ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse182952020.pdf
https://doi.org/10.30534/ijatcse/2020/182952020

Personal popular Name Identification Through Twitter Data

Najirah Umar[1], Yuyun[2], Hazriani[3], Herman[4]
[1,]STMIK Handayani Makassar, Indonesia, najirah@handayani.ac.id
[2,]STMIK Handayani Makassar, Indonesia, yuyunwabula@handayani.ac.id
[3]STMIK Handayani Makassar, Indonesia, hazriani@handayani.ac.id
[4] Balai Besar Pengkajian dan Pengembangan Komunikasi dan Informatika Makassar, Indonesia, herman@kominfo.go.id

ABSTRACT

A name has important meaning to someone's life. Since, a name describes specific identities to distinguish someone from others, and represent the uniqueness of the person as well. This paper addresses popular personal names which are preferred to be used by people in Makassar City, by investigating the typology of first name, middle name, and last name. Around 14,500 personal names were collected from social networking Twitter. To ensure data integrity, preprocessing data was conducted by implementing Min-Max normalization which functions to replace missing or false values. Based on data analysis with the K-means algorithm, three clusters were formed, namely *very popular*, *popular* and *quite popular*. Finally, through a ranking method, we discover a list of a hundred names popularly used by people in Makassar city.

Key words: twitter-name, popular-name, k-means-clustering, ranking.

1. INTRODUCTION

The research of personal names is referred to as anthroponomy [1]. A name represents a person's social, cultural, and legal identity [2][3], [4]. When a child is given a name, parents select a name that actually not only identifies the child as a person but describes the socio-cultural [5]. In this case, giving a new kid name, parents consciously hope that their child will become a good person according to the inherent name, and declare a hope or prayer[6]. What hides behind the label of personal identity? Names can represent self-descriptions and self-knowledge[7]. Moreover, a name often considered as prized property to identify house belonging, familial belonging, descent group, village, local group, and ethnic group belonging [8]. In terms of sex, a name explained gender characteristics [9], identify the character, self and identity [10], and development of the personality [11], [12]. In the family tree, the name indicated birth order [13] and the parent-child relationship [14]. While in the education field, name functions as the initial step in recognizing who a person is. calling student by name depict respect, they sense recognized as an individual, and developing a fell of community in the class room[15].

Current research on a personal name, methods in data collection are grouped into two categories. *First*, focused on direct research such as analyzed the list name on the inscriptions; read up the historical documents; encyclopedias book; the yearbooks; and inscriptions on funerary monuments. *Second*, by analyzing and interpreting anthroponomical data, such as the geographical method; the comparative-historical method; the linguistic analysis; the statistical method; the computing method; and the cartographic method [16]. Furthermore, we addressed some researcher's use direct observation to investigated personal names, for instance, a socio-cultural and linguistic analysis of Arabic personal names [9], identify parental naming activity in multilingual Manchester [17], the development of personal names in Kudus [18], describe the development of the form of personal names in the Semarang community [19], Javanese names during the height of the Hindu-Buddhist Kingdoms in Java [20]. Another researcher used corpus data sourced from the student database names to describe the trend of the naming system on Javanese people which is beginning to switch from the Javanese to Arabic[21].

In recent years, the user of social media has been increased drastically. Millions of data are stored in these sites every day. Through social networking, a user leaves digital traces data. The digital trace data contains a period (time-stamped), user's status text, location category, as well as individual information of the social media account owner [22]. In this case, individual footprint e.g. the user's name is attached in the social media user's profile. The Importance of name on the user profile to facilitate people to find us.

A lot of benefits can be explored by social networking data, such as quantifying the mobility of urban inhabitant[23], understanding mobility patterns [24], inferring individual lifestyle patterns [25], sensing urban land use for urban planning application [26], identifying the city center [27], estimating user location [28], urban population [29], students' group collaborative [30], privacy awareness [31], consumer behavior[32] and predict mental illness [33]. In general, prior studies have focused on the field text (user posted text) and geo-location (location check-in) as the criteria to make the measurement.

In this research, our focus is to analyze a name attached in the social media user's profiles as exemplified in (Figure. 1). The research of the user's profile has studied by some scientistssuch as a study of the user profile to obtain user's interest using K-means and Spectral clustering to cluster user interest [34], analyze username to classify gender and language by applied the Morfessor algorithm [35]. Then [36], revealed fraud profiles using machine learning for further detection of spamming profiles. While [37], presented a string similarity algorithm to measure the redundant information between different display names of the same individual. But In previous research has not discussed the popular individual names on social media.

