IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SOCIAL NETWORK USING OPTIMIZED CLUSTERING ALGORITHM

Akshay Gupta
Department of Computer Science and Engineering
Abdul Kalam Technical University
Lucknow, India

Mr.Rahul Chaturvedi
Asst. Professor Department of Computer Science and Engineering
Abdul Kalam Technical University
Lucknow, India

Kelvin
Department of Computer Science and Engineering
Abdul Kalam Technical University
Lucknow, India

Chirag Sanwal
Department of Computer Science and Engineering
Abdul Kalam Technical University
Lucknow, India

Atul Tiwari
Department of Computer Science and Engineering
Abdul Kalam Technical University
Lucknow, India line

***Abstract*—In the current pandemic erra there has been a boom in the rize of people using social media platforms to be able to connect and interact with the world virtually, making the social media platforms a go to place for consuming day to day information about various topics related to an individual's preference.This information expands vast and diverse in its domain making most of it almost noisy data, irrelevant to an individual preference.**

Practical implementation approach to clustering in social media websites to group similar groups of tags that can be used to optimize the preference search and feeds of the users and optimize the content delivery of our platform using k-means clustering algorithm and existing popular technology stack known as MERN stack.

Keywords—MERN stack, k-means, clustering, Machine Learning, Data Mining.

I. INTRODUCTION

The main objective of this paper is to explore the methods and technologies existing in current era that can be used to implement the clustering on diverse heterogeneous ideas and topic of interest which can be used for creating a platform for a diverse ideas and topic of interests which can be used for creating a platform for a diverse audience that can connect on the common base ideologies.

In this paper we describe the use of MERN stack to implement the K-means unsupervised learning algorithm essential for a rapidly growing diverse collection of topics.

The unsupervised iterative k-means algorithm generates distant clusters in the database of diverse topics, generating an output of percentage of liking of a cluster by a group or individuals.

The k-means algorithm succeeds in doing this by mapping each observation value in the input dataset to a point in the n-dimensional space.

In k-means clustering, a center point value is contained in each cluster. At the time of model training, the k-means algorithm uses the distance of the point that relates to each observation value in the dataset to the cluster centers as the basis for clustering. The number of clusters to be created are predefined by the observer.

This clustering approach in our paper is structured upon the most popular programming language of the 21st century after the python language known as javascript and the technologies environment based primarily on the javascript language.

II. LITERATURE Survey

In the past few years, a number of clustering algorithms for big data have been proposed which are derived on the basis of distributed and parallel computation. In 1967 Mac Queen was the first to propose this technique. The first standard algorithm was proposed by Stuart Lloyd in 1957 as an approach for pulse-code modulation. Oftentimes it is referred to as Lloyd-Forgy because in 1965, E.W.Forgy published essentially the same method. According to K. A. Abdul Nazeer all the major drawback of the k-means algorithm is about choosing the initial centroids which produce different clusters. But eventually the final cluster quality of algorithms depends on the selection of initial centroids, chosen at the time of computation. Two phases included in the original k-means algorithm: first for determining initial centroids and second for assigning data points to the nearest clusters and then recalculating the clustering mean. But this enhanced clustering method uses both the phases of the original k-means algorithm. This algorithm combines a systematic method for finding initial centroids and an efficient way for assigning data points to clusters. But still there is a limitation in this enhanced algorithm that is the value of k, the number of desired clusters, is still required to be given as an input, regardless of the distribution of the data points.

According to Y. S. Thakare et al., the performance of k-means algorithm which is evaluated with various databases such as Iris, Wine, Vowel, Ionosphere and Crude oil data Set and various distance metrics. It is concluded that performance of k-means clustering is dependent on the data base used as well as distance metrics. Soumi Ghosh et al. proposed a comparative discussion of two clustering algorithms namely centroid based K-Means and representative object based Fuzzy C-Means clustering algorithms. This discussion is on the basis of performance evaluation of the efficiency of clustering output by applying these algorithms. The result of this comparative study is that FCM produces a closer result to the K-means but still computation time is more than k-means due to involvement of the fuzzy measure calculations. Sakthi et al. proposed that due to the increment in the amount of data across the world, analysis of the data turns out to be a very difficult task. To understand and learn the data, classify those data into remarkable collections. So, there is a need for data mining techniques. Amutha et al. proposed that when two or more algorithms of the same category of clustering technique is used then best results will be acquired. Two k-means algorithms: Parallel k/h-Means Clustering for Large Data Sets and A Novel K-Means Based Clustering Algorithm or High Dimensional Data Sets. Parallel k/h-Means algorithm is designed to deal with very large data sets. Novel K-Means Based Clustering provides the advantages of using both HCand K-Means. Using these two algorithms, space and similarity between the data sets present each node is extended. Nidhi Singh et al. proposed the comparative analysis of one

