Clump Approach for Animal Concern System

Aruna Pavate [1], Pranav Nerurkar[2] Anita Chaudhary[3],
Rajesh Bansode[4] Mansi Shah[5]

[1,3]*(Research Scholar,IT Thakur college of Engineering, Mumbai, India)*
[2]*(Research Scholar,IT Veermata Jijabai Technological* Institute, Mumbai, India)
[4] *(IT department, Thakur college of Engineering ,Mumbai, India)*
[5] *(CMPN department, Rizvi college of Engineering,   Mumbai, India)*

Abstract: *This Machine learning approach have proved best results in many applications from healthcare, Videos Surveillance, Social Media Services, Email Spam and Malware Filtering, Search Engine Result Refining, product recommendations, online fraud detection etc. Machine learning can provide an important part of the solution for cutting the cost and improving the quality with fewer efforts. This work implements machine learning technique to create a network of users who can assist an organization and or an individual to spread awareness by helping them to reach proper animal care center. Here adopted unsupervised learning for clustering animal care centers. The developed model is claimed to make predictions with relatively high accuracy. Furthermore, the result of the clustering-guided approach of prediction is discussed in general.*
Keywords: *Machine learning, unsupervised learning, k-means algorithm, animal care system, clustering*

I.   Introduction

This Clustering is a type of unsupervised technique which helps to group the data samples based on similarities of the data. The result of clustering gives the groups of samples clumped together based on the similarities with labels [1]. Clustering methods developed on the data without labels i.e. unknown data class. Therefore, clustering is often classified as an unsupervised learning method [2]. Now days, good and legal vet hospitals are difficult to find. The designed system helps the individuals to find the legal vet hospitals/animal centers. Animal care system allows individuals to create a network of users who can assist an organization /an individual, for spreading the awareness by helping the strays reach a proper care center, which is difficult in current scenario. Direct contact to the vets is not possible, a user or an organization can connect too many different animal lovers through this system. It's not only changes the environment but does also helps the various systems to enact upon it and have a complete healthcare system for animal in one roof. It is necessary to get information of the all useful vet hospitals so animals can be admitted. Animal care system helps care center to manage the information about types of animals wanted to adopt by individuals. Also, a system informs the availability of animals in the specific center and allows the persons to get through the center. The randomness of the Animal Center managed and controlled, to put a limit on the number of iterations to be carried out in the conventional algorithm.

a.   Significance of the study

There are very less Animal care centers available in India. Many pet animals lost. It is necessary to ensure the return of missing animals and also identification of animals in the form of licenses. It is difficult to search missing pets, if the full time service centers available which helps in searching pet. Also an animal care center helps the officers to assist animals in the field. An individual helps in the placement of shelter of animals and also give fundraising support and valuable administrative. Animal care centers helps animals to rescue in need, create a threat to public safety, injured animals ,stray dogs that need extra help and also can assist to answer to phone calls from the public have queries regarding animals. It is requisite to deal with the issues such animal testing, Animals in zoos and related facilities, abandoned pets, hunting, poaching, Cruelty to animals etc.

Cluster Analysis is one of the most important data mining techniques which help the researchers to analyze the data and categorize the attribute into various groups [3]. K Means is one the frequent partitioning algorithm used in clustering. The improvement in the K- means algorithm clustering can be done by choosing appropriate initial cluster centers to converge quickly to the local optimum [2]. The paper suggests that initialization of the cluster centers cannot be separated from effectiveness and the concept of success and failure. The randomness of the cluster centers needs to be managed and controlled to put a limit on the number of iterations to be carried out in the conventional algorithm with decreased complexity and increased accuracy.

b. Applications

- Individuals can operate this application from any place to adopt a pet by contacting vet care centre.
- When an individual comes across an injured stray, he/she can use this application to call an ambulance or take them to a nearby vet hospital/animal centre.
- Animal care center as well as NGOs can spread the awareness about the strays with the help of this application.
- In case of misplaced pet an individual post the details of the pet on the portal for assistance as well as can contact to the doctor in illness or in case of injury.

The study initially started with the aim of grouping vet care centers from a data-driven standpoint [5]. In this concern, unsupervised clustering using the K-Means algorithm was used. The learned clusters were then utilized to train the forecast models as presented at the following sections in detail.