The aim of this research is to find out the human popular names and identify the characters of the first letters of personal name using Twitter social media data. This paper proposes K-Means clustering to group the user profile name. Meanwhile, we offer the ranking method to identify popular personal name. K-means method used to cluster the names into three groups, namely: very popular, popular and quite popular. While the ranking method to sort a hundred popular names by categories first name, middle name, and last name. We regard that research aim and method proposed in this paper distinguish our work from others'**.**

Figure 1:A user name profile on Twitter

2. RESEARCH METHOD

2.1Normalization, Ranking and K-Means Methods

After collecting names through Twitter user's profiles, the next step is performing data cleaning and normalization. Data cleaning is performed to verify and correct inconsistent data, such as equalize letters in a sentence into lower case, while normalization aimed to scale the data into a certain range to ease data processing. Since, data recapitulation on the user's name shows a huge data gap, which ranges between 1 to 1.500, data normalization is required. For that purpose, we use the min-max method. A min-max method is a normalization method by performing linear transformation into data source (real data), to obtain new (normalized) value in a range between 0 to 1 using. The min-max method was implemented using Formula 1.

$$X_{new} = \frac{X_i - Min(X)}{Max(X) - Min(X)} \quad (1)$$

Where *Xnew* is normalization result, *Xi* is the current value, *Min(X)* is lowest value and *Max(X)* is the highest value of data.

Furthermore, after the normalization result has obtained, the next step is performing ranking to rank data from the highest to the lowest name frequency weight using sort method. Through this method, a list of hundred names order comprises first name, middle name, and the last name was obtained. These names then were classified into three clusters using the K-Means method, namely: very popular, popular and quite popular names. K-Means was chosen since it has popularly known that it can be used to analyze as well as processing data with high dimension, efficiency, scalability, and simplicity [38], [39].

K-Meansmethod classifies the data into groups based on certain characteristics so that each group holds specific characteristics that differ with other groups. There are two separate stages in the K-means clustering process. *First*, calculates the k centroid; and *second*, calculates the distance of each cluster point (based on centroid value) from the respective data point [40] until stable cluster's members are defined. There are several methods for calculating distances between centroids, and the most commonly used is *Euclidean distance.* In this paper, we select Euclidean distance function as the distance method, due to its function which is effective for a small set of data [41]. The steps of K-means clustering are described by itemization a to e [42], while the flowchart for applying the K-Means algorithm and ranking method is explained in Figure 2.

1. Choose the number of clusters (*k*=3)
2. Choose *Ci* centroids arbitrarily as the initial centers of the clusters.
3. Grouping the objects into groups which has nearby centroid. Assign each object to their closest cluster center using Euclidean distance. The formula of Euclidean distance as shown in Formula 2.

$$D(p,c)_n = \sqrt[2]{\sum_{i=0}^{n} (pi - ci)^2} \quad (2)$$

Where *p* is data, *c* is centroid (midpoint), *n* all data and *i* is iteration.

4. Count a center of new cluster by compute average points. If the membership object still changes, go back to step b
5. Stop when the object membership no change in eachcluster.

2.2 Data Collection

Social Media has experienced dramatic growth in the past few years. It can provide space to expand our social relationships between individuals and people. A lot of new features are added to enhance the convenience of social media users. One of them is a user profile name feature. The feature as the identity of a person on social media.

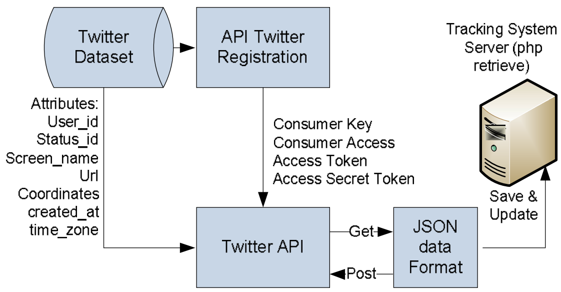

Figure 3: Data collection method architecture

To collect Tweet data, we used the Twitter API. The APIs provide various information as an example; id, name, screen_name, created_at and location. The Twitter API requires consumer key, consumer access, access token, and secret access tokens obtained by registering the Twitter API app at http://dev.twitter.com (see Figure 3). The created system works based on the location coordinates parameters. A 14.500 data is analyzed from Twitter containing the user name focused on Makassar City, Indonesia.

