

## A Methodology for Brand Name Hierarchical Clustering Based on Social Media Data

Tasweer Hussain Shah<sup>1</sup>, Nasir Naveed<sup>1</sup>, Zahid Rauf<sup>2</sup>

<sup>1</sup>Department of Computer Science and Information Technology, Virtual University of Pakistan, <sup>2</sup>Department of Electrical Engineering, Faculty of Information and Communication Technology, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan

### Abstract

*Clustering of brand names is a common technique for identification of different groups of brand names with maximum effectiveness and similarities. In this study, we proposed a methodology for brand name clustering, based on data from social media like Twitters. Dataset consist of tweets mentioning various garment brands in them. In order to cluster the brand names, we have proposed an algorithm, named BNACA (Brand Names Agglomerate Clustering Algorithm), an extension to the standard hierarchical clustering algorithm. In the proposed algorithm, we used single linkage as a similarity measure. The proposed clustering algorithm provides consistent clustered results for various sets of brand names of garment industry. Finally, clusters of garment brand names are visualized through dendogram. The dendogram clearly shows the maximum similarities among garment brand names.*

**Key words:** Social Media Data; Hierarchical Clustering; Explicit Semantic Analysis; Twitter

**Corresponding author's email:** [nasir@vu.edu.pk](mailto:nasir@vu.edu.pk)

### INTRODUCTION

Social media such as Twitter, LinkedIn and Facebook provide new platforms to users for expressing their opinion. Different organizations collect data from various social media platforms to help brands understanding their relationship with users. These organizations provide different application programming interface (API) for different tasks including sentiment analysis. Normally, the main aim of different organizations is to establish a relationship with their customers under a customer relationship management (CRM). In this research work, the effect of Twitter has been studied. We collected and manually annotated a data set of tweets. Our data set is based on posts about garment brand names. Clustering techniques offer precise methods for better reasoning in the social media data. In hierarchical clustering, hierarchical algorithm creates a hierarchical classification of different objects. There are two types of hierarchical clustering, Agglomerate and Divisive. In general, the agglomerative methods are mostly used. On each step, the pair of clusters with smallest cluster-to-cluster distance is merged into a single cluster and finally all the objects are grouped into a single cluster. In this paper, we proposed an algorithm BNACA which provides distances (similarities) between different garment brand names at each step. The proposed methodology helps advertisers to determine different groups of brand names on the basis of customers' opinions. It also helps consumers to find similarities between different brand names. This information is further used to develop effective marketing programs. It also increases the interest in finding similarities in the brand names of an industry and assists advertisers on the basis of social media data to increase customer's satisfaction. In this way, advertisers can achieve the goal of establishing meaningful business. In the last 20 years, a drastic change has taken place as to how business is conducted. The development of personal computers, the internet and e-commerce has a great impact for function and endorsement of marketing. The preface of social networking sites tools is increasing and has the same impact on business. As new technology is increasing day by day, businessman learns to use these technologies. Some important technology driven companies are Microsoft, eBay, Amazon and Google etc. The communication through social media platforms has become easier and user friendly; that is why social media grows very fast (Brown and Kietzmann, 2018). The previous research can be divided into two main categories. The first category illustrates the gathering of social media data and second explains the effect of the social media on brand images through different clustering techniques (Kang et al., 2015).

### Brands Analysis for Social Media Data

In order to better understand about brands and how a brand is carried out in the marketplace, organizations pay an extensive variety of measures. In this regard, the consumer-based reviews regularly play a very important role (Poecze et al., 2018). James and Rowley (2017) developed the social media marketing (SMM) framework. For this, they conducted interviews from social media marketing agencies. The framework consists of six important stages i.e. setting assessment goals, identifying key performance indicator (KPI), classifying into different levels, information gathering and processing, preparation of report. Challenges related to all stages of the framework are recognized and explained to making a better decision related to social media techniques. Shamsher et al. (2016) identified the role of social media marketing for developing brands. They also identified influence of consumers in order to make online buying. Their study consisted of the data from 256 respondents. Their factor analysis for marketing consists of factors like Brand impact, Brand recognition, Brand approval, Brand association and Brand reliability. These factors show the role of social media marketing in building brands. Chakraborty et al. (2016) collected data from social media like Twitter and Facebook in the form of public online posts related to fast fashion brands. They proposed a novel based collaborative model on the base of the N – grams in order to deal with the diverse nature of data collected from these social media platforms. Their framework is based on Naïve Bayes classifier with N-grams and consists of two algorithms for text analysis. They compared these different N-gram models in the different numerical performance metrics such as Accuracy, Precision, Recall and F1- score. This framework provides a good road - map for text analysis and text classification. Kang et al. (2015) designed a framework, to analyze the brands from social media data. She also performed a case study for supporting her framework. However, this framework uses data from social media without any text analysis. Carbonell et al. (2014) suggested a method for measuring the total number of drugs brands and its features, by using Twitters data. In their study, they used Twitter API (Application Programming Interface) and Drug bank database for the collection of data from social media like Twitter. It shows an analysis and detection of side effects of drug and an interaction between drug brand names. This analysis provides a strong platform for drug associations. Results showed that different messages on Twitters are good for useful drug related information. Vendors have many choices within the social media platforms for the marketing of brands. For example, advertising of brands and products, brand Variety, emerging different occasions for customer participation, and distributing content about different brands in social media sites (Tuten and Solomon, 2013). Vries et al. (2012) examined the effects of social media marketing using the brands posts. These brands posts consist of video clip, messages, quizzes and information about the brands from social media. Their research shows that a huge number of user's effects on the numbers of "Likes" and "comments" on brands. These "likes" and "comments" on brands posts about social media show brand post popularity. Customer engagement is another very important factor in the marketing. A research in 2012 showed that about 78% of vendors' discussions on the social media platforms help to increase customer branded engagement. Customer engagement about brands and products is established on customers' behavior and firm focus on their customers etc. Customers may engage due to five characteristics including value, form, scope, effect and aims (Doorn et al., 2012). In social media platforms, engaged users contribute and communicate information about products, brands etc. Participation of users may be passive or active. Passive participation includes simply consuming the social media content and active participation includes behaviors like submitting user generated stories (Hutton and Fosdick, 2011). According to Jackson (2011), social media platforms enable organizations to interconnect with their users. Social media platforms also allow users to connect with each other. These interconnections establish a brands consistency between organizations and their customers. This helps for the advertising of products and other facilities. This also establishes online groups among brand fans. Moreover, sharing among different brand customers helps firm to increase the more information about brands and products (Gunelius, 2011). Constantinides et al. (2011) studied different strategies of marketing. In their study, they explained strategies of marketing like Marketing Intelligence, Promotions of Brands, Customer interconnection, Firm and customer management and Communications for marketing etc. According to them, these strategies help for exploring social media because usage of internet is creating an interest among users and users considered data shared on social media as more consistent than directly to different organizations. Thus, users may communicate their feelings about the brands and products with others. In the social media platforms, response of users is dignified through users' contacts, bookmarks, blogs, discussions, clicks, comments, subscribes, and buys the brand (Falls, 2010). Bernoff and Li (2008) stated that advertising of brands is not restricted to the up - graduation of products and facilities. It includes the elevation of brands and likings of

