

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

Shuchita Mudgil

Kalaniketan Polytechnic College Jabalpur

Abstract:

The COVID-19 epidemics have made it challenging for consumers and families to sustain a hale and hearty daily life, eminence of living, and well-being. Preliminary proof has recommended that higher probability of both mortality and severity of the COVID-19 be directly connected to unhealthy daily life behaviors. In this research article, we identify impact of epidemic Corona virus on consumer sentiment thoughts or feelings on their daily life style. Sentiment analysis on various consumers have been to gather data , in addition, by the latest advancement in machine learning algorithms, the precision of our sentiment analysis predictions is able to get better. In this report, we will try to perform sentiment analysis on "lifestyles" using various rapid miner tools. According to literature survey done in respect to sentiment analysis, there are two techniques i.e. semantic orientation and machine learning which are important. The semantic orientation of a view suggests whether the view is positive, negative or neutral whereas machine learning is a technique of data analysis which automates logical building of a model. In this paper, we will use a variety of fast mining tools to perform sentiment analysis on "lifestyles." We attempt to categorize the polarity of a lifestyle as positive, negative, or neutral. According to a literature review on sentiment analysis, there are two methodologies that are important: semantic orientation and machine learning. Machine learning is a data analysis technology that automates the logical creation of a model, whereas semantic orientation of a view indicates whether the view is positive, negative, or neutral. The two Algorithm of Machine learning are used K-means Clustering & X-Means Clustering calculating the correct number of centroids.

Keywords: Machine Learning, Big Data, Sentiment Analysis, K-means Algorithm, X-means Algorithm

Introduction:[1][2]

The automated technique of evaluating whether a text indicates a positive, negative, or neutral attitude about a product or issue is known as **sentiment analysis**. Sentiment analysis is the process of analyzing consumer sentiment utilizing natural language processing, text analysis, and statistics. To assess and comprehend consumer sentiment, including what individuals say, how they say it, and

what they mean. Tweets, comments, reviews, and other sites where people mention your goods might reveal consumer sentiment. Sentiment Analysis is the domain of using software to interpret these feelings, and it's a must-know for developers and business executives in today's workplace.

Covid-19 has drastically transformed our life over the last two years. People were locked up at home owing to induced lockdown at the same time as the environment began to heal. Significant modifications in lifestyle were seen. Residents of societies where social connection was forbidden were educated to avoid it. Citizens begin to adapt, and a new way of life emerges as a result. We have realized it; the corona virus has caused so many changes in our daily life that would not have occurred otherwise.

NLP (Natural Language Processing), biometrics, text analysis, and computational linguistics are all used in opinion mining to detect, extract, and refer subjective information. Sentiment analysis is used to determine a writer's attitude toward a topic or the complete polarity of a work. The conduct could be the author's evaluation, judgment, or affective state, or the emotional communication or interlocutor's emotional communication. It is the calculated analysis of a user's thoughts, feelings, and behavior toward an object. Sentiment mining is a technique for gathering positive, negative, or neutral feedback on a product.

Sentiment analysis is performed at three levels using studies by Bandanna, et al [3], namely:

- **Document level analysis:** The goal here is to figure out what the document's general opinion is. At the document level, sentiment analysis implies that each document communicates ideas about a particular entity.

- **Sentence-level analysis:** At this level, the goal is to figure out if each sentence has communicated an opinion. This level distinguishes between objective and subjective sentences, which communicate facts and opinions, respectively. In this scenario, the treatments are twofold: first, determine whether or not the sentence expresses an opinion, and then assess the polarity of that opinion. The fundamental problem arises from the fact that objective sentences can also include an opinion.

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

Shuchita Mudgil
Kalaniketan Polytechnic College Jabalpur

• **Aspect level analysis:** This level conducts a more detailed investigation and necessitates the use of natural language processing. Opinion is characterized by polarity and a target of opinion at this level. Treatments in this scenario are twofold: first, identify the thing and its aspects, and then assess the opinion on each aspect.

