



SURVEY ON ADVERTISEMENT ANALYSIS USING CLASSIFICATION AND CLUSTERING ALGORITHMS IN DATA MINING

Mrs.C. Kalpana M.E.,^[1] Madhu Mathi Rani.L^[2], Vidhya Prabhavathi.S^[3], Sri Saranya.S^[4]

^[1] Assistant Professor ^[2, 3, 4] B. Tech Student

^[1, 2, 3, 4] Department of Information Technology

^[1, 2, 3, 4] KGiSL Institute of Technology, Coimbatore, India.

Abstract- Internet has become the part of our day to day life, in which social media plays a major role by becoming one of the important aspects of digital marketing. Through social media, millions of information is shared among people. Thus it has become the more impacting media. Social media has a power to influence people in multiple ways, it is highly used as a marketing strategy. It is also an easy way to connect with customers and business sector. When an organisation posts advertisements on social media, the response of the advertisement depends on the parameters considered. To increase the traffic of the business, various parameters like number of views, product rating etc., are taken and analysed using five different algorithm techniques. And also with the respondent's data, accuracy of five different algorithms is concluded. With the rapid growth of social media many people are connected to internet providing an outlet for companies to reach buyers. For publishers it is important to get maximum profit from the advertisements posted. In order to reach out to large audience, an organisation ends up posting the same ads on different social media. Hence in order to predict and conclude on the basis of large amount of data k-nearest neighbour, k-means, k-medoids, linear model, SVR algorithms are used to analyse the advertisements.

I. Introduction

The history of advertising are often discovered to ancient civilizations. It became a major inverting economies in the mid-19th century, based on newspapers and magazines. The advertisement grew rapidly with the new technologies such as radio, mail, Television in the 20th century. Throughout the 20th century means of advertising had expanded drastically too many forms. First the radio was introduced in the year 1920 and 1930's, allowing advertisement to publish all over the world. The radio expanded on the idea of targeting in modern advertising. In 1950's television was another medium for advertising. Hence television was an advertisement in effective advertising. As personal computers and the World Wide Web became popular in 1990's, a final means of advertising appeared in the twentieth century. Online advertising allowed users to not only to see and hear advertisement, but also learn more about the company's growth. Although there had been major progress made in the 20th century, advertisers were still only able to advertise to groups and not by individuals. As data mining is defined as the combination of statistics, machine learning and artificial intelligence in order to draw prediction and conclusion on large data set we use three clustering algorithms.

II. Data Mining

As more and more technologies are becoming computerized our ability to generate and collect data is expanding rapidly, data mining was introduced in 1990. Data mining is also called as "Knowledge discovery in database", the process of discovering interesting and useful patterns and relationships in large volume of data. Data mining produce a technique to analyze big set of data has created a need for new methods and which will convert big data to useful knowledge. This demands in driving the expansion and modernization of the practice of data mining. More precisely, data mining is the process of discovering patterns in large sets of data. Data processing in some ways is fundamentally the difference of machine learning techniques to business applications. Data processing is best described as union of historical and up to date development in statistics, AI and machine learning. These technique are then used together to study and find a pattern to it.

III. Clustering In Data Mining:

A cluster is a subset of similar objects. A subset of data such that the distance between any two objects is less than the distance between any object in the cluster and any object that is not located inside the cluster. Clustering is the method of converting a group of objects into classes of similar objects. Clustering could be a process of partitioning a group of information or objects into a group of sub classes called clusters. Clustering algorithm have been made to standardize the data that can perform in all situations.

1. K-Means Algorithm

The k-means clustering algorithm is a simple unsupervised algorithm that is used to predict the classification from an unlabelled dataset. The prediction is based on the number of cluster center present (k) and the nearest mean value. Finally k-means algorithm aims at minimizing an objective function mentioned as a mean squared error function given by:

where,

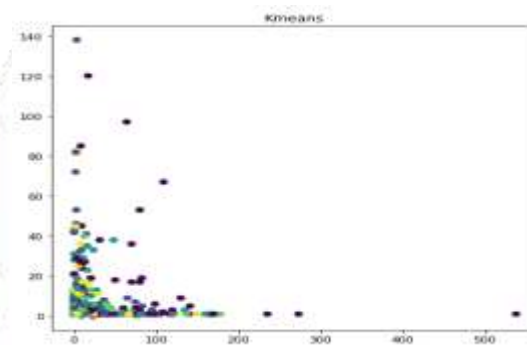
' $\|x_i - v_j\|$ ' is that the Euclidean distance between x_i and v_j .

' c_i ' is that the number of knowledge points in i th cluster.

' c ' is the number of cluster center.

Let $X = \{x_1, x_2, x_3, x_4, \dots, x_n\}$ will be the set of data points and $V = \{v_1, v_2, v_3, v_4, \dots, v_n\}$ will be the set of centers.