partition clustering algorithm (k means) and one hierarchical clustering algorithm (agglomerative).On the basis of accuracy and running time the performance of k-means and hierarchical clustering algorithmics calculated using WEKA tools. This work results that accuracy of k-means is higher than the hierarchical clustering for iris dataset which have real attributes and accuracy of hierarchical clustering is higher k-means for diabetes dataset which have integer, real attributes. So for large datasets k means algorithm is good. Shi Na et al.

Proposed the analysis of shortcomings of the standard k-means algorithm. The k-means algorithm has to calculate the distance between each data object and all cluster centers in each iteration. This repetitive process affects the efficiency of clustering algorithms. Cui, Xiaoli, et al. proposed optimized big data K-Means using Map-Reduce in which they claimed to counter the iteration dependence of Map-Reduce jobs. They used a sequence of three Map-Reduce (MR) jobs. However, in their approach sampling technique is used in the first M-R job.

- In data mining, two learning methods used to mine data i.e. supervised learning and unsupervised learning. Supervised learning: In this learning, data includes together the input and the desired result. It is a fast and perfect learning method. The accurate results are known and are given in inputs to the model during the learning procedure. Neural network, Multilayer perceptron, Decision tree are supervised models. Unsupervised learning: The desired result is not provided to the unsupervised model during the learning procedure. This method can be used to cluster the input data in classes on the basis of their statistical properties only. These models are for various types of clustering, k-means, distances and normalization, self- organizing maps.
- K-Means clustering

  The K-means clustering algorithm is a famous clustering technique. It is used in many areas such as information retrieval, computer vision and pattern recognition. K-means clustering assigns n data points into k clusters so that similar data points can be grouped together. It is an iterative method which assigns each point to the cluster whose centroid is the nearest. Then it again calculates the centroid of these groups by taking its average. Properties of k-means algorithm:

  - Large data sets are efficiently processed.
  - It often terminates at a local optimum.
  - It supports numeric values.

A. Objectives

Data mining is a process of deriving required data from a collection of large datasets and making analysis on collected data. Clustering is a technique for grouping of similar datasets in which data within clusters have similar properties. K-Means is a widely used clustering algorithm whose uniform effect is producing clusters with relatively uniform size even if the input data have different cluster size is the main advantage.

The implementation of k-means in this paper was explored through MERN stack which is a javascript dependent development stack capable of designing fullstack web apps.

JavaScript is a versatile scripting language that requires no compiler to run its code base while simultaneously fully capable of creating programing logic that are feasible for practical implementations

The JavaScript codes are reusable, they have good tutorials, they are relatively easy to learn and implement, numerous resources can be found in Stack Overflow and GitHub projects, they offer faster development, great distribution through npm.

B. Scope

The thing is real life data is almost always complex, disorganized and noisy. Situations in the real world rarely reflect clear conditions in which to apply these types of algorithms right out of the shelf. In the case of the K-means algorithm it will be expected that at least one of its assumptions gets violated, so we need not only to identify this, but to know what to do in such a case.

C. Limitations

Our Projects Input data required for computation is dependent on the people's survey opinion.

D. Delimitations

This algorithm is very interesting and easy to implement given that we have an affordably small network with not too complex clustering.

E. Problem statement

Now-a-days social media is used to introduce new issues and discussion on social media. More users participate in discussion via social media. Different users belong to different kinds of groups. Positive and negative comments will be posted by the user and they will participate in discussion . Here we proposed a system to group different kinds of users and the system specifies from which category they belong to. For example the film industry, politicians etc. Once the social media data such as user messages are parsed and network relationships are identified, data mining techniques can be applied to groups of different types of communities. We used the K-Means clustering algorithm to cluster data. In this system we detect communities by clustering

messages from large streams of social data. Our proposed algorithm gives better clustering results and provides a novel use-case of grouping user communities based on their activities.

F. Project Design

A data-mining analysis about opinions of people on social media sites based on k-mean algorithms condensed to be used in a practical implementation in designing a basic structure for a social media network on a web app platform using current existing technology.

IV Clustering

A. Overview

Clustering is the method of partitioning the population or data points into a number of groups such that those data points in the same groups are more similar to other data points in the same group than those in the other groups. In simple words, the aim is to segregate groups with specifically similar traits and assign them into clusters.