Sentiment Mining has become extremely popular in the field of research technique. A lot of research has already been done but still certain challenges to sentiment mining still exist related to unstructured data. As per the study of various published papers, it can be accomplished that supervised techniques provide much better accurate result in comparison to dictionary technique.

The main aims of this paper are as follows:

- The main goal of this research is to develop a deep understanding of the various opinion mining and sentiment analysis approaches performed on individual view of text analytics
- Evaluation of sentiments of unstructured data using Clustering in Rapid miner tool

A.) Types of sentiment analysis [12]

1. **Manual processing:** Human interpretation of the sentiment must be accurate.

2. **Keyword processing:** Assign positivity or negativity to individual words and calculates the overall percentage score to the post.

3. **Natural language processing (NLP):** Also called text analytics, computational linguistics.

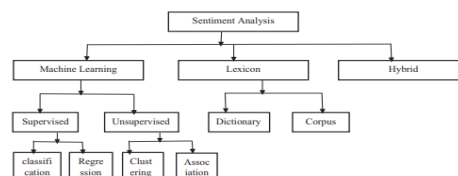
NLP is superior to keyword processing. NLP works by analyzing language for its meaning. The information what the vendors get from sentiment analysis provides them to improve their marketing strategy. By sentiment analysis, the researcher can see the positive or negative discussions among their audience. By sentiment analysis, the researcher knows the customer's opinions about their views.

The opinion are not judged by their functionality, instead of how well it is presented on the online reviews. Sentiment analysis can be measured using They are

- machine learning,
- lexicon based, and
- Hybrid-based approaches.

In the **machine learning approach**, the supervised learning model can be easily trained, and the unsupervised model can be easily categorized the data. The **lexicon-based approach** can be easily calculating the sentiment scores for each word. The **hybrid** is a combination of both machine learning and lexicon-based approaches and measures the sentiment for noisy and less sensitive data.[10][11][12] [14]

The sentiment analysis can be divided into different categories as shown in Figure 1



B.) Overview:

Nowadays and since the last two decades, digital data is generated on a massive scale, this phenomenon is known as Big Data (BD). This phenomenon supposes a change in the way of managing and drawing conclusions from data. Moreover, techniques and methods used in artificial intelligence shape new ways of analysis considering BD. Sentiment Analysis (SA) or Opinion Mining (OM) is a topic widely studied for the last few years due to its potential in extracting value from data. As the amount of the data generated by users increases, due to the improvement of internet connections worldwide, the challenges to handle this amount of data increase too. For this reason, tools like machine learning (ML) can help

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

Shuchita Mudgil

Kalaniketan Polytechnic College Jabalpur

organizations and individuals to handle and take advantage of data generated by users.

In Attribute level sentiment analysis, a sentiment for each entity in a sentence is provided. In sentence level sentiment analysis, the overall sentiment of each and every sentence in a document is provided. In Document level sentiment analysis, the overall sentiment of the complete document is provided. According to literature survey done in respect to sentiment analysis, there are two techniques i.e. semantic orientation and machine learning which are important. The semantic orientation of a view suggests whether the view is positive, negative or neutral whereas machine learning is a technique of data analysis which automates logical building of a model. The techniques are shown in in figure: **(Opinion Mining Techniques)[2]**

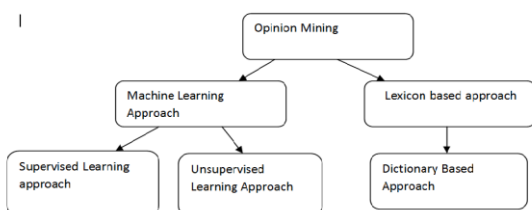


Fig 2.1 Opinion Mining Techniques

- In the **supervised learning** approach of machine learning, pseudo codes are trained using descriptive examples, as an input in which the desired output is already known. It is basically used in applications where historical data is used to predict forthcoming data.
- In the **unsupervised learning** approach of machine learning, there is no historical data. The objective is to investigate the data and to make some useful information within it .