customers about products and facilities. This creates a challenge for organizations whose marketing strategies must incorporate a new reality where the consumers often desire a link with the organizations or the brands. Communicating with the consumers involves using the many tools available to the organization. These tools are targeted in an integrated manner so that messages remain focused, differentiated, and relevant. While traditional media such as TV, radio and print etc. have become less effective, new tools have developed to supplement that media, e.g. Web 2.0 (Singh et al., 2008). Hence, advertising managers have long trusted the estimate of consumers observation of brand. These brands along with the attributes of interest are used to show marketing strategy (Lehmann et al., 2008).

### **Brands and Clustering Analysis**

Data mining is a very important field where knowledge is extracted from the large database. For this purpose, many techniques are used. Clustering is one of the important techniques used for the data mining. Clustering technique is an adaptive methodology in which objects are merged together (Salloum et al., 2018). Basically, Cluster analysis is a technique used for division of statistics. In this technique, each part of the organization is called a cluster. It provides maximum number of the clusters' prototypes (Chitral and Maheswari, 2017). Cluster evaluation is a decision making method and a creativity technique in getting ideas and thoughts. The hierarchical clustering algorithm is one from different clustering algorithms which is used for cluster evaluation (Goyal et al., 2016). Anaza et al. (2016) performed a cluster analysis through the heterogeneous nature of social project consumers. This analysis is based on consumers' usage of the social media sites. In the study, they used 305 consumers from numerous social projects and divided these consumers into four unique groups called clusters. These groups are observers, contributors, connectors and contributors. These groups are based on three important characteristics of social media usage. These characteristics are creating content, connecting with others and managing the users. On the basis of survey analysis, they showed that the four segments are different in their supporting behavior of consumers. Kang et al. (2015) designed a framework to analyze the brands from social media data and perform a K – mean cluster analysis on these brands. They also performed a case study for supporting their framework. However, this framework used data from social media without any text analysis. Kyung et al. (2014) suggested a car designed model for different automobile manufacturing. This model has two important objectives. The first objective is to find similarities between different models from different car brands. It is necessary to design a distinctive car model. Secondly, the classification of the design properties is performed through hierarchical clustering. Their case study consisted of 120 to 123 car brands which contribute an important role for design decision. The most functional clustering approach in economy is hierarchical clustering analysis. Hierarchical clustering is constructed on a proximity matrix. This matrix contains the similarity calculation for all elements of entities. Furthermore, several techniques such as single linkage technique can be used for measurement of the cluster similarity. Apart from providing an opportunity to analyze document files with qualitative variables, the most important benefit of this type of clustering technique is a graphical output (Rezankova, 2014). Mohbey and Thakur (2013) accomplished a test on two dimensional spaces where several objects were accessible and combined in clusters by Euclidean distance method. In their study, they used MATLAB™ to compute the minimum distance among several objects and constructed distance matrix. They also prepared dendrogram. This dendrogram is like a minimum spanning tree. It is constructed using minimum distance among several objects. Hierarchical clustering technique is divided into agglomerative clustering technique and divisive clustering technique. The agglomerative hierarchical clustering technique has bottom- up approach and starts with each object as a unique cluster and merges them into larger clusters. This clustering is also called bottom-up clustering. On the other hand, divisive hierarchical clustering technique has the top-down approach. This technique begins with a single cluster containing all elements, and then splits this cluster into resulting clusters (Verma et al., 2012). Saraf and Patil (2013) presented market-basket study using Agglomerative hierarchical technique for clustering a retail item. In their study, they explained that market-basket analysis is widely used in retail stores to make the task of planning efficient. A retail store is considered the best place for achieving market-basket analysis. Because, the items in retail stores are organized in suitable and systematic manner so that a customer can select the product from many options available and buy according to his/her needs. Their work is based on the second type of agglomerative hierarchical clustering techniques where cluster centers are specified to consider dissimilarities between two different items. Lee and Bradlow (2011) examined how to aggregate brands and products attributes in a hierarchical way. They used Meta – attributes to show the package of product attribute at a high level. However, they addressed only single layer for Meta - attributes. For example, in case of digital camera, they present three products Meta – attributes such as auto focus,