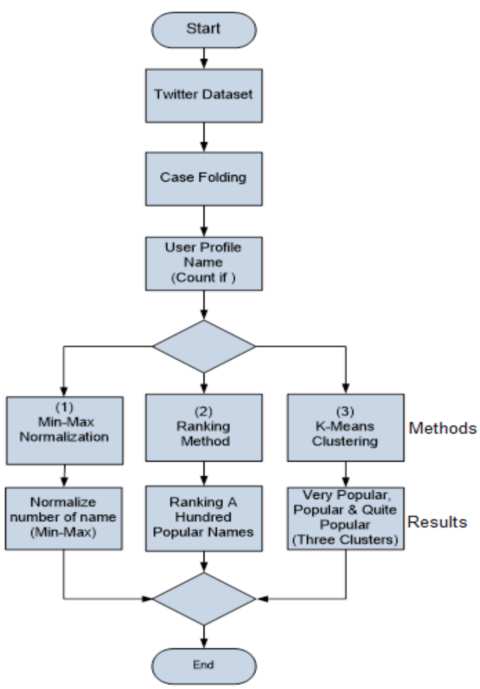

Figure 2: Research flowchart methodology

3. RESULT AND DISCUSSION

3.1Names Grouping With K-Means

Figure 4 shows the result of the name clustering using the K-means method. Names grouped into three clusters, namely: cluster-2 as *very popular* name, cluster-1 *popular* name, while cluster-0 as a *quite popular* name. Cluster-2 is a *very popular name* group has two members (as can be seen in Table 1, 2 and III). It can be addressed that two very popular first names are "Andi" with frequency 1255 (18.3%) users dan "Muhammad" with frequency 941 (12,7%) users. While two very popular middle names are "Nur" with frequency 137 (2%) users and "Didit" with frequency 126 (2%) users. Lastly, for very popular last names are "Ayunita" with frequency 150 (2%) users dan "Zulnizar" with frequency 122 (1.6%) users. These numbers representing the total person using each name in our entire sample.

Table 1: Members of cluster 2 (very popular) of first name

No	First Name	Total	Normalization	%
1	Andi	1355	1	18,3
2	Muhammad	941	0,69	12,7

Table 2: Members of cluster 1 (very popular) of middle name

No	Middle Name	Total	Normalization	%
1	Nur	137	1	2,0
2	Didit	126	0,911	2,0

Table 3: Members of cluster 1 (very popular) of last name

No	Last Name	Total	Normalization	%
1	Ayunita	150	1,0	2,0%
2	Zulnizar	122	0,807	1,6%

Cluster-1 as the *popular name* group has eight members. The number of users using the name in this group for the first name ranged between 103 to 652 (4% to 8.8%) users, for middle name 58 to 120 (1% to 2%) users, and for last name 29 to 87 (0,4% to 1.2%) users. In normalization form, weight values of cluster-1 are ranged between 0,063-0,474 for first name, 0,36 to 0,86 for middle name, and 0,166 to 056 for last name. In this cluster, the name "Nur" listed as the highest for first name, "Ramadhan" for a middle name, and "Pratiwi" for the last name. While for lowest, "Siti" for first name, "Eka" for a middle name, and "Pratama" for the last name. Detail information regarding members of cluster-1 is presented in table 4, 5 and 6.

Table 4: Members of cluster 1 (popular) of first name

No	First Name	Total	Normalization	%
3	Nur	652	0,474	8,8
4	Ayu	282	0,197	3,8
5	Nurul	264	0,183	3,6
6	Indah	238	0,164	3,2
7	Dian	205	0,139	2,8
8	Sri	130	0,083	1,8
9	Rahma	124	0,079	1,7
10	Siti	103	0,063	1,4

Table 5: Members of cluster 1 (popular) of middle name

No	Middle Name	Total	Normalization	%
3	Ramadhan	120	0,863	2
4	Putri	108	0,766	1
5	Dwi	96	0,669	1
6	Hidayat	83	0,565	1
7	Pratiwi	69	0,452	1
8	Ayu	66	0,427	1
9	Nurul	64	0,411	1
10	Eka	58	0,363	1

Table 6: Members of cluster 1 (popular) of last name

No	Last Name	Total	Normalization	%
3	Pratiwi	87	0,566	1,2
4	Sari	79	0,510	1,1
5	Putri	75	0,483	1,0
6	Putra	65	0,414	0,9
7	Victory	40	0,241	0,5
8	Lestari	39	0,234	0,5
9	Nur	33	0,193	0,4
10	Pratama	29	0,166	0,4