Let's understand clustering with an example. Lets say,, you are the head of a rental store and want to understand the preferences of your customers to scale up and scale out your business. Is it possible and feasible for you to look at the details of each customer and devise a unique business strategy for each and every one of them? Definitely not. But, what can be done is to cluster all of your customers into lets say 10 groups based on their purchasing habits and spending habits and use a separate strategy for customers in each of these 10 groups. And this is called a method of clustering.

B. Types of Clustering

Statistically speaking, the methods of clustering are divided into two subcategories

Hard Clustering: In hard clustering, every data point either belongs to a cluster completely or not. For example, in the above given example every customer is put into one group out of the 10 groups.

Soft Clustering: In soft clustering, in place of putting each data point in a separate cluster, a probability or likelihood of that data point to be in these clusters is assigned. For example, from the above scenario each customer is assigned a probability to be in either of 10 clusters of the retail store.

C. Types of clustering models

Since the result of clustering is subjective, this means that there are multiple possible methods that can be used for achieving this goal. Every methodology follows a different set of rules for defining the '*similarity'* between data points.

Connectivity Models

As the name suggests, these models are based on the idea that the data points closer in data space show more similarity to each other than the data points situated farther away. These models follow two following approaches. In the first approach, they start with classifying all data points into separate clusters & then aggregating them as the distance between these data points decreases. In the second approach, all data points are classified as a single cluster and then partitioned as the distance between the data points increases. Also, the choice of distance function is totally subjective. These models are very easy to interpret but lack the scalability for handling huge datasets. Examples of these models are hierarchical clustering algorithms and its variants.

Centroid Models

These are iterative clustering algorithms in which the idea of similarity is derived by the closeness of a data point to the centroid of the clusters. K-Means clustering algorithm is a popular algorithm that falls into this category. In these models, the number of clusters required at the end have to be mentioned beforehand, which makes it important to have prior knowledge of the dataset. These models run iteratively to locate the local optima.

Distribution Models

These clustering models are based on the idea of how probable it is that all the data points in the cluster belong to the same distribution (For example: Normal or Gaussian). These models often suffer from overfitting. A popular example of these models is the Expectation-maximization(EM) algorithm which makes use of multivariate normal distributions.

Density Models

These models search the data space for areas of different density of data points in the given data space. It isolates various different density regions and assigns the data points within these regions in the same cluster. Popular examples of density models are DBSCAN and OPTICS.

V. K-means clustering Algorithm

K-means is a partitioning algorithm that uses iterative steps for breaking data sets into groups. The contents of these groups may or may not be related to that particular group so to correct this error the process of clustering is done in x number of steps to reach the optimal solution where the majority of the identical solution is considered as optimal and the non-identical solution is considered as the error in the cluster generating process.

K-means is also used for minimizing the total squared errors in the data sets.

K-means can be used for the cases where there is no prior knowledge of classes to cluster the data together.

Used for suggesting groups based on patterns in a data set.

The formed groups can be distinguished on a varying scale of degree calculated by distances between these classified or grouped data using distance calculating algorithms such as Hamming distance, Euclidean distance etc.

K-Means is an algorithm that takes in a dataset and a constant k and returns k centroids (which define clusters of data in the dataset which are similar to one another).

For analysis a k-means is capable of visual representation of the output results in n numbers of dimensions where each centroid is the geometric mean of the points that have that centroid's label. Important: If a centroid is empty (no points have that centroid's label) you should randomly re-initialize it.

K-Means is much faster if you write the update functions using operations on arrays, instead of manually looping over the arrays and updating the values yourself in terms of programmatically implementing the algorithm.

K-means can be paired with many other algorithms to increase the robustness of the output result like adding the probabilistic approach to the output result. It can also be paired with the estimation and maximization algorithm to obtain the average output values in terms of its best and worst performing results.

III. METHODOLOGY

Our technology stack is encircled around the javascript language, in the early days the javascript used to be a scripting language designed to work using a web browser created by Brendan Eich in 1995 until Ryan Dahl in 2009 who wrote node.js which is a JavaScript runtime environment in c and c++ language that runs on v8 google javascript engine and is capable of executing the javascript code outside of a web browser.

The technology stack and techniques used are as given

below.