2. Methodology :

The main goal of this research is to develop a deep understanding of the various opinion mining and sentiment analysis approaches performed on individual view of text analytics. This study advocates the ways applications are present and

utilized in many areas in the society. The technique used in the study is the systematic literature review. A systematic review is completely based on an evidently framed question, presents relevant studies, evaluates their findings, and summarizes the data by means of clear methodology. This explicit and methodical approach makes systematic reviews different from the traditional reviews

The blend of evidence from the current literature can create new knowledge in the existing studies, which is as important as conducting new research .Rousseau *et al.* [18] maintained an argument that systematic reviews are different from tradition reviews in that systematic literature reviews are comprehensive in nature, use transparent and fair analysis, and apply certain criteria for understanding of the findings that provided in the previous literature. In addition, systematic literature reviews mainly focus on objectivity and reproducibility of results [11]. The process of review starts with framing the questions and conducting a systematic and step-by-step process and applying a replicable method to answer these questions [11]. Thus, the evidence generated from the exact approach of identifying, selecting, and analyzing the data can have a significant impact on the body of knowledge collected, but the supreme concern of this practice is synthesizing the results produced through this systematic process [10], [11].

The methodology used in the study is a five-step process shown in Figure, as proposed by [10]. It is systematic in nature, clear and reproducible, and involves identifying, examining, synthesizing, deducing, and reporting the evidence from the existing sentiment analysis and opinion mining literature.

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

Shuchita Mudgil

Kalaniketan Polytechnic College Jabalpur

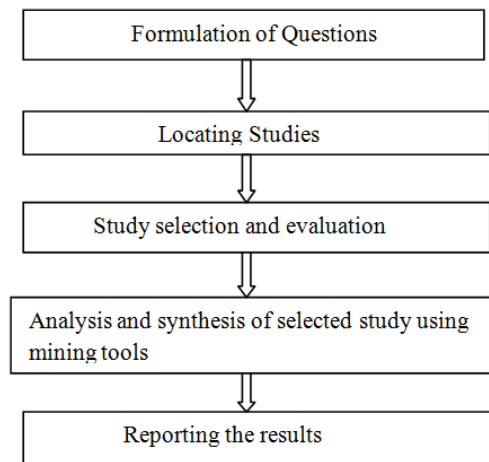


Fig :Research methodology of systematic literature review

1.QUESTION FORMULATION :

A deep and insightful literature review should start with the development of a clear understanding about your objectives [18]. Therefore, to ascertain this, we clearly formulated and considered research questions to evade doubts in our study [19]. The purpose of the research is to discuss the methodological and application side of opinion mining and sentiment analysis and explore whether the intervention of opinion mining and sentiment analysis would be applicable to humans or in an organization as a whole

<https://forms.gle/Kp9vcxqBUXinjUyUA>

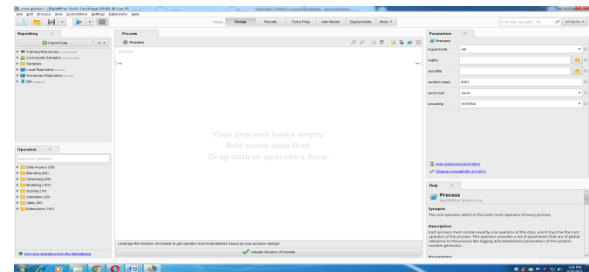
The above link is having the list of questionnaire of data collection of my research topic

B. LOCATING STUDIES

The objective of identifying an suitable academic journal articles is to develop a list of all related articles to our research questions. I have prepared a list of selected questionnaire as a core database. Since the study is based on opinion mining and sentiment analysis, we used different strings to identify relevant papers.