The last cluster or cluster-0, depicted as a group of a *quite popular name* has ninety members. The number of users using the name in this group for a first name ranged between 18 to 102 users, middle name 13 to 55users, and last name 5 to 129 users. In cluster-0, the name “Ahmad” has the highest weight for the first name with 1,37%, followed by “Muhammad” for a middle name with 0,74%, and “Syam” for the last name with 0.39%. While for lowest weight, the name “Aisyah” uses by 18 (0,24%) users for a first name, followed by the name “Azis” uses by 13 (0,18%) users for a middle name, and the name “Fauzan” uses by 5 (0.07%) users for the last name. Complete information about member of cluster-0 is provided in table 7, 8 and 9

Table 7: Members of cluster 0 (quite popular) of first name

No	First Name	Total	Normalization	%
11	Ahmad	102	0,06	1,37
12	Dwi	97	0,06	1,31
13	Putri	90	0,05	1,21
14	Dewi	89	0,05	1,20
15	Eki	81	0,05	1,09
16	Dinda	80	0,05	1,08
17	Ria	70	0,04	0,94
18	Annisa	69	0,04	0,93
19	Widya	69	0,04	0,93
20	Aris	55	0,03	0,74
⋮	⋮	⋮	⋮	⋮
98	Arif	18	0,00	0,24
99	Akbar	18	0,00	0,24
100	Aiya	18	0,00	0,24

Table 8: Members of cluster 0 (quite popular) of middle name

No	Middle Name	Total	Normalization	%
11	Muhammad	55	0,34	0,74
12	Ramadhani	55	0,34	0,74
13	Amalia	51	0,31	0,69
14	Indah	51	0,31	0,69
15	Wahyuni	51	0,31	0,69
16	Sari	49	0,29	0,66
17	Aulia	48	0,28	0,65
18	Pratama	48	0,28	0,65
19	Ramadhan	47	0,27	0,63
20	Akbar	45	0,26	0,61
⋮	⋮	⋮	⋮	⋮
98	Faisal	13	0,00	0,18
99	Chandra	13	0,00	0,18
100	Azis	13	0,00	0,18

Table 9:Members of cluster 0 (quite popular) of last name

No	Last Name	Total	Normalization	%
11	Syam	29	0,17	0,39
12	Saputra	28	0,16	0,38
13	Utami	27	0,15	0,36
14	Utami	26	0,14	0,35
15	Makassar	25	0,14	0,34
16	Azis	21	0,11	0,28
17	Qifly	19	0,10	0,26
18	Rahman	18	0,09	0,24
19	Wahyuni	17	0,08	0,23
20	Ahmad	16	0,08	0,22
⋮	⋮	⋮	⋮	⋮
98	Darwis	5	0,00	0,07
99	Fadilah	5	0,00	0,07
100	Fauzan	5	0,00	0,07

3.2Ranking Result

To identify which names are prioritizing to be used by people, we implement a ranking method. This approach to short names based on their frequency, from the highest to the lowest score. The result then ordered using numbers (1, 2, 3 and so on) from the top to the last (See table 1, 2, and 3). Figure 5 presents a graphic of the top ten popular names, including first name, middle name as well as last name.

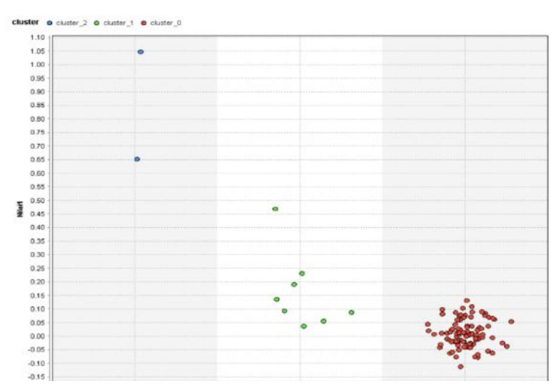

Figure 4: K-Means clustering for names; very popular (cluster 2,), popular (cluster 1,) and quite popular (cluster 0)

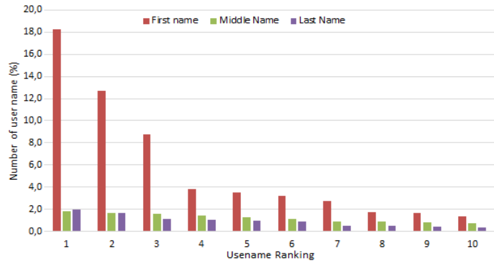

Figure 5: Graphic of the top ten popular names: first names, middle and last name based on ranking method