A. MERN

MERN stack is a JavaScript stack that is used for easier and quicker deployment of full-stack web based applications. It simplifies the work flow as MERN stack implies focus on using a single programming language (i.e javascript) for developing API's, algorithms and data packets for storage. A MERN stack brings a complete suite of technology that is more than capable enough to build and deploy any web-app with completely responsive frontend and backend with asynchronous input and output operations on stored data. MERN also has a very important functionality that it supports JSON format of data storage natively. MERN stack brings a suite of functionality which are well suited for creating cloud native apps that are JSON heavy and have dynamic web interfaces. Since, MERN stack uses a single programming logic it eliminates the need of designing engines required for data interpretations between front end and back end of our app. The MERN architecture allows you to easily construct a 3-tier architecture (frontend, backend, database) entirely using JavaScript and JSON.

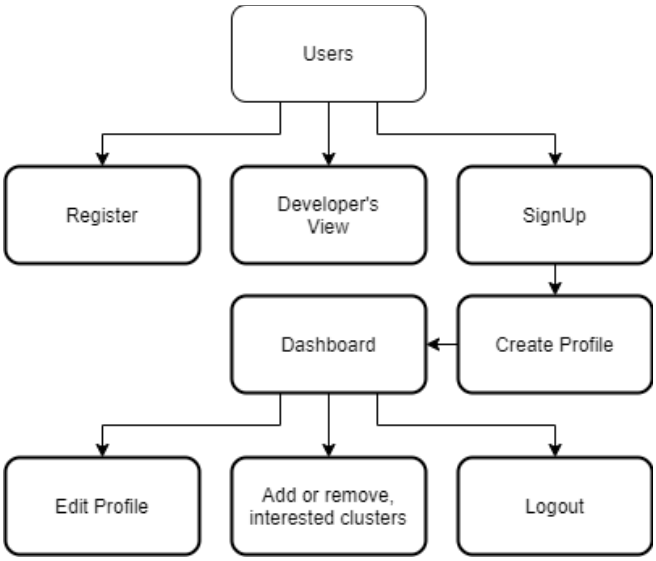
MERN stack is simply a group of four technologies expanding along with their individual functionalities:

- MongoDB,
- Express,
- React and
- Node.js.

MERN stack can easily component the backend and frontend into sub components independently contributing to the entire code base of a project ensuring the ease of debugging modification, and updations of the code base in the individual components.

MERN is an open source technology with availability of solutions for issues in developing a project through community support guides and ample of open source contributors willing to help even with any previously unknown issues in design and code base of a project.

It is designed to make the development process smoother and better.



B. K-Means algorithm

K-Means Algorithm is a clustering algorithm to partition the number of observations into clusters in which each observation belongs to the cluster with the nearest mean. K-means takes two variables as inputs:

The first variable is the observations that we want to cluster.

The second variable is the size of the cluster.

C. K-Means Implementation in Node.js Using node-k means

We will build an API REST server using Node.js and create a function to cluster our observations using K-Means Algorithm. We can use a library named node-k means. In our case, the target user that we had to arrange contained their interest information. We would consider this information as the first variable. Second target user who has the same interest as the first would be our second variable.

K-Means will cluster the users into three groups. Each group may not contain the same number of users, but we can be sure that the interests of the users in the same group are close to each other.

D. Node.js API server

This API server will contain functions to implement K-Means Algorithm and routes for connections with the ability to asynchronously perform input and output operations contributing to faster run time of the entire system with all the computations running even on exponentially increasing data sets required for our clustering algorithm.

Node is the best and the fastest JavaScript runtime environment engine built on chrome v8 engine which is written in c++ , enabling the javascript run outside the browser. Node Js also brings the advantage of a massive open source community with no cost of using it while simultaneously supporting most of the common operating system platforms like Windows, Linux, Unix, Mac OS X, etc.

Node Js establishes the operations of create, delete, write, read, open, close files on the web servers using api and route calls.

Node Js is also capable of executing functions that can trigger database operation through programing logic to delete, add, and modify data in a database. Node Js also removes the waiting for file operations to be executed first before initiating next operations by simply moving with the next request. Hence establishing the ability of asynchronously performing file operations using single thread non blocking processes which brings an efficient use of memory and computation resources.

Node uses npm as its package manager for the JavaScript programming language. npm, Inc. is a subsidiary of GitHub, that provides hosting for software development and version control with the usage of Git. npm is the default package manager for the JavaScript runtime environment Node.js.

Node Js also supports most of the existing and current latest generation databases and can easily be paired to create and manage an efficient backend. Node Js operations are broken into tasks that are executed when a certain event is triggered.