- 1) Randomly select 'c' cluster centers.
- 2) Calculate the space between each datum and cluster centers.
- 3) Allocate the information point to the cluster center whose distance from the cluster center is minimum of all the cluster center.
- 4) Recalculate the new cluster center using: where, ' c_i ' represents the quantity of knowledge points in i th cluster.
- 5) Recalculate the space between each datum and new obtained cluster centers.
- 6) If no datum was reassigned then stop, otherwise repeat from step 3.



2. k-Nearest Neighbouring Algorithm

K-Nearest Neighbouring algorithm is one of the simplest Machine Learning algorithm based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new data and the available data. K-NN may be a non-parametric algorithm, which suggests it doesn't make any assumption on underlying data. It is also called as a "Lazy Learner Algorithm" because it does not learn from the training set instead it stores the dataset and at the time of classification it performs an action on the dataset. K-NN can be used for regression as well as for classification problems.

The K-Nearest Neighbor algorithm working can be explained on the basis of the below steps:

Step-1: Select the number K of the neighbours

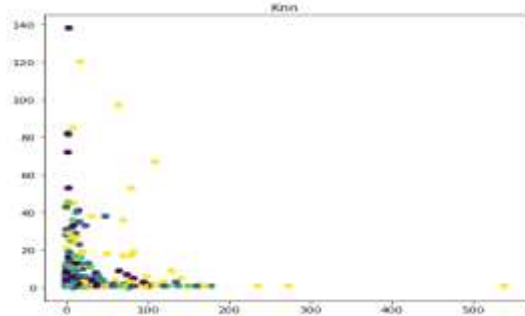
Step-2: Calculate the Euclidean distance of K number of neighbours

Step-3: Take the K nearest neighbours as per the calculated Euclidean distance.

Step-4: Among these k neighbours, count the number of the data points in each category.

Step-5: Assign the new information points to that category for which the number of the neighbour is maximum.

Step-6: Our model is ready.



3 .K- medoids

Partitioning around Medoids or the K-medoids algorithm may be a partitional clustering algorithm which is slightly modified from the K-means algorithm. They both plan to minimize the squared-error but the K-medoids algorithm is more robust to noise than K-means algorithm. In K-means algorithm, they choose means because the centroids but within the K-medoids, data points are chosen to be the medoids. A medoid are often defined as that object of a cluster, whose average dissimilarity to all or any the objects within the cluster is minimal.

The algorithm proceeds in two steps:

- **BUILD-step:** This step sequentially selects k "centrally located" objects, to be used as initial medoids
- **SWAP-step:** If the target function are often reduced by interchanging (swapping) a specific object with an unselected object, then the swap is administered. This is continued till the target function cannot be decreased.

The algorithm is as follows:

Step-1: Initially select k random points because the medoids from the given n data points of the information set.

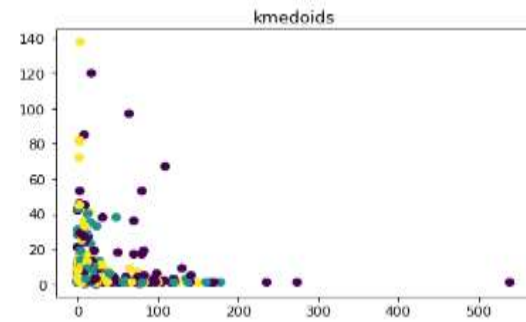
Step-2: Associate each datum to the closest medoid by using any of the foremost common distance metrics.

Step-3: For each pair of non-selected object h and selected object i calculate the entire swapping cost

TC_{ih} .

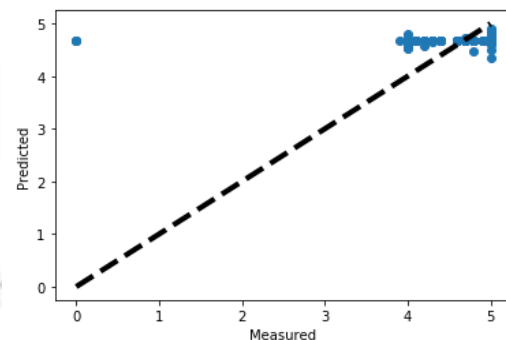
replaced i by h , If $TC_{ih} < 0$.

Step-4: Repeat the steps 2-3 until there's no change of the medoids.



4. SVR Algorithm

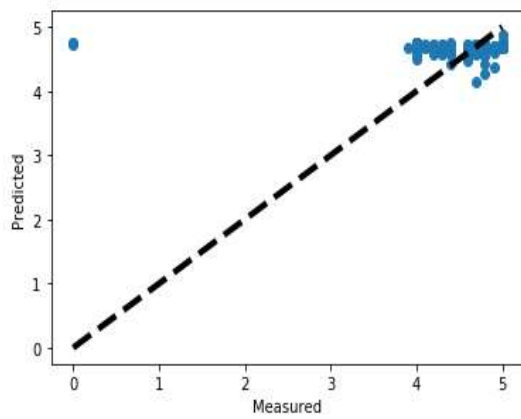
Support Vector Machine can also be used as a regression method. SVR uses the identical principle as SVM for classification, with only few minor differences. The Support Vector Machine might be a discriminative classifier formally defined by a separating hyper plane. When the output might be a push number it becomes very difficult to predict the info, which has infinite possibilities. Within the case of regression, a margin of tolerance (epsilon) is prepared within the approximation to the SVM which could have already requested from the given problem. The foremost idea of this algorithm is to cut back the error, individualize the hyper plane which maximizes the margin and to tolerate the error occurred.