But in this research the positive and negative states are highlighted with the help of tool.

Rapid Miner 9.8.001 is data science platforms that combine with analytics and artificial intelligence. This tool includes text mining also.



Rapid Miner studio is visual workflow designer, which is used for predictive analysis. This combines machine learning and big data science for the better analysis. Understanding of the rapid miner studio is important to design the process diagram. The IDE (Integrated Development Environment) has repository, Operators, Views, Global Search, Parameters, help panel, Functional ports and Process panel with usual tool such as file, edit.

The benefits of the Rapid Miner explained in detail as it offers a robust and very powerful tool and have plenty of features. In rapid miner tool each component is user-friendly environment interface which helps to realize the users about massive productivity gains. It is a visual work-flow designer tool is specially contributing the users with an environment. This environment enables the user to create, design and deploy analytics processes, visual presentations, and models. The guidance of Rapid Miner tool operators uncluttered, meaningless, disorganized, expressions and seemingly useless data also enhanced into very valuable and visualizable. The system facilitates data access, manage accessed data, load and evaluate of data such as texts, images, and audio tracks. Rapid Miner allows the user to structure data to be mending to arrange. Rapid Miner used to create models and plans with a strong set of tools and functionalities so that the user has the possibilities to extricate or extract critical statistics and information. The user is fully allowed to utilize data exploration capabilities and descriptive statistics like graphics and visualization. This software is more powerful which offers analytics related to real-world data transformation settings which gives the mastery of formatting and creating the most favorable data set for user predictive analytics[20] [21].

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

Shuchita Mudgil

Kalaniketan Polytechnic College Jabalpur

Explicit Features of the Rapid Miner tool is explained in the below: [20][21]

1. Visual Workflow Designer - Rich library with over of 1500 machine learning algorithms, drop and drop visual interface, pre-built templates, possibility of proactive recommendations.
2. Data Access and Management- Accessing files including SAS (Statistical Analysis System), ARFF(Attribute-Relation File Format), STATA (software for statistics and Data science) and via URL.
3. Data Exploration- A label in column target variable.
4. Descriptive Statistics- Uni-variate statistics and plots, Distribution plots, Bi-variate statistics and plots.
5. Graphs and Visualization- visualization module is created as an alternative to the well known and old module Plot View.
6. Data Prep- Turbo or data Prep is a new module in Rapid Miner that speed up time-consuming data prep tasks to speed the productivity of analytics teams.
7. Data Sampling -The number of examples in the sample can be specified on absolute, relative or probability basis depending on the setting of the sample parameter. Moreover, the samples are generated randomly.
8. Data Partitioning - In this partitioning the parameter is used to separate the number of subsets or partitions and the relative size of each partition. It is named as partitions parameter.
9. Data Replacement - The data replacement operator enables to select attributes to use replacements in and to a specify regular expression. Attribute values of selected attributes that match this regular expression are replaced by the specified replacement. The replacement can be empty and can contain capturing groups.
10. Weighting and Selection - Select by Weights operator allows selecting attributes using the weights of the attribute.
11. Similarity Calculation - A similarity calculate object measure the calculated similarity between each data to similarity and similarity to data. The Data to Similarity operator and vice versa can generate a similarity calculation or measure object.
12. Clustering - Clustering groups Examples together which are similar to each other. As no Label Attribute is necessary, Clustering can be used on

unlabeled data and is an algorithm of unsupervised machine learning.

13. Market Basket Analysis - Market Basket Analysis is a association analysis method which is used to discover attractive relationship between the variables in the given set of data set. The modeling of association between variables is based on the set of items, frequently appeared together.

14. Bayesian Modeling - Naive Bayes model is a high-bias and low-variance classifier. It has the capability to build a better model even with a minimum number of data. It is very simple to use and understand. The use cases mostly involved with text categorization, spam detection, sentiment analysis, and popularly known for recommender systems.