E. React web application

This will be served as a user interface where one user will greet with some another user of the same interests.It was used to develop a user-friendly interface of the application. Also, to reduce the stress of doing CSS styling from scratch and to re-use all the components within the project React was used.

Notable feature of react is the use of a virtual Document Object Model, or virtual DOM. React creates an in-memory data-structure cache, computes the resulting differences, and then updates the browser's displayed DOM efficiently. This process is called reconciliation. This allows the programmer to write code as if the entire page is rendered on each change, while the React libraries only render subcomponents that actually change. This selective rendering provides a major performance boost. It saves the effort of recalculating the CSS style, layout for the page and rendering for the entire page

React code is deployed on the server side to be loaded as front end through node js by rendering the code as client side event which pushes the front end code to be displayed on client side web browser.

A react web app initializes the code base as frames with components inside it that are reloaded instead of reloading of the entire front end code making it displaying dynamically loading content, this content state can be managed using redux or angular which are javascript libraries that can efficiently display content as user interface to client side.

F. Mongoose

The use of mongoose enables server and database communicate with each other. Mongoose is installed on the node js server to enable communication with our database server i.e mongo db.

Mongoose provides a straight-forward, schema-based solution to model your application data. It includes built-in type casting, validation, query building, business logic hooks and more, out of the box.

Mongoose is a MongoDB object modeling tool, or ODM (Object Document Mapper), written in JavaScript and designed to work in an asynchronous environment.

G. MongoDB

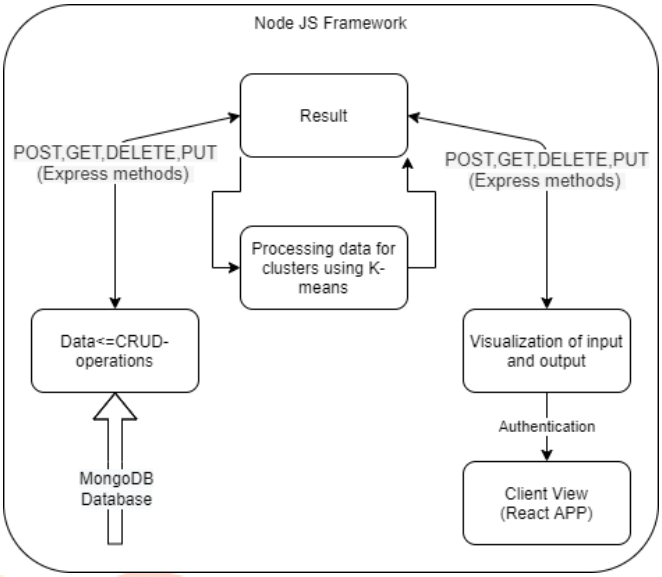
This is the database of our app to hold all the raw and processed data. An Account was created in MongoDB website. A free MongoDB cluster was set up after creating a new project. Apart from this, read and write access to the database was created which was used in the NodeJS application to read and write the data from and to the database. Also, the current IP address of the machine was added to the IP whitelist so that the app can communicate with the server. Finally, the MongoDB driver was installed in the application using the command "npm install mongodb''.

The mongo db supports the json format useful in structured storage of data easing the retrieval of useful information from the database for further processing.

A record in MongoDB is a document, which is a data structure composed of field and value pairs. MongoDB documents are similar to JSON objects. The values of fields may include other documents, arrays, and arrays of documents.

H. Nodemon

Nodemon concurrently is a tool that helps to run more than one command inside the terminal. This tool is handy when running the front end and back end of the project simultaneously.



I. JSON

JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language.

JSON is built on two structures:

- A collection of name/value pairs. In various languages, this is realized as an object, record, struct, dictionary, hash table, keyed list, or associative array.
- An ordered list of values. In most languages, this is realized as an array, vector, list, or sequence.

J. Naming Conventions in JS

Since we describe the use of Javascript as our primary programming language, this language finds it difficult to interpret the variable names with white spaces, so camel casing helps in making it efficient in identifying and using the variable names without dealing with white space issues in Javascript.

K. Working of K-mean clustering

K-mean is the simplest unsupervised learning algorithm that solves the well-known clustering problem. It's a method of vector quantization, originally from signal processing, that is popular for clustering analysis in data mining. The procedure follows a simple way to classify a given data set through a certain number of clusters fixed initially. The distance among the formed clusters is calculated using methods such as euclidean distance which supports the idea of distinguishing the formed groups to distinguish the apart on a mathematical scale of values. These mathematical values can be assigned to any type of data on initial stages and can be converted back to its original data type mapping to the

actual meaning of its original. The algorithm aims at minimizing an objective function known as squared function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (| \ |x_i - y_j| \ |)^2$$

Where,

||xi – vj || = Euclidian distance between xi and vj

Ci = number of data points in ith cluster

C = number of cluster centers

Step 1: Randomly select 'c' cluster centers.