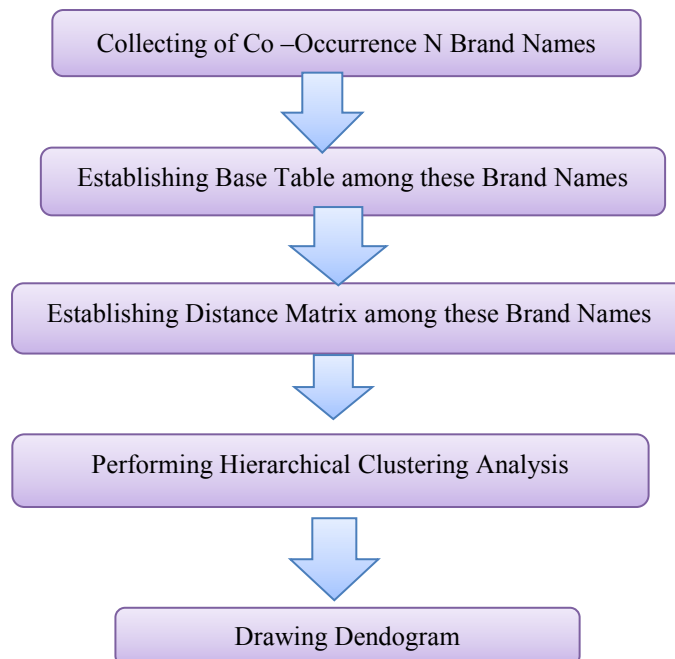
manual focus and PC connect. These three attributes are separate and controlled by different physical components in a camera, yet auto focus and manual focus likely lie in much closer proximity in consumers' perceptual spaces than do autofocus and PC connect, because they refer to the same usage situation i.e. focusing before taking pictures. Shelestov (2008) explained the notion of clustering and agglomerative hierarchical clustering algorithm. He showed how a clustering is applied for the analysis of stocks and traded in order to identify similar temporal behavior of the traded stocks. As inputs he took data for 16 stocks from the Bulgarian Stock Exchange in a single day. It contains information for each stock as the code and the name of the company, prices, the change in price in comparison to the previous day and the traded amount of this kind of stock. He used the data mining tool named XLMiner™ for MS Excel. He selected the agglomerative method of hierarchical clustering to find clusters of stocks. He performed five variants of agglomerative method of hierarchical clustering and found that the average linkage method will give the best results. Hence, clustering is one of the techniques used to grouped objects together. It based on the principle of optimizing the inside class similarity reducing the class – class similarity (Chen et al., 2004).

## MATERIALS AND METHODS

This research focuses on the use of hierarchical clustering of brand names in social media such as Twitter. According to modern research on social media platforms, it is explained that the social media such as Twitter is the best source for providing timely information. It provides original response from customers to the brands, mostly regarding their involvements, views and queries etc. (Rutter et al., 2016).

To find maximum similarities among brand names in the social media data, the proposed method is given in figure 1. It has following steps.

- It collects N co-occurrence brand names.
- It establishes the base table among brand names.
- It establishes the distance matrix among brand names.
- It shows maximum similarities among different brand names through hierarchical clustering.
- It draws dendrogram.



**Figure 1:** Proposed method for hirarchical clustering of brand names

### Case study

To evaluate the performance of the proposed framework, we chose the brand names of garment industry. Twitter data set for the garment brand names is collected. It is better to select an industry which has optimal number of brand names for clustering, otherwise, small numbers of brand names may result in minor clusters and a very large number of brand names may become more complex for practical use.

### Data collection and distribution

We collected data in the form of tweets from Twitter over a period of 2 months. The ten (10) selected garment brands were Converse, Vans, Reebok, Adidas, New Balance, Nike, Puma, Brooks, Levis and Jordan. The official Twitter search API was used to fetch the tweets. Tweets containing selected brand names and hash tags with brand names are selected. Data was preprocessed for filtering such as, cleaning data of URLs, HTML tags, punctuation marks, emotions symbols and similar noise. Each tweet was manually annotated for brand name mentioned. For test phase, each post is assigned either of the two-brand names Co – Occurrence or non Co – Occurrence depending on the content.

**Table 1: Distribution of Training Data**

Brand Names	Brand Names	Co – occurrence	Non Co-occurrence	Total
Converse	Vans	406	516	922
	Reebok	33	40	73
	Adidas	110	30	140
	New – Balance	91	50	141
	Nike	109	06	115
	Puma	125	201	326
	Brooks	02	02	04
	Levis	04	100	104
Vans	Jordan	139	10	149
	Reebok	10	06	16
	Adidas	250	210	460
	New – Balance	10	07	17
	Nike	108	02	110
	Puma	37	51	88
	Brooks	10	04	14
	Levis	10	14	24
Reebok	Jordan	160	152	312
	Adidas	200	150	350
	New – Balance	50	14	64
	Nike	300	35	335
	Puma	10	09	19
	Brooks	86	44	130
	Levis	01	03	04
	Jordan	05	09	14
Adidas	New – Balance	20	22	42
	Nike	704	308	1012
	Puma	77	24	101
	Brooks	08	02	10
	Levis	09	02	11
	Jordan	04	03	07
New balance	Nike	135	76	211
	Puma	07	06	13
	Brooks	10	23	33
	Levis	03	02	05
	Jordan	106	09	115
Nike	Puma	89	02	91
	Brooks	10	15	25
	Levis	07	02	09
	Jordan	09	19	28
Puma	Brooks	10	07	17
	Levis	03	13	26
	Jordan	10	03	13
Brooks	Levis	08	81	89
	Jordan	03	36	39
Levis	Jordan	08	15	23
Total		3506	2335	5841

Table 1 lists the distribution of training data (posts). The garment brand names distribution is shown through total number of co – occurrences and non co - occurrences between brand names. The data set consists of 5841 garment brand names posts.