3.3 Name Characteristic

Another issue addressed in this research in the name characteristics i.e. name length (number of characters) and the first letter of each name. Figure 6 shows the classification of name based on its length which identified with C2 to C11. Where, C2 means name consist of two characters, C3 consists of three characters, until C11 which means name consist of eleven characters. From the whole data, name belongs to C2 (two characters) are seven users located at the last name. The name belongs to C3 on first names are 1658 users, middle names 322 users, and last names 55 users. Followed by C4 with 2586, 164, and 231; C5 with 1883, 917, and 422; C6 with 257, 475, and 173; C7 with 47, 544 and 523; and C8 with 966, 338, 170 users for the first name, middle name and last name respectively. While for C9 only in middle name with 246 users and last name and 25 users. In contrast, C10 only in the first name with 22 users and last name 42 users. Lastly, the name consists of eleven characters (C11) only in the middle name with 14 users.

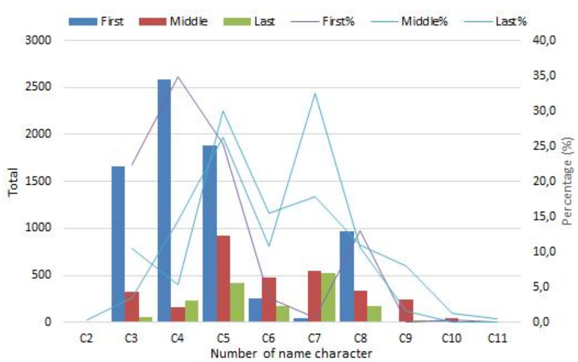

Figure 6: Minimum and maximum character length of first, middle and last name

Furthermore, the result of the comparison for the first letter of each name is presented in Figure 7. It shows the percentage of each letter (A-Z) appeared as the first letter in the first name, middle names as well as last names. For instance, some name begins with character “A”, “B”, and so on until “Z”. As can be seen in Tables 1, to 9, it can be addressed that character “A” is the most usable letter for the first name with percentage 34%, followed by character “M” and “N” with percentage 15% dan 14% for each (See Figure 8). While for the Middle names, character “A” also becomes the most usable first letter with a percentage 22%, followed by character “R” with 14% and character “P” with 10% (See Figure 9). Finally, for the last name also dominated by the character “A” with a percentage 25%, followed by character “P” with 18% and character “S” with 12% (See Figure 10).

3.4Discussion

The proposed methods can group personal names according to their cluster. Moreover, the methods used can also identify the most widely used personal names for the first, second, and third names. Then we count individual name character length to find out the number of letters in each name. Last, displayed the percentage of the first letter character for each class; first, middle and last names. The importance of this research can be a recommendation for anyone, specifically looking for a popular name based on first, middle, and last name to be named in a person. The results of this study reveal that a name with certain identity and characteristic remains is the main priority of parents to name to their children. for example, names with religious identity and culture or ethnic identity are the main choice. In the location of this research, name of Muhammad and Andi are two names sourced from religion and ethnic identities (see table 1).

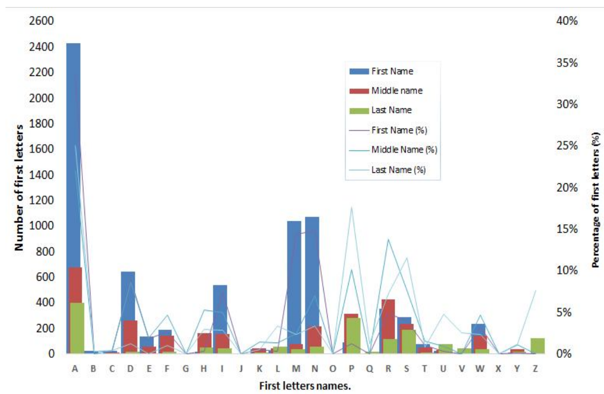

Figure 7 Comparison graphics of the first letter of user’s first name, middle names and last names.

