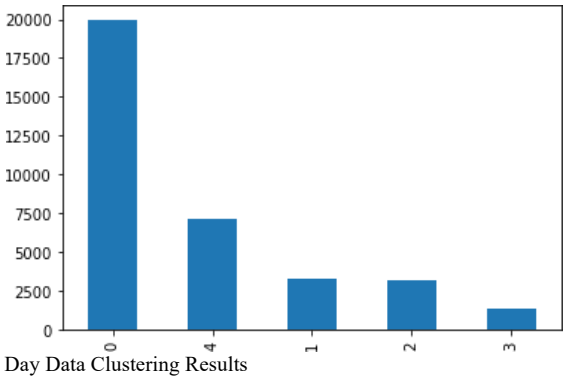


Fig 6 First Day Data Clustering Results

The following is a chart of Kmeans Clustering results on the first day of data:

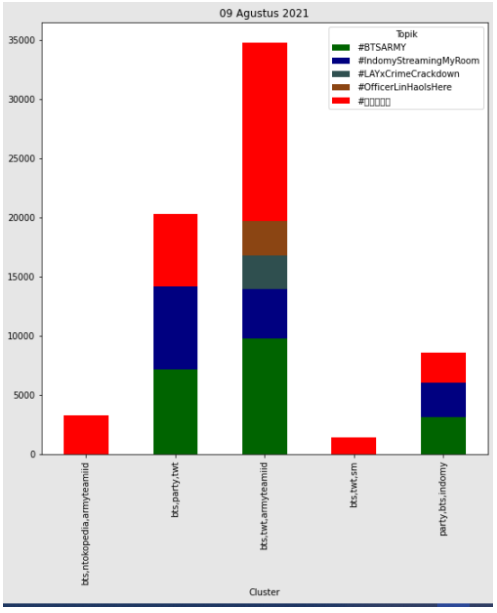
The following is a list of cluster labels taken from the top 3 the best term in each cluster:

Table 8. Cluster Labeling

Cluster	Label
0	bts,twt,armyteamiid
1	bts,ntokopedia,armyteamiid
2	party,bts,indomy
3	bts,twt,sm
4	bts,party,twt

4.7. Topic Distribution Chart in Cluster

At this stage, each cluster data that has been formed will be visualized the source or origin of the dominant topic in each cluster. The following is a chart of the distribution of topic sources on the first day of data clusters:



*Corresponding author
E-mail addresss : trisakti_agung06@trisakti.ac.id

Fig 7. Topic Distribution Chart in Cluster

Based on the chart above, clusters with the labels "bts,ntokopedia,armyteamiid" , "bts,twt,armyteamiid", "bts,twt,armyteamiid" are dominantly derived from the topic or keyword "#방탄소년단". The cluster with the label "party,bts,indomy" comes from the topic or keyword "#BTSARMY". And in the Cluster "bts,party,twt" also dominantly comes from the topic or keyword "#BTSARMY".

4.8. Membership Movement Chart

Each cluster that has been formed is stored in a CSV containing the username column and also the label column that has been determined in the previous stage. At this stage, it will be visualized and explained about the movement of membership / members / users from day to day. In this experiment, using cluster data on the first day (August 09, 2021) and cluster data on the second day (August 10, 2021) which have also gone through the same process. The following is a chart of membership transfer on the first day of the cluster to the second day of the cluster:

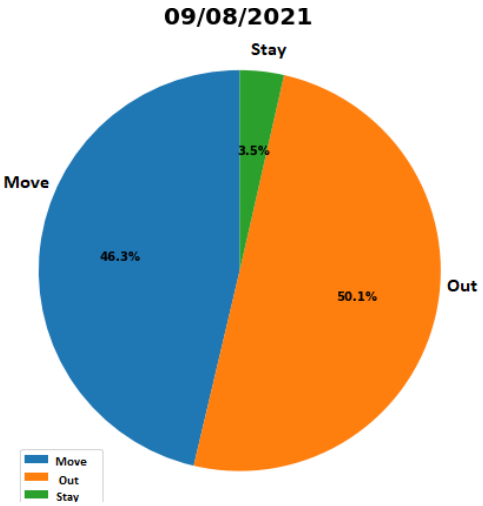


Fig 8. Pie Chart Membership Movement

The chart above shows that there are 50.1% membership / users who leave. The word "out" in the chart above explains that the membership / user contained in the first day cluster data (9 August 2021) is not found in the second day cluster data (10 August 2021). Then followed by 46.3% membership / user who moved. The word "move" in the chart above explains that membership / users who are in the data cluster on the first day discuss a topic / label but on the second day cluster data do not discuss the same topic / label but are on another topic / label. Finally, there are 3.5% membership / permanent users. The word "stay" in the chart above explains that there is a membership / user who in the first day cluster data also discusses the same topic / label on the second day cluster data.

4.9. Discussion

The analysis phase in this study uses the Anaconda Navigator software using Jupyter Notebook using the Python programming language. The stages of this research process begin with Crawling Data with a total of 34 topics or keywords originating from 5 top trending topics every day for 7 days (August 09, 2021 – August 15, 2021). The total tweet record data obtained for 7 days is 286,559 tweets with an average of 40,937 tweets per day.

After getting the tweet data using the Twitter Data Crawling method, the next step is the Text Preprocessing stage. This stage is the stage of converting data into structured data. In this process, there are several processes, including: Cleaning, Casefolding, Tokenizing, Normalization of Words, Elimination of Stopwords, and Stemming.