### Algorithm

The algorithms have been presented in the form of a pseudo code where each step has been clearly mentioned. It has been further grouped into separate training and testing phase. The algorithm takes N brand name sample data sets from social media as input and outputs total number of co – occurrence among N ( $N=1$  to  $n$ ) brand names.

It consists of following steps.

- Take N brand name data sets with brand label.
- Take a brand name  $B_i$  from sample data sets of brand names where  $i= 1$  to  $n$
- For a given test data calculate the co - occurrence as  $(O_i, j, B_j)$ , by comparing with brand name  $B_i$ . Where  $O_i, j$  is the Co - occurrence values output of the  $i$ th brand name with  $B_j$  brand name. where  $j= 1$  to  $n$
- Calculate the maximum values of  $O_i, j$ , to the final brand name.
- Repeat step 3, 4 and 5 until given brand names co - occurrences are calculated.

The output of algorithm is in the form of base table of co – occurrences of garment brand names as shown in the table 2. It validates the test data of garment brand names.

### Base table between the brand names

The base table showed the co - occurrences among ten garment brand names. The size of base table is '10'. The base table ' $B_{[10, 10]}$ ' is symmetric table, which means co – occurrences between two brands like Adidas and Nike is same as co – occurrence between Nike and Adidas.

**Table 2: Base table  $B_{[10, 10]}$**

Brand Names	Converse	Vans	Reebok	Adidas	New Balance	Nike	Puma	Brooks	Levis	Jordan
Converse	-	406	33	110	91	109	125	02	04	139
Vans	406	-	10	250	10	108	37	10	10	160
Reebok	33	10	-	200	50	300	10	86	1	05
Adidas	110	250	200	-	20	704	77	08	09	04
New Balance	91	10	50	20	-	135	07	10	03	106
Nike	109	108	300	704	135	-	89	10	07	10
Puma	125	37	10	77	07	89	-	09	03	08
Brooks	2	10	86	08	10	10	09	-	03	03
Levis	04	10	1	09	03	07	03	03	-	08
Jordan	139	160	05	04	106	10	08	03	08	-

### Distance matrix 'D'

The distance matrix ' $D_{[10, 10]}$ ' is established between garment brand names from the base table. The distance matrix is a square matrix that consists of ten garment brand names. So, its order is '10×10'. The distance matrix denotes the distances (similarities) among different garment brand names as presented in the table 3.

### Hierarchical clustering analysis

In this step, each brand name of garment industry is represented by one cluster. The purpose is to convert these ten clusters into one whole cluster, which shows these ten brand names. Here, single link clustering technique is used. It is most suitable for finding maximum similarities between brand names. We proposed the Algorithm BNACA (Brand Names Agglomerate Clustering Algorithm) for clustering analysis on brand names for finding maximum similarities between brand names.



**Table 3: Distance matrix D** <sub>[10,10]</sub>

Brand Names	Converse	Vans	Reebok	Adidas	New Balance	Nike	Puma	Brooks	Levis	Jordan
Converse	0	406	33	110	91	109	125	02	04	139
Vans	406	0	10	250	10	108	37	10	10	160
Reebok	33	10	0	200	50	300	10	86	1	05
Adidas	110	250	200	0	20	704	77	08	09	04
New Balance	91	10	50	20	0	135	07	10	03	106
Nike	109	108	300	704	135	0	89	10	07	10
Puma	125	37	10	77	07	89	0	09	03	08
Brooks	2	10	86	08	10	10	09	0	03	03
Levis	04	10	1	09	03	07	03	03	0	08
Jordan	139	160	05	04	106	10	08	03	08	0

**Algorithm: BNACA**

1. Begin with Brand Name Clusters Matrix D.
2. Initializing, clustering level as  $L(0) = 0$  and sequence number  $s = 0$
3. Find the pair of Brand Name clusters which have minimum distance in the current clusters matrix  
Say Brand Name clusters pair (A), (B) according to  $d(A), (B) = \min d(i, j)$   
Where  $\min d(i, j)$  is complete pairs of Brand Names clusters in the current clustering.
4. Increment in sequence number  $s = s + 1$
5. Merge pair Brand Name clusters (A) and (B) into a single cluster.
6. Set level of the clustering into  $L(s) = d(A), (B)$ .
7. Update Matrix,  $B_{[i,j]}$ ,  
By deleting the entries from row and column of corresponding to clusters (A), (B)  
And adding a row and column corresponding to newly formed cluster denoted (A, B) and old cluster (K) is defined as  
 $d[(K), (A, B)] = \min d[(K), (A)], d[(K), (B)]$
8. If all the Brand Names clusters are merged into one cluster, stop, else go to step 2.