II. Literature Review

There are many applications where K-means algorithm applied for clustering data. A few of them are mentioned below.

Mahmoud Elbattah, Owen Molloy implemented unsupervised and supervised approach for clustering patients individually in Ireland. This study describes the machine learning problem in healthcare. K means algorithm suggested three coherent clusters of patients. The study adopted random forest approach for classification and regression. Author claimed that approach applied provide more accurate predictions compared to non-clustering-based predictions[6].

Parasian D.P Silitonga, used K-means algorithm to cluster the patients disease at Haji Adam Malik Hospital in Medan. K means clustering algorithm applied using Weka tool. Results showed that the pattern of disease tendency of patients like Mild Septicemia Disease with the highest tendency pattern. The results showed that in future hospitals can predict the priority of service based on the example of trends received [7].

Pranav Nerurkar, Archana Shirke, Madhav Chandane, Sunil Bhirud , have studied different clustering approaches in the context of massive data set with their strength and weaknesses. Results showed that current clustering algorithms needed parameter tuning as well as sensitive to noise [8].

Shenghui Wang, Rob Koopman, used unsupervised learning such as K-Means and the Louvain community detection algorithm to cluster the document which helps in topic identification. This work concentrated on semantic representation of an article from the entities associated with it [9].

Marco Capo, Aritz Pérez, Jose Lozano proposed recursive and parallel algorithm to cluster massive data. Author showed that K means algorithm works well on both the problems that are dimensionality and number of instances without affecting the quality of approximation & presented a new Boundary weighted K-means algorithm[11].

There are variety of clustering algorithms used among that, the k means algorithm remains one of the most popular algorithms[10][12][13][14][15].

Oyelade, O. J , Oladipupo, O. O , Obagbuwa, I. C designed system for analyzing students results using clustering technique. In this work, k means algorithm combined with deterministic model to analyses the student result. Author claimed that this model is improved over the previous model such as Fuzzy logic and reduced the limitations. Clustering algorithms serves as best algorithms and enhances the decision-making power [16].

Author's discussed the shortcomings of K-means algorithm such as efficiency of the clustering algorithm, calculation of distance between the object. In this work they proposed an improved k means algorithm to eliminate the shortcomings. Modified algorithm improved speed of clustering and reduced the complexity of the standard K means algorithm [17].

Huda Hamdan Ali , Lubna Emad Kadhum, discussed K means algorithms used in various applications from economic science to Image processing. This work focused on the image processing and data mining applications such as efficient customer segmentation, Pattern Recognition, to detect an infected part of fruit, Leukemia Image Segmentation and discussed the advantages and disadvantages of the algorithm[18].

Author proposed new algorithm based on standard K-means algorithm for face extraction. Improved algorithm helps reduction in intra cluster distance, number of iterations, processing times, and an increase in the accuracy rate [19].

Unnati R. Raval, Chaita Jani, suggested new algorithm based on the traditional K means algorithm. Author claimed that the time complexity of the proposed algorithm is much faster than traditional K-means. There are many advantages of the K means algorithm ad compared to disadvantages of, but it still needs some improvements. Also showed that the value of K affects the clustering decision [20].

Author addressed the problem of the traditional k means algorithm. In this work the classical k means algorithm modified by keeping the advantages of the original and concentrated on improving centroids. The updated algorithm computes new centroids by finding the mean of all data points within the same cluster. The

updated and classical K means algorithm applied on ten different datasets and comparative analysis done based on the measurement metrics recall, F1-score, accuracy, precision. Also it is proved that classical k –means failed to support outliers[21].

Wang proposed updated k-means algorithm which is feasible and effective for grid simulation analysis. The working of the proposed algorithm showed on the simulator. The results showed that there is difference between the theoretical results and the actual results. The proposed generative model detects the anomalous data sequences [22].

This work clusters the text with high similarity. The initial cluster selected randomly, and the two samples chosen based on the largest distance from the initial cluster. The work showed that the selection of the initial cluster center affects the time and precision of the clustering. Author proposed the vector space model to cluster the similar text. The result showed that the total time utilized for clustering has been reduced as well as the accuracy of the system has been improved [23].