After producing clean and tidy data, the next step is to enter the TFIDF weighting stage. After the TFIDF weighting stage is successful, then proceed with the matrix dimension reduction stage using the NMF (Non-Negative Matrix Factorization) method. Next is to determine the best K value with the best Silhoutte Score. Here are the top 5 terms that often appear in the cluster for 7 days:

Table 9. Top 5 Terms in Clusters for 7 Days

Term	Occurrence Frequency
bts	37
twt	16
armyteamiid	7
astro	7
lord	7

V. CONCLUSION

Based on the results of the analysis and discussion, 34 topics or keywords were obtained from the top 5 trending topics in Indonesia on Twitter for 7 days (9 August – 15 August 2021) with a total record of 286,559 tweets resulting in 116 clusters. Topic movement (clusters) for 7 days is dominated by topics that are always new or topics that do not last more than 1 day. Topic movements that last more than 1 day are in topics or clusters with the label "uwu, short, namjoon" (12 – 13 August 2021) and in the cluster with the label “yuk,wetv,tonton” (13 – 14 August 2021). Membership movement for 7 days, namely: first day (9 August 2021) 50.1% Exit (did not move to any topic on the second day), 46.3% Moved (moved to another topic on the second day) and 3.5% Stayed (moved). to the same topic on the second day). Day two 68% out and 32% moved out; Third day 67.8% out and 32.3% moved; Fourth day 62.5% out and 37.5% moved; Day five 80.7% out and 19.3% moved; Day six 70.9% moved in and 29.1% left. By this finding it study can be adopted to an computer application that monitor topics movement and its memberships in order to know discourse in society. In practical, it can be used in targeted marketing, social policy, etc.[11][12].

Based on the completed discussion, the authors have suggestions, namely: In this study, the authors only use data within a week (August 09, 2021 – August 15, 2021), for further research, the time period can be added. In addition, in this study, the authors conducted an analysis of topic movements and daily membership movements, for further research they could analyze weekly or monthly periods.

REFERENECES

[1] Sánchez-Rada,J.Fernando , and Carlos A. Iglesias , “ Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison “, Information Fusion , vol.52 , pp. 344- 356, 2019.

[2] Gustientiedinaa , M.Hasmil Adiyaa, and Yenny Desnelitab, “ Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru “ , JURNAL NASIONAL TEKNOLOGI DAN SISTEM INFORMASI , VOL. 05 , NO. 01 , 2019.

[3] Earle ,Leon and Tony Earle , “ Social Context Theory “, South Pacific Journal of Psychology , 2015.

[4] Rheza Muhammad Ali , Farindika Metandi , “ IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK PENENTUAN JENIS KOMENTAR PADA TWEET PSSI “ , JUST TI: (Jurnal Sains Terapan Teknologi Informasi) , Vol. 12 , No. 2 , 2020.

[5] Alnajran ,Noufa , Keeley Crockett, David McLean and Annabel Latham , “ Cluster Analysis of Twitter Data: A Review of Algorithms ”, Proceedings of the 9th International Conference on Agents and Artificial Intelligence , Vol.1 ,2017.

[6] Sanusi, Wahidah , Ahmad Zaky ,and Besse Nur Afni , “ Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di

Provinsi Sulawesi Selatan Berdasarkan Faktorfaktor Penyebab Gizi Buruk “ , Journal of Mathematics, Computations, and Statistics , Vol.2 , No.1 , pp. 47 - 54 , 2019.

[7] Darojat , Hamid . “ CLUSTERING DAN GEOVISUALISASI DATA TEKS TWITTER DENGAN ALGORITME K-MEANS UNTUK KASUS KEBAKARAN HUTAN DAN BENCANA ALAM “ . Skripsi . FMIPA. Ilmu Komputer , INSTITUT PERTANIAN BOGOR, Bogor ,2017.

[8] Ahmed, Mohiuddin, Raihan Seraj, and Syed Mohammed Shamsul Islam. 2020. "The *k-means* Algorithm: A Comprehensive Survey and Performance Evaluation" *Electronics* 9, no. 8:1295. https://doi.org/10.3390/electronics9081295

[9] T. M. Kodinariya and P. R. Makwana, “Review on determining number of cluster in K-Means s in data: An introduction to cluster analysis, USA: Wiley Series in Probability and Statistics

[10] B. Ryu, A. R. Rivera, J. Kim, and O. Chae, “Local Directional Ternary Pattern for Facial Expression Recognition,” IEEE Trans. Image Process., vol. 26, no. 12, pp. 6006–6018, 2017, doi: 10.1109/TIP.2017.2726010.

[11] Mundt, M., Ross, K., & Burnett, C. M. (2018). Scaling Social Movements Through Social Media: The Case of Black Lives Matter. Social Media + Society, 4(4). https://doi.org/10.1177/2056305118807911

[12] Sutan, A.J., Nurmandi, A., Mutiarin, D., Salahudin, S. (2021). Using Social Media as Tools of Social Movement and Social Protest in Omnibus Law of Job Creation Bill Policy-Making Process in Indonesia. In: Antipova, T. (eds) Advances in Digital Science. ICADS 2021. Advances in Intelligent Systems and Computing, vol 1352. Springer, Cham. https://doi.org/10.1007/978-3-030-71782-7_24

*Corresponding author
E-mail addresss : trisakti_agung06@trisakti.ac.id