**Example**

For simplicity of calculation, garment brand names are considered as:

Converse = "1", Vans = "2", Reebok = "3", Adidas = "4", New Balance = "5", Nike = "6", Puma = "7", Brooks = "8", Levis = "9" and Jordan = "10".

In every step, we checked the minimum distances among brand names from matrix. The minimum distance between cluster '3' and '9' is '1'. These two clusters get merged into one cluster "(3, 9)". Using the input distance matrix, distance between cluster (3) and cluster (9) is computed as: -

$$\begin{aligned}
 D_{(3,9) \rightarrow 1} &= \min(d_{31}, d_{91}) \\
 &= \min(33, 04) \\
 &= 04 \\
 D_{(3,9) \rightarrow 2} &= \min(d_{32}, d_{92}) \\
 &= \min(10, 10) \\
 &= 10 \\
 D_{(3,9) \rightarrow 4} &= \min(d_{34}, d_{94}) \\
 &= \min(200, 09) \\
 &= 09 \\
 D_{(3,9) \rightarrow 5} &= \min(d_{35}, d_{95}) \\
 &= \min(50, 03) \\
 &= 03
 \end{aligned}$$

$$\begin{aligned}
 D_{(3,9) \rightarrow 6} &= \min(d_{36}, d_{96}) \\
 &= \min(300, 07) \\
 &= 07 \\
 D_{(3,9) \rightarrow 7} &= \min(d_{37}, d_{97}) \\
 &= \min(10, 03) \\
 &= 03 \\
 D_{(3,9) \rightarrow 8} &= \min(d_{38}, d_{98}) \\
 &= \min(86, 03) \\
 &= 03 \\
 D_{(3,9) \rightarrow 10} &= \min(d_{310}, d_{910}) \\
 &= \min(05, 08) = 05
 \end{aligned}$$

Table 4: DATASET

Brand Names	1	2	3	4	5	6	7	8	9	10
1	0	406	33	110	91	109	125	02	04	139
2	406	0	10	250	10	108	37	10	10	160
3	33	10	0	200	50	300	10	86	1	05
4	110	250	200	0	20	704	77	08	09	04
5	91	10	50	20	0	135	07	10	03	106
6	109	108	300	704	135	0	89	10	07	10
7	125	37	10	77	07	89	0	09	03	08
8	2	10	86	08	10	10	09	0	03	03
9	04	10	1	09	03	07	03	03	0	08
10	139	160	05	04	106	10	08	03	08	0

Table 5: Minimum Distance Between Two Clusters (3 and 9)

B/N	1	2	3,9	4	5	6	7	8	10
1	0	406	<b>04</b>	110	91	109	125	02	139
2	406	0	<b>10</b>	250	10	108	37	10	160
<b>3,9</b>	<b>04</b>	<b>10</b>	<b>0</b>	<b>09</b>	<b>03</b>	<b>07</b>	<b>03</b>	<b>03</b>	<b>05</b>
4	110	250	<b>09</b>	0	20	704	77	08	04
5	91	10	<b>03</b>	20	0	135	07	10	106
6	109	108	<b>07</b>	704	135	0	89	10	10
7	125	37	<b>03</b>	77	07	89	0	09	08
8	2	10	<b>03</b>	08	10	10	09	0	03
10	139	160	<b>05</b>	04	106	10	08	03	0



Similarly, minimum distance between clusters '1' and '8' is '2'. These two clusters get merged into one cluster "(1, 8)". The minimum distance between cluster (1, 8) and other clusters in the distance matrix is calculated as shown in the below table 6.

**Table 6: Minimum distance between two clusters (1, 8)**

<b>B/N</b>	<b>1,8</b>	<b>2</b>	<b>3,9</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>10</b>
<b>1,8</b>	<b>0</b>	<b>10</b>	<b>03</b>	<b>8</b>	<b>10</b>	<b>10</b>	<b>9</b>	<b>03</b>
<b>2</b>	<b>10</b>	0	10	250	10	108	37	160
<b>3,9</b>	<b>03</b>	10	0	09	03	07	03	05
<b>4</b>	<b>08</b>	250	09	0	20	704	77	04
<b>5</b>	<b>10</b>	10	03	20	0	135	07	106
<b>6</b>	<b>10</b>	108	07	704	135	0	89	10
<b>7</b>	<b>09</b>	37	03	77	07	89	0	08
<b>10</b>	<b>03</b>	160	05	04	106	10	08	0

The minimum distance between cluster '(3, 9)' and '(1, 8)' is '3'. These two clusters get merged into one cluster "((3, 9), (1, 8))". The minimum distance between cluster "((3, 9), (1, 8))" and other clusters in the distance matrix is calculated as shown in the below table 7.

**Table 7: Minimum distance between two clusters (3, 9) and (1, 8)**

<b>B/N</b>	<b>(1, 8), (3,9)</b>	<b>2</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>10</b>
<b>(1, 8), (3,9)</b>	<b>0</b>	<b>10</b>	<b>09</b>	<b>03</b>	<b>07</b>	<b>03</b>	<b>03</b>
<b>2</b>	<b>10</b>	0	250	10	108	37	160
<b>4</b>	<b>09</b>	250	0	20	704	77	04
<b>5</b>	<b>03</b>	10	20	0	135	07	106
<b>6</b>	<b>07</b>	108	704	135	0	89	10
<b>7</b>	<b>03</b>	37	77	07	89	0	08
<b>10</b>	<b>03</b>	160	04	106	10	08	0

The minimum distance between cluster '((3, 9), (1, 8))' and '5' is '3'. These two clusters get merged into one cluster "((3, 9), (1, 8), 5)". The minimum distance between cluster "((3, 9), (1, 8), 5)" and other clusters in the distance matrix is calculated as shown in the below table 8.