15. Scoring - Rapid Miner Time Scoring is an additional service to Rapid Miner. It is designed for fast scoring use cases through the web services which fulfilled by the components, they are called as Scoring Agents. It access minimal memory and have fast response times.

16. Automation and Process Control - It execute multiple process in parallel. The long-term process can be run in the background, while the other process execute in the front. The main services are background process execution, automatic optimization, scripting in data preparation, logging in process, macros, process control and process based reporting.

C .STUDY SELECTION AND EVALUATION

In order to guarantee and maintain the quality of the paper, we have constrained our selection of articles to only peer reviewed journals. Peer-reviewed journals have strict quality control and have gone through methodical, accurate processes and have rigorous requirements for publication, which leads to better research output [20]. The process began with studying and scanning of selected articles from the journal database. The initial criterion of selection was based on choosing the keywords ``opinion mining" and ``sentiment analysis."

The next step involved reading the abstracts to assess whether it was applicable to my research topic. Initially a people read it, but to justify its rigor, an independent consumer to improve its impartiality and strength read the same number of articles. Intellectual outputs that did not align with my research questions or that seemed inappropriate and non-substantive

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

Shuchita Mudgil

Kalaniketan Polytechnic College Jabalpur

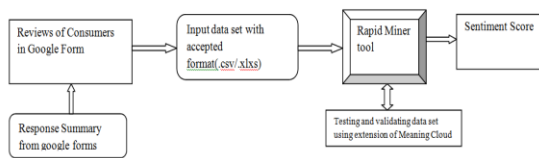
were excluded. The articles that were included exhibit good fit with the objective of the study.

The selected papers were then examined in detail and synthesize to answer the research questions. For the preferred papers, I created the taxonomy which is represented in following sections (dataset, methods, application, and major challenges).

3. DATASETS:

A more in detail analysis was done regarding the sources of datasets were from the created questionnaire containing total 12 questions. The main source of data is from the articles from Google. The Google form was asked to fill by various consumers /person, nearly 1010 entries were recorded to perform the analysis for big data

4. Method and Proposed framework :



Methodology Diagram for Proposed Framework

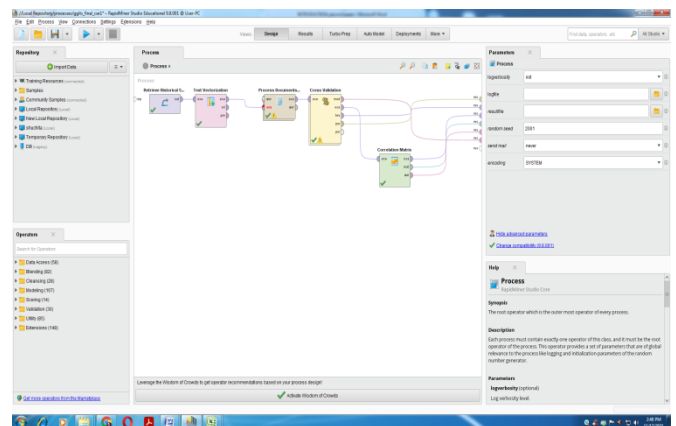
Statistical text mining using rapid mining model, includes loading the data, Pre-processing the data, generating term-by-document matrix, building models and lastly applying the model on new data to predict the outcome. A process diagram also created for similarity-based methods and clustering techniques for measuring the similarity between the documents. The above figure explains the idea of proposed framework clearly. First, the data needed to be analysis are gathered from the various entries from Google form.. The review data are gathered in the acceptable Google form and then the response summary is downloaded in the excel sheet. The acceptable dataset is entered the tool to analysis based on the training data if the algorithms are depending on unsupervised learning.