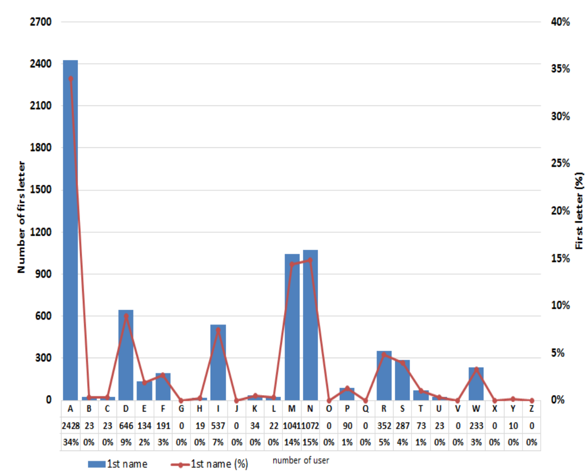

Figure 8: Percentage of the first letter in the first name

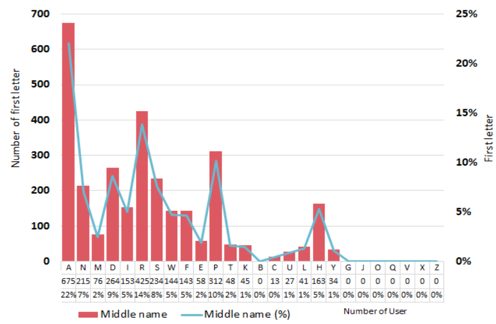

Figure 9: Percentage of the first letter in the middle name

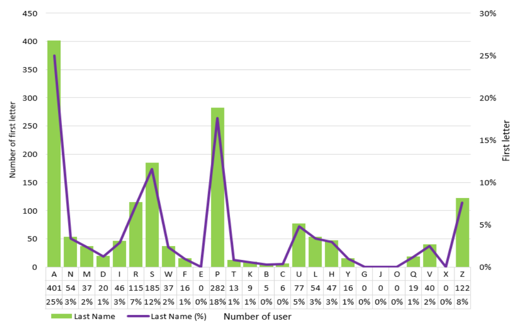

Figure 10: Percentage of the first letter in the last name

5. CONCLUTION

There are three clusters obtained in grouping names extracted from social media Twitter user's profiles in Makassar, namely: cluster-2, cluster-1, and cluster-0. These cluster representing a group of very popular names of cluster two, popular names of cluster one and quite popular names respectively of cluster 0. Each cluster defined based on three criterias, namely: first name, middle name and last name. Based on the research results it is found that (from a hundred names incorporate in this research) cluster-2 has two members, cluster-1 with eight members, and cluster-0 with ninety members. With ranking method, this research produced top ten popular names, where first top 3 are distributed in cluster-2 and cluster-1, top five are also distributed in cluster-2 and cluster-1, and top ten names spread in cluster-2, cluster-1, and cluster-0 as well (see table 1 to 6). This implied that clustering with K-Means resulting in equal cluster's members distribution. Furthermore, it is found that the minimum length of a name in this research is two characters and the longest name length is eleven characters. While for the first character of the name, most name begins with characters A, N, dan M in the first name, followed by characters A, R, dan P for the middle name, and characters A, P, dan S for the last name (see figure 7 to 10).

AUTHOR CONTRIBUTIONS

Najirah umar wrote most of the article, Yuyun collected the data, Hazrianicomposed and designed the analysis, and Herman Efendi Performed the analysis. All authors are major contributors

REFERENCES

[1] W. Deprez, S. Indiana, and N. March, "**Trustees of Indiana University Review**," vol. 59, no. 1, pp. 81–82, 2015.

[2] S. Brink, "**Transferred Names and Analogy in Nameformation**," January 2015, pp. 158–166, 2015.

[3] H. Zhu, "**A Cultural Comparison of English and Chinese Family Names**," *Asian Soc. Sci.*, vol. 6, no. 11, pp. 226–228, 2010, doi: 10.5539/ass.v6n11p226.

[4] A. Aksholakova, "**Proper Name as a Clue Symbol of Identity**," *Procedia - Soc. Behav. Sci.*, vol. 112, no. Iceepsy 2013, pp. 465–471, 2014, doi: 10.1016/j.sbspro.2014.01.1190.

[5] S. Mutunda, "**Luvale Personal Names and Naming Practices : A Socio-Cultural Analysis**," vol. 1, no. 3, pp. 75–81, 2017, doi: 10.11648/j.ijecs.20160103.12.

[6] F. Ushuluddin, D. A. N. Humaniora, U. Islam, and N. Walisongo, "**Pemberian nama kepada anak menurut hadis**," 2017.

[7] M. Watzlawik *et al.*, "**First Names as Signs for Personal Identity – An Intercultural Comparison**," vol. 8, no. 1, pp. 1–21, 1997.

[8] J. S. Odell, "**Names of Persons Interviewed,**" ***U.S. Int. Monet. Policy***, no. January, pp. xiv–xvi, 2015, doi: 10.1515/9781400856299.xiv.