The minimum distance between cluster '((3, 9), (1, 8), 5)' and '7' is '3'. These two clusters get merged into one cluster "((3, 9), (1, 8), 5, 7)". The minimum distance between cluster "((3, 9), (1, 8), 5, 7))" and other clusters in the distance matrix is calculated as show in the below table 9.

**Table 8: Minimum distance between two clusters (((3, 9), (1, 8)) and 5)**

<b>B/N</b>	<b>(((3, 9), (1, 8)) and 5)</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>7</b>	<b>10</b>
<b>(((3, 9), (1, 8)) and 5)</b>	<b>0</b>	<b>10</b>	<b>09</b>	<b>07</b>	<b>03</b>	<b>03</b>
2	<b>10</b>	0	250	108	37	160
4	<b>09</b>	250	0	704	77	04
6	<b>07</b>	108	704	0	89	10
7	<b>03</b>	37	77	89	0	08
10	<b>03</b>	160	04	10	08	0

**Table 9: Minimum distance between two clusters (((3, 9), (1, 8), 5) and 7)**

<b>B/N</b>	<b>(((3, 9), (1, 8), 5) and 7)</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>
<b>(((3, 9), (1, 8), 5) and 7)</b>	<b>0</b>	<b>10</b>	<b>09</b>	<b>07</b>	<b>03</b>
2	<b>10</b>	0	250	108	160
4	<b>09</b>	250	0	704	04
6	<b>07</b>	108	704	0	10
10	<b>03</b>	160	04	10	0

The minimum distance between cluster '((3, 9), (1, 8), 5, 7)' and '10' is '3'. These two clusters get merged into one cluster "((3, 9), (1, 8), 5, 7, 10)". The minimum distance between cluster "((3, 9), (1, 8), 5, 7, 10))" and other clusters in the distance matrix is calculated as shown in the below table 10.

**Table 10: Minimum distance between two clusters (((3, 9), (1, 8), 5, 7) and 10)**

<b>B/N</b>	<b>(((3, 9), (1, 8), 5, 7) and 10)</b>	<b>2</b>	<b>4</b>	<b>6</b>
<b>(((3, 9), (1, 8), 5, 7) and 10)</b>	<b>0</b>	<b>10</b>	<b>4</b>	<b>07</b>
2	<b>10</b>	0	250	108
4	<b>04</b>	250	0	704
6	<b>07</b>	108	704	0

The minimum distance between cluster '((3, 9), (1, 8), 5, 7, 10)' and '4' is 4. These two clusters get merged into one cluster "((3, 9), (1, 8), 5, 7, 10, 4)". The minimum distance between cluster "((3, 9), (1, 8), 5, 7, 10, 4))" and other clusters in the distance matrix is calculated as shown in the below table 11.

**Table 11: Minimum distance between two clusters (((3, 9), (1, 8), 5, 7, 10) and 4)**

B/N	(((3, 9), (1, 8), 5, 7, 10) and 4)	2	6
(((3, 9), (1, 8), 5, 7, 10) and 4)	0	10	07
2	10	0	108
6	07	108	0

The minimum distance between cluster '((3, 9), (1, 8), 5, 7, 10, 4)' and '6' is '7'. These two clusters get merged into one cluster "(((3, 9), (1, 8), 5, 7, 10, 4, 6))". The minimum distance between cluster "(((3, 9), (1, 8), 5, 7, 10, 4, 6))" and other clusters in the distance matrix is calculated as shown in the below table 12.

**Table 12: Minimum distance between two clusters (((3, 9), (1, 8), 5, 7, 10, 4) and 6)**

B/N	(((3, 9), (1, 8), 5, 7, 10, 4) and 6)	2
(((3, 9), (1, 8), 5, 7, 10, 4) and 6)	0	10
2	10	0

Finally, the minimum distance between cluster '(((3, 9), (1, 8), 5, 7, 10, 4, 6))' and '2' is '10'. These two clusters get merged into one cluster "((((3, 9), (1, 8), 5, 7, 10, 4, 6)), 2)", as shown in the below table 13.

**Table 13: Minimum distance between two clusters (((3, 9), (1, 8), 5, 7, 10, 4, 6) and 2)**

B/N	(((3, 9), (1, 8), 5, 7, 10, 4, 6) and 2)
(((3, 9), (1, 8), 5, 7, 10, 4, 6) and 2)	0

## RESULTS AND DISCUSSION

1. In the beginning we have clusters as: 1,2,3,4,5,6,7,8,9 and 10.
2. Clusters '3' and '9' are merged into cluster (3, 9) at distance "1".
3. Clusters '1' and '8' are merged into cluster (1, 8) at distance "2".
4. Clusters '(3, 9)' and '(1, 8)' are merged into cluster '((3, 9), (1, 8))' at distance "3".
5. Clusters '(((3, 9), (1, 8)))' and '5' are merged into cluster '((((3, 9), (1, 8)), 5)' at distance "3".
6. Clusters '(((3, 9), (1, 8), 5))' and '7' are merged into cluster '((((3, 9), (1, 8), 5), 7)' at distance "3".
7. Clusters '(((3, 9), (1, 8), 5, 7))' and '10' are merged into cluster '((((3, 9), (1, 8), 5, 7), 10)' at distance "3".
8. Clusters '(((3, 9), (1, 8), 5, 7, 10))' and '4' are merged into cluster '((((3, 9), (1, 8), 5, 7, 10), 4)' at distance "4".
9. Clusters '((((3, 9), (1, 8), 5, 7, 10, 4)))' and '6' are merged into cluster '((((((3, 9), (1, 8), 5, 7, 10, 4)), 6)' at distance "7".
10. Clusters '((((((3, 9), (1, 8), 5, 7, 10, 4, 6))))' and '2' are merged into cluster '((((((((3, 9), (1, 8), 5, 7, 10, 4, 6)), 2)' at distance.