Dhyaa Shaheed, worked on clustering of Arabic words by using k- means algorithm. Clustering of words done based on the characteristic like syntactic, morphological, and Semantic. The performance of the system

III. Approach Overview

In Machine learning, learning can be possible with three ways: 1. Supervised Learning, 2. Unsupervised Learning and 3. Reinforcement Learning, Algorithms related to Unsupervised Learning have no attribute to predict fixed to the data. In place of output the data has an input which would be multiple attribute that describe the data.

The study implemented an approach that composed of an unsupervised learning. Initially, the study aimed to explore the potential presence of coherent clusters of Vet care centers. The K-Means algorithm was used to discover the groups of vet care center nearby location. The prediction model developed to clump the vet care centers and helps the centers as well as the individuals to access the information about animals.
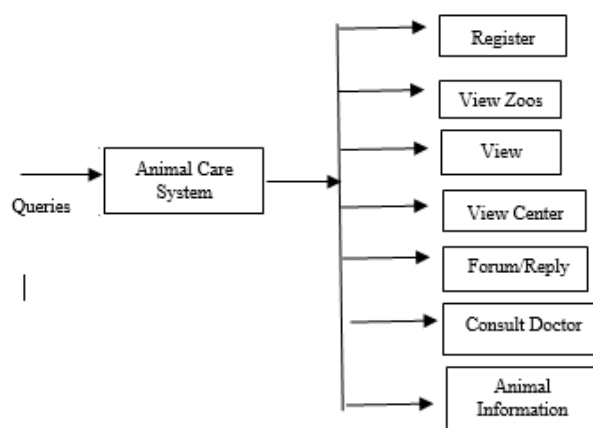

Fig. 1. Approach Overview

The figure 1 describes various modules such as Register, view zoos, etc. which can be accessed by the User through the proposed Complete Animal Care system. The User can Register and then Login to access the various modules.

K-Means is based on the concept of centroid-based clustering. A centroid is an any data point at the center of a cluster. In K-means clustering, clusters are represented by a centroid. It is not necessary that centroid contains in the dataset.

A. Data Description

Table 1 lists the attributes that were initially considered as probable features for the clustering and prediction models as well. Dataset consists of the database records of Veterinary hospitals, adoption centers, forum details record and doctor forum.

TABLE I.	ATTRIBUTES EXPLORED AS POSSIBLE FEATURES.	
Attributes Explored		
NGO Id	Address	Forum Id
Name	Operational Since	Creation Date
Registration id	Logo	Title
Phone number	Website	Description
Email id	Area of Interest	Gender
Google Map Link	Doctor Id	Solution
Concern Id	Query	User Id

B. Methodology

The main opinion in partition-based clustering algorithms is that a cluster must comprise at least one attribute and that each attribute must belongs to exactly one group as known as hard clustering[8]. For the development of k-mean clustering algorithm the following steps were followed:

- **Step 1**. Loading the dataset

Fig. 2. Approach overview : Dataset Attributes 1

Fig. 3.

- S**tep 2.** Cleaning the dataset to remove all the null values..
- **Step 3.** First 300 records taken from the data set for analysis.

Fig. 4. Approach overview : Dataset Attributes 2

- **Step 4 :** Attributes NGO name and location (googleMaplink) used to make clusters.
- **Step 5** Determined the number of clusters
- **Step 6.** Analysed cluster formation
- **Step 7.** Examined the coordinates of the cluster centroids
- **Step 8.** Examined how the cluster assignment relates to individual characteristics.

C. Result and Analysis

Fig. 5.   Homescreen

Fig. 6.   Simulation Environment for system

The working and simulation environment of animal care system implemented using Android studio and the Website designed using PHP environment as shown in Figure 5.

Fig. 7.   Adoption center details   Fig .7.   Home Page Android Environment

Figure 6 shows the Homepage in Android Environment. After successful login we get this page where user can check all the centers, a user can ask some queries in the forum section and can check animal information.

Fig 7. Adaption center details

Figure 8 is the page where an admin can add, update or delete any Adoption Centers.

IV. Conclusion And Future Work

In this paper, we designed animal care system and analy Home Page Android Environment zed clustering algorithm (K-means) in order to obtain optimal clusters for an animal care data set.

The future work mainly concentrates on improving the healthcare of all the animals. To give them a proper medical facility, to give them a house, to help other stray find animal centers, to make everything in one portal. Currently, good and legal vet hospitals are difficult to find and with our proposed System we shall be able to find the legal vet Hospitals/animal centers. We can create a network of users who can assist an organization and or an individual, for spreading the awareness by helping the strays reach a proper care center, which is difficult in current scenario. A user or organization can connect to many different animal lovers. Direct contact to the vets which was not possible can be made available in the current system.