[9] A. W. Q. G. Al-Zumor, "**A socio-cultural and linguistic analysis of Yemeni Arabic personal names**," *GEMA Online J. Lang. Stud.*, vol. 9, no. 2, pp. 15–27, 2009.

[10] R. R. Troiden, ***Self, Self-Concept, Identity, and Homosexual Identity***, vol. 10, no. 3–4. 1984.

[11] M. E. Mabuza, "**Individual Names and Personality: A Consideration of some Beliefs**," pp. 833–841, doi: 10.2436/15.8040.01.86.

[12] B. Helleland, "**Personal Names and Identity in Literary Contexts,**" *Oslo Stud. Lang.*, vol. 4, no. 2, pp. 273–284, 2012, [Online]. Available: http://www.journals.uio.no/osla.

[13] P. Marrone, "**Chambers, RT,**" ***Etica e Polit.***, vol. 15, no. 1, pp. 583–605, 2013, doi: 10.1093/acprof.

[14] M. Dixson, E. Bermes, and S. Fair, "**An instrument to investigate expectations about and experiences of the parent-child relationship: The parent-child relationship schema scale**," *Soc. Sci.*, vol. 3, no. 1, pp. 84–114, 2014, doi: 10.3390/socsci3010084.

[15] T. Glenz, "**The Importance of Learning Students' Names**," *J. Best Teach. Pract.*, no. April, pp. 21–22, 2014, [Online]. Available: http://teachingonpurpose.org/wp-content/uploads/2015/03/Glenz-T.-2014.-The-importance-of-learning-students-names.pdf.

[16] I. Boamfa, "**Preliminary Considerations on the Anthroponymy Research Methodology**," *4th SGEM Int. Multidiscip. Sci. Conf. Soc. Sci. ARTS Proceedings4th, Sci. Humanit.*, vol. 4, no. September, 2017, doi: 10.5593/sgemsocial2017/41/s15.008.

[17] A. Portwood, "**All in a Name : Parental Naming Activity in Multilingual Manchester**," no. October, 2017.

[18] S. T. Widodo, "**The Development of Personal Names in Kudus, Central Java, Indonesia**," *Procedia - Soc. Behav. Sci.*, vol. 134, no. 271, pp.

154–160, 2014, doi: 10.1016/j.sbspro.2014.04.234.

[19] S. T. Widodo, "**Personal Names as an Inter-Ethnic Model of Acculturation in Indonesia**," *Asian J. Soc. Sci. Humanit.*, vol. 4, no. 1, pp. 126–133, 2015.

[20] S. T. Widodo, "**Javanese names during the height of the hindu-buddhist kingdoms in java: An ethnolinguistic study**," *Kemanusiaan*, vol. 20, no. 2, pp. 81–89, 2013.

[21] E. K. Aribowo and N. Herawati, "**Trends in Naming System on Javanese Society: A Shift From Javanese to Arabic**," *Ling. Cult.*, vol. 10, no. 2, p. 117, 2016, doi: 10.21512/lc.v10i2.1730.

[22] Yuyun, F. A. Nuzir, and B. J. Dewancker, "**Dynamic land-use map based on twitter data**," *Sustain.*, vol. 9, no. 12, pp. 1–20, 2017, doi: 10.3390/su9122158.

[23] Yuyun, F. A. Nuzir, and B. J. Dewancker, "**Quantifying Mobility of Urban Inhabitant Based on Social Media Data**," vol. 12, no. 6, pp. 401–405, 2018, doi: doi.org/10.5281/zenodo.1316837

[24] S. Hasan, X. Zhan, and S. V Ukkusuri, "**Understanding Urban Human Activity and Mobility Patterns Using Large-scale Location-based Data from Online Social Media**," *Proc. 2Nd ACM SIGKDD Int. Work. Urban Comput.*, pp. 6:1--6:8, 2013, doi: 10.1145/2505821.2505823.

[25] S. Hasan and S. V. Ukkusuri, "**Location contexts of user check-ins to model urban geo life-style patterns**," *PLoS One*, vol. 10, no. 5, pp. 1–19, 2015, doi: 10.1371/journal.pone.0124819.

[26] V. Frias-Martinez and E. Frias-Martinez, "**Spectral clustering for sensing urban land use using Twitter activity,**" *Eng. Appl. Artif. Intell.*, vol. 35, pp. 237–245, 2014, doi: 10.1016/j.engappai.2014.06.019.