Prior to that, the admissible data-set is processed using pre-processing procedures. Pre-processing

entails transforming the given case into an accepted case, as well as Text Vectorization, document processing, and cross validation. Clustering, which can be performed on unlabeled data and is an unsupervised machine learning algorithm, is a building of the algorithm utilizing the k-means algorithm(Clustering groups Examples together that are similar to each other.). The output of the k-means algorithm is fed into X-Means, a clustering technique that uses a heuristic to calculate the optimal number of centroids. It starts with a small number of centroids and then iterates to see if adding more makes sense based on the data. The data model is created by combining the k-means technique and the X-Means algorithms for dataset cross validation. Finally, the Correlation matrix for sentiments is created, which determines the correlation between all attributes and may be used to generate a weights vector based on these correlations.

5. Experiment and Performance Analysis :[29]

Rapid Miner is a fantastic application that includes text processing skills as well as a third-party Application Programming Interface (API) that is simple to connect to. The process diagram for sentiment analysis is depicted in the graphic below. Covid-19's impact on consumers' daily life



The sentiment analysis in Rapid Miner begins with the Retrieve Operator, which imports the excel sheet to be analyzed. This Operator has access to the Repository's stored data and can load it into the

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

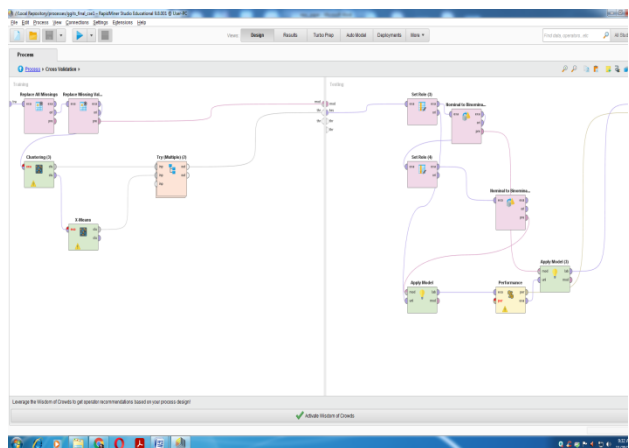
Shuchita Mudgil

Kalaniketan Polytechnic College Jabalpur

Process. This method of data retrieval also returns the Rapid Miner Object's Meta data. There are 1010 examples in the Example collection, each with three special features and 14 regular attributes.

The next stage is to do text vectorization before extracting/generating word vectors from string attributes in process documents from data in order to perform cross validation to assess a learning model's statistical performance.

A nested Operator is the Cross Validation Operator. It has two subprocesses: one for training and the other for testing. A model is trained using the Training subprocess. In the Testing subprocess, the trained model is used. During the Testing phase, the model's performance is evaluated.



The k-means technique is used by this Operator to cluster data. Examples that are similar to each other are grouped together in a cluster. Clustering is an unsupervised machine learning approach that can be performed on unlabeled data because no Label Attribute is required. The k-means method creates a set of k clusters and assigns each Example to one of them. The clusters are made up of examples that are comparable. A distance measure between Examples is used to determine how similar they are. The position of the centre in the n-dimensional space of the n Attributes of the Example Set determines a cluster in the k-means algorithm. The centroid is the name given to this location. The **k-means algorithm** begins with a set of k points that serve as the centroid of k potential clusters. If determine good start values

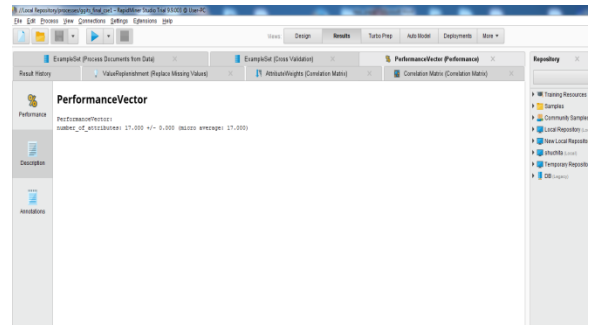