Step 2: Calculate the distance between each data point and cluster centers.

Step 3: Assign the data point to the cluster center whose distance from the cluster center is the minimum of all the cluster centers.

Step 4: Recalculate the new cluster center using:

$$v_i = \frac{1}{c_i} \sum_{j=1}^{c_i} x_i$$

Where, ci represents the number of data points in ith cluster.

Step 5: Recalculate the distance between each data point and new obtained cluster centers.

Step 6: If no data point was reassigned then stop, otherwise repeat from step 3.

Here euclidean space is the length of a line segment between the two points. It can be calculated from the Cartesian coordinates of the points using the Pythagorean theorem, therefore occasionally being called the Pythagorean distance.

IV. CONCLUSION

The MERN stack is a reasonable choice to work with the k-means algorithm to develop a social media platform.

The clustering strategy for exponentially increasing data makes an unsupervised clustering algorithm a viable choice where k-means clustering algorithm comes into picture.

The outcome of this paper gives us the structural approach to build a social media platform that is capable of grouping similar ideas in a pile of data and enable social interaction on the platform.

The iterative algorithm used in MERN stack generates poor runtime to overcome this issue; the correct initial conditions should be supervised as the iterative techniques are very sensitive to the initial starting conditions.

Overlapping groups should be identified in the database in earlier stages so as an effective k-means clustering algorithm to generate minimal, discrete and diverse number of centroids as far as possible from one another i.e maximum Euclidean distance.

V. REFERENCES

[1] Refining Initial Points for K-Means Clustering
Paul S. Bradley Usama M. Fayyad, Appears in Proceedings of the 15th International Conference on Machine Learning (ICML98), J. Shavlik (ed.), pp. 91-99. Morgan Kaufmann, San Francisco, 1998.

[2] *A Comparative Study of K-means and K-medoid Clustering for Social Media Text Mining Volume 2 ||Issue 11 ||JUNE 2017||ISSN.*

[3] Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M., & Vrgoč, D. (2016). Foundations of JSON Schema. Proceedings of the 25th International Conference on World Wide Web - WWW '16. doi:10.1145/2872427.2883029

[4] Analysis and Visualization of Twitter Data using k-means Clustering, International Conference on Intelligent Computing and Control Systems ICICCS 2017.

[5] Kalra, M., Lal, N., & Qamar, S. (2017). K-Mean Clustering Algorithm Approach for Data Mining of Heterogeneous Data. Lecture Notes in Networks and Systems, 61–70. doi:10.1007/978-981-10-3920-1_7

[6] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: analysis and implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), 881–892. doi:10.1109/tpami.2002.1017616

[7] Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 219(1), 103–119. doi:10.1243/095440605x8298

[8] Wang, X., Tang, L., Gao, H., & Liu, H. (2010). Discovering Overlapping Groups in Social Media. 2010 IEEE International Conference on Data Mining. doi:10.1109/icdm.2010.48

[9] Villamizar, M., Garces, O., Ochoa, L., Castro, H., Salamanca, L., Verano, M., … Lang, M. (2016). Infrastructure Cost Comparison of Running Web Applications in the Cloud Using AWS Lambda and Monolithic and Microservice Architectures. 2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid). doi:10.1109/ccgrid.2016.37

[10] Zhao, G., Huang, W., Liang, S., & Tang, Y. (2013). Modeling MongoDB with Relational Model. 2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies. doi:10.1109/eidwt.2013.25

[11] Gyorodi, C., Gyorodi, R., Pecherle, G., & Olah, A. (2015). A comparative study: MongoDB vs. MySQL. 2015 13th International Conference on Engineering of Modern Electric Systems (EMES). doi:10.1109/emes.2015.7158433

[12] https://nodejs.org/docs/latest-v15.x/api/

[13] https://reactjs.org/docs/getting-started.html

[14] https://docs.mongodb.com/manual/

[15] Binkley, D., Davis, M., Lawrie, D., & Morrell, C. (2009). To camelcase or under_score. 2009 IEEE 17th International Conference on Program Comprehension. doi:10.1109/icpc.2009.5090039