[27] Y. Sun, H. Fan, M. Li, and A. Zipf, "**Identifying the city center using human travel flows generated from location-based social networking data**," *Environ. Plan. B Plan. Des.*, vol. 43, no. 3, pp. 480–498, 2016, doi: 10.1177/0265813515617642.

[28] S. Chandra, L. Khan, and F. Bin Muhaya, "**Estimating twitter user location using social interactions - A content based approach**," *Proc. - 2011 IEEE Int. Conf. Privacy, Secur. Risk Trust IEEE Int. Conf. Soc. Comput. PASSAT/SocialCom 2011*, pp. 838–843, 2011, doi: 10.1109/PASSAT/SocialCom.2011.120.

[29] Yuyun, F. A. Nuzir, and B. J. Dewancker, "**Analysis of Urban Population Using Twitter Distribution Data: Case Study of Makassar City, Indonesia**," vol. 10, no. 10, pp. 1811–1815, 2016, [Online]. Available: https://publications.waset.org/10009070/quantifying-mobility-of-urban-inhabitant-based-on-social-media-data.

[30] T. M. Hiele, A. E. Widjaja, J. V. Chen, and T. Hariguna, "I**nvestigating students' collaborative work to continue to use the social networking site**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.5 Special Issue, pp. 375–386, 2019, doi: 10.30534/ijatcse/2019/6181.52019.

[31] Candiwan and F. Savindraputra, "**Is information privacy awareness important for Indonesian social media instagram users?**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.5 Special Issue, pp. 281–287, 2019, doi: 10.30534/ijatcse/2019/4981.52019.

[32] I. K. Sumerta, I. G. P. A. Widyagoca, and I. W. Meryawan, "**Online consumer behavior on using social media on E-commerce, based on the AISAS model approach. Case study; Bukalapak, Tokopedia and Blili.com**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.5 Special Issue, pp. 234–242, 2019, doi: 10.30534/ijatcse/2019/4281.52019.

[33] N. N. Kamaruzaman, N. A. Husin, N. Mustapha, and R. Yaakob, "**Towards a multimodal analysis to predict mental illness in twitter platform**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.4 S1, pp. 126–130, 2019, doi: 10.30534/ijatcse/2019/1981.42019.

[34] Y. YASLAN, H. GÜLAÇAR, and M. N. KOÇ, "**User Profile Analysis Using an Online Social Network Integrated Quiz Game**," *Süleyman Demirel Üniversitesi Fen Bilim. Enstitüsü Derg.*, vol. 21, no. 3, p. 696, 2017, doi: 10.19113/sdufbed.20506.

[35] A. Jaech and M. Ostendorf, "**What your username says about you**," *Conf. Proc. - EMNLP 2015 Conf. Empir. Methods Nat. Lang. Process.*, no. September, pp. 2032–2037, 2015, doi: 10.18653/v1/d15-1240.

[36] A. Romanov, A. Semenov, and J. Veijalainen, "**Revealing fake profiles in social networks by longitudinal data analysis**," *WEBIST 2017 - Proc. 13th Int. Conf. Web Inf. Syst. Technol.*, no. Webist, pp. 51–58, 2017, doi: 10.5220/0006243900510058.

[37] Y. Li, Y. Peng, Z. Zhang, Q. Xu, and H. Yin, "U**nderstanding the user display names across social networks**," *26th Int. World Wide Web Conf. 2017, WWW 2017 Companion*, pp. 1319–1326, 2019, doi: 10.1145/3041021.3051146.

[38] M. A. A. Riyadi, D. S. Pratiwi, A. R. Irawan, and K. Fithriasari, "**Clustering stationary and non-stationary time series based on autocorrelation distance of hierarchical and K-means algorithms**," *Int. J. Adv. Intell. Informatics*, vol. 3, no. 3, pp. 154–160, 2017, doi: 10.26555/ijain.v3i3.98.

[39] D. P. Ismi, S. Panchoo, and Murinto, "**K-means clustering based filter feature selection on high dimensional data**," *Int. J. Adv. Intell. Informatics*, vol. 2, no. 1, pp. 38–45, 2016, doi: 10.26555/ijain.v2i1.54.

[40] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "**Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm**," *Procedia Comput. Sci.*, vol. 54, pp. 764–771, 2015, doi: 10.1016/j.procs.2015.06.090.

[41] M. Hasan, "**Comparison of Euclidean Distance Function and Manhattan Distance Function Using K-Mediods**," vol. 13, no. 10, pp. 61–71, 2015.

[42] P. Lama, "**Clustering System Based K-Means Algorithm** – News headlines clustering," 2013.