The result showed that ten clusters are merged into one cluster. It means that ten garment brand names clusters Converse, Vans, Reebok, Adidas, New Balance, Nike, Puma, Brooks, Levis and Jordan are merged

into cluster ((((((Reebok, Levis), (Converse, Brooks)), New Balance), Puma), Jordan), Adidas), Nike), Vans). The main aim of cluster analysis here is to find maximum similarities between garment brand names. The dendrogram showed minimum distance at which similar entities (brand names) are grouped together to form a cluster. This similarity between clusters is represented by minimum distance through agglomerate hierarchical clustering.

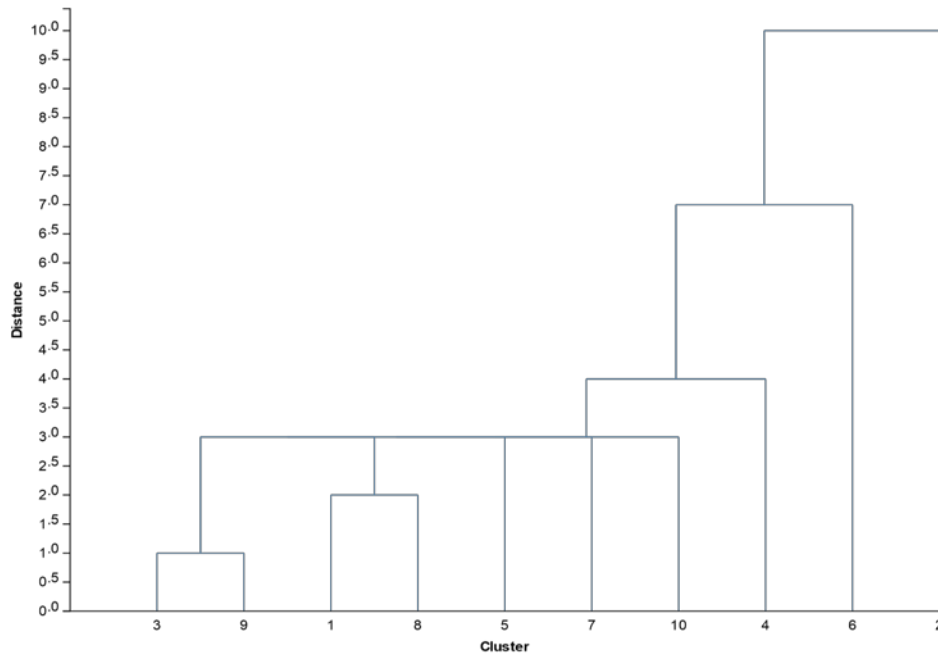


Figure 2: Dendrogram

While conducting experiments during the implementation work through MATLAB, the desired characteristics of an agglomerate clustering algorithm with K – mean clustering technique have been observed at different levels of clustering. These characteristics are dependent on a particular problem under consideration. The following is a list of characteristics:

1. The agglomerative hierarchical clustering form a cluster hierarchy that is commonly displayed as a tree hierarchy called a dendrogram. They begin with each object in a separate cluster. At each step, the two clusters that are most similar are merged into a single new cluster. While k-means algorithm gives us a simple or flat partition, because it just gives us a single set of clusters, with no particular organization or structure within them. In the K mean clustering technique, some closely clusters could be related to other clusters and more distantly related to others.
2. In the hierarchical clustering, historical data about the number of clusters is not required. While, K mean clustering technique needs historical data about the number of clusters.

As, in the case of garment brand names clustering analysis, there is no need of historical data about the number of clusters and similarities of garment brand name is calculated on the bases of minimum distance among closet pair of brand names. Hence, hierarchical clustering technique is more suitable for clustering analysis of garment brand names.

## CONCLUSION

As organizations give much time for measuring the effectiveness of their social media activities, it is very necessary for these organizations to attach with high level business objectives. Business through social media provides a suitable approach for firms to get the customer's idea about brands. In present days,

different organizations use social networking sites for their marketing. This practice is very handy for advertisers to establish brand's sales, etc. Brand name analysis is a very important part of the modern business world. The satisfaction of customers is the main issue in the marketing. In order to minimize this issue, it is important to collect the main interest of customer in special brand name. The proposed methodology offers a new way for finding similarities among brand names using social media data. The performance of method is verified using garment brand names from social media data. We believe that this work will assist advertisers to develop more care about customer's preference in online business world.