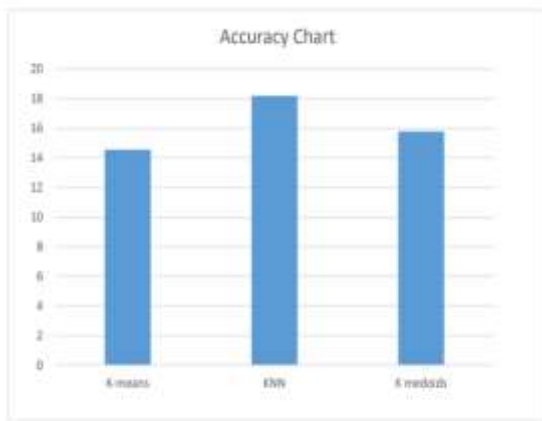
5. Linear Model Algorithm

Linear Regression may be a machine learning algorithm supported supervised learning. It performs a regression task. Linear model is target prediction value supported independent variables. It's mostly used for locating out the connection between variables and forecasting. Linear model performs the task to predict a variable value(y) supported a given independent variable(x). So, this system finds out a linear relationship between x (input) and y (output). Linear algorithm predict the output supported input features from the info fed within the system. This algorithm enables users to summarise and study

relationships between two continuous (quantitative) variables



Accuracy Chart



V. Conclusion

Precision advertisement system is concentrated on “precision”, therefore, the acceptable classification algorithm is particularly important to differentiate different users. On the other hand, the important time of the advertising is additionally very important. Pushing to the user’s ad must be the newly ad. This paper is concentrated on precision advertisement system’s design and realization, especially the classification of different users. We are going to implement by using three clustering algorithms, k-means; k-NN; k-medoids to compare the highest accuracy based on the advertisement posted by an organisation.

V1. REFERENCES

[1] A. Y. Zhou, M. Q. Zhou and X. Q. Gong, “Computational Advertising: A Data-Centric Comprehensive Web Application,” Chinese Journal of Computers, vol.34 (10), pp. 1805–1819, Oct. 2011.

[2] J. M. He, Sh. L. Yang and Z. Liu, “Study of Two-stage Advertising Investment Model between Manufacturers and Agents,” Journal of Systems Engineering, vol. 25 (4), pp. 473-478, Aug. 2010.

[3] Sh. P. Yu and G. Chen, “An Efficient Algorithm of Behaviour targeted Advertising,” Computer Applications and Software, vol. 28(1), pp. 4-7, Jan. 2011.

[4] N. Li, “The Research of Precision Advertising System Based on Incremental Learning,” Shanxi University of Finance & Economics, 2012.

[5] Jerzy Surma, “Selecting Data Mining Model for Web Advertising in Virtual Communities”, the first International Conference on Advances in Information Mining and Management: IMMM, 2011.

[6] Priyanka Saini, Sweta Rai, Ajit Kumar Jain, “Data Mining Application in Advertisement Management of Higher Educational Institutes”, International Journal of Computer-Aided technologies, Vol-1, April 2014.

[7] Maria Diapouli, Stelios Kapetanakis, Roger Evans, Miltos Petridis, “Behavioural Analytics using Process Mining in Online Advertising”, University of Middlesex, UK.

[8] Izabela A. Wozczko, “Skills and Vacancy Analysis with Data Mining Techniques”, Institute of Technology Blanchardstown, Ireland, Informatics-2015.

[9] Charles X. Ling, Chenghui Li, “Data Mining for Direct Marketing: Problems and Solutions”, University of Western Ontario, London.

[10] Prof. Dr. Raj B. Kulkarni, Ms Shankari V. Gajul, “E-Marketing Techniques using Data Mining Approach”, International Journal of Advanced Research in Computer and Communication Engineering, Vol-4, Issue-10, October 2015.

[11] Avinash Sharma, Swati V. Kulkarni, Dhanajay Kalbande, Surekha Dholay, “Cost Optimized Hybrid System in Digital Advertising using Machine Learning”, International Journal of Innovative Technology and Exploring Engineering(IJITEE), Volume-8, Issue-8S2, June 2019.

[12] Elangovan.D, Dr. Subedha. V, Sathishkumar. R, “A Survey: Data Mining Techniques for Social Media Analysis”, Advances in Engineering Research (AER), Volume-14.