References

[1]. Jiawei Han and Micheline Kember, Data Mining: Concepts and Techniques Second Edition , Morgan Kaufmann Publishers. San Francisco: 2006.
[2]. Ian. H and Eibe Frank, Data Mining Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers. San Francisco: 2005.
[3]. Merlin Jacob, Anina John, "Improved Clustering of Documents using K-means Algorithm ",International Journal of Engineering Research & Technology ,Volume. 5-Issue.07, July-2016.
[4]. Sudhir Singh and Nasib Singh Gill, "Analysis and Study of K-Means Clustering Algorithm", International Journal of Engineering Research & Technology, Vol.2 -Issue 7 (July - 2013).
[5]. Gracy Joseph, Shravan Rao, Spurthi Shetty, Prof. Aruna Pavate," Complete Animal Care System using Clustering Algorithm: CACS", IJERT ICIATE – 2017 (Volume 5 – Issue 01)
[6]. Mahmoud Elbattah, Owen Molloy," Clustering-Aided Approach for Predicting Patient Outcomes with Application to Elderly Healthcare in Ireland", The AAAI-17 Joint Workshop on Health Intelligence WS-17-09,pp 533-541
[7]. Parasian D.P Silitonga,"Clustering of Patient Disease Data by Using K-Means Clustering", International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 7, July 2017 pp 219-221
[8]. Pranav Nerurkar, Archana Shirke, Madhav Chandane ,sunil Bhirud," Empirical Analysis of Data Clustering Algorithms", 6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8
[9]. December 2017, Kurukshetra, India procedia Computer Science 00 (2018) 000–00
[10]. Shenghui Wang, Rob Koopman "Clustering articles based on semantic similarity" , Journal Scitometrics archive Volume 111 Issue 2, May 2017 Pages 1017-1031 Springer-Verlag New York, Inc. Secaucus, NJ USA
[11]. Jain A. K.: Data clustering: 50 years beyond K-means. Pattern recognition letters 31(8) 651−666(2010).
[12]. Marco Capó, Aritz Pérez, and Jose A. Lozano," An efficient K -means clustering algorithm for massive data", JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015
[13]. Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on (pp. 63-67). IEEE.
[14]. Godwin Ogbuabor1 and Ugwoke, F. N2," Clustering Algorithm For A Healthcare Dataset Using Silhouette Score Value", International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018
[15]. Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications, 41(4), 1476-1482.
[16]. Oyelade, O. J , Oladipupo, O. O , Obagbuwa, I. C ,"Application of k-Means Clustering algorithm for prediction of Students' Academic Performance " , (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, o. 1, 2010
[17]. S. Na, L. Xumin and G. Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, Jinggangshan, 2010, pp. 63-67
[18]. Huda Hamdan Ali1 , Lubna Emad Kadhum2," K- Means Clustering Algorithm Applications in Data Mining and Pattern Recognition ", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2015): 78.96
[19]. Yousef Farhang ," Face Extraction from Image based on K-Means Clustering Algorithms", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 9, 2017

[20]. Unnati R. Raval, Chaita Jani," Implementing and Improvisation of K-means Clustering ", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.11, November- 2015, pg. 72-76

[21]. Peter O. Olukanmi , Bhekisipho Twala," K-means-sharp: modified centroid update for outlier robust k-means clustering ", Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech) Bloemfontein, South Africa, November 29 - December 1, 2017

[22]. Wang, T., Bucci, D. J., Liang, Y., Chen, B., & Varshney, P. K. "Exponentially Consistent K-Means Clustering Algorithm Based on Kolmogrov-Smirnov Test.", 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).doi:10.1109/icassp.2018.8461730

[23]. Huan, Z., Peng zhou, Z., & Zeyang, G. (2018)," K-means Text Dynamic Clustering Algorithm Based on KL Divergence", 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS).doi:10.1109/icis.2018.8466385

[24]. Al-Azzawy, D. S., & Al-Rufaye, F. M. L. (2017), " Arabic words clustering by using K-means algorithm.", 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT).doi:10.1109/ntict.2017.7976098