## REFERENCES

- Chung T, Anaza NA, Park J, Hall-Phillips A. (2016): Who's behind the screen? Segmenting social venture consumers through social media usage. *Journal of Retailing and Consumer Services* 28:288-295.
- Brown A, Kietzmann C. (2018). Strategic knowledge management and enterprise social media. *Journal of Knowledge Management* 22(6): 1288-1309.
- Chakraborty R, Kundu S, Agarwal P. (2016). Fashioning Data – A Social Media Perspective on Fast Fashion Brands. *Proceedings of NAACL-HLT 2016*, pp. 26 – 35.
- Carbonell P, Mayera M, Bravo A. (2015). Exploring Brand-Name Drug Mentions on Twitter for Pharmacovigilance. *A Research Program in Biomedical Informatics (GRIB)*, pp. 55 – 59.
- Chen Y, Tang K, Shen R, Hua Y. (2004). Market basket analysis in a multiple store environment, decision Support Systems. Department of Information Management, National Central University, Chung-Li, 320 Taiwan, ROC b Krannert Graduate School of Management, Purdue University, West Lafayette, IN 47907, USA. Pp. 339 – 354.
- Chitral K, Maheswari D. (2017). A Comparative Study of Various Clustering Algorithms in Data Mining. *International Journal of computer science and Mobile Computing*. ISSN 2320–088X IMPACT FACTOR: 6.017. *IJCSMC*, Vol. 6. pp.109 – 115
- Constantinides E, Lorenzo C, Alarcon –Del – Amo M. (2011). Consumer adoption of social networking sites: Implications for theory and practice. *Journal of Research in Interactive - Marketing* 5:170-188.
- Falls J. (2010). What is engagement and how do we measure it? *Social Media Explorer*.
- Gunelius S. (2011). 30 minute social media marketing: Step by step techniques to spread the words about your business. New York: McGraw-Hill. ISBN13: 9780071743815. (New York, USA).
- Goyal R, Srivastava K. (2016). A Study on the cluster analysis technique - hierarchical algorithm. *Imperial Journal of Interdisciplinary Research (IJIR)* 2(9): 1274 – 1279.
- Hyun K, Lee J, Kim M, cho S. (2014). Style analysis methodology: identifying the car brand design trends through hierarchical clustering. *Proceedings of the 19th International Conference on Computer Aided Architectural Design Research in Asia CAADRIA*. (Daejeon, Korea). Pp. 327–336.
- Jackson N. (2011). Infographic: Using social media to build brand loyalty.
- James B, Rowley J. (2017). Evaluation and decision-making in social media marketing. *Management Decision* 55(1):15-31.
- Kang HN, Yong HR, Hwang HS. (2016). Brand Clustering Based on Social Big Data: A Case Study. *International Journal of Software Engineering and Its Applications* 10(4): 27-36.

- Larson L, Watson R. (2011). The value of social media: toward measuring social media strategies. Thirty Second International Conference on Information Systems, Shanghai 2011.
- Li C, Bernoff J. (2008). Groundswell. Boston: Harvard Business Review Press.
- Leis A. (2013). How Twitter Is Used in International Health Events: World Aids Day Case Study. In Medicine 2.0 Conference. JMIR Publications Inc., Toronto, Canada.
- Murtagh F, Contrears P. (2011). Methods of Hierarchical Clustering. Pp 1-21.
- Mohbey K, Thakur GS. (2013). An Experimental Survey on Single Linkage Clustering. International Journal of Computer Applications 76(17):6-10.
- Hutton G, Fosdick M. (2011). The Globalization of Social Media. Consumer Relationships with Brands Evolve in the Digital Space. J. Advert.Res. 51(4): 564-574.
- Paquette H. (2013). Social Media as a Marketing Tool: A Literature Review. Major Papers by Master of Science Students. Pp 1-26.
- Poecze F, Ebster C, Strauss C. (2018). Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. 9th International Conference on Ambient Systems, Networks and Technologies, ANT-2018. Pp 660 – 666.
- Rutter R, Roperb S, Letticec F. (2016). Social media interaction, the university brand and recruitment performance. Journal of Business Research 69(8):3096-3104.
- Lehmann R, Lane DK, Farley JU. (2008). The structure of survey-based brand metrics. Journal of International Marketing 16 (4):29 – 56.
- Rezankova H. (2014). Cluster Analysis of Economic Datas. Statistika: Statistics and Economy Journal 94(1):73-86.
- Singh M, Peszynski K. (2008). Organisational Value of Social Technologies: An Australian Study. Electronic Journal Information Systems Evaluation 17 (1):88 – 99.
- Soni N, Ganta A. (2012). Categorization of Several Clustering Algorithms from Different Perspective: A Review. International Journal of Advanced Research in Computer Science and Software Engineering 2(8): 63-68.
- Sing S, Sao A, Nagare TB, Dharmarajan A. (2016). Role of Social Media Marketing In Brand Building: The New Age Marketing strategy. International journal of scientific research 5(9): 458- 463.
- Shelestov A. 2008. Using the agglomerative method of hierarchical clustering as a data mining tool in capital market. International Journal Information Theories & Applications 15:282 – 286.
- Salloum S A, Al – Emran M., Monem AA, Shaalan K. (2018). Using Text Mining Techniques for Extracting Information from Research Articles. In Intelligent Natural Language Processing: Trends and Applications, Springer, Pp 373 – 397.
- Tuten TL, Solomon MR. (2013). Social Media Marketing. Sage. 2<sup>nd</sup> Edition. Pp 1-352.
- De Vries L, Gensler S, Leeflang PS. (2012). Popularity of Brand Posts of Brand Fan Pages: An Investigation of the Effects of Social Media Marketing 26(2): 83 - 91.
- Verma M, Srivastava M, Chack N, Kumar A, Gupta N. (2012). A Comparative Study of Various Clustering Algorithms in Data Mining. International Journal of Engineering Research and Applications (IJERA) 2(3):1379- 1384.
- Van - Droon J, Lemon K, Mittal V. (2012). Customer Engagement Behavior: Theoretical Foundations and Research. Journal of Service Research 13(3):253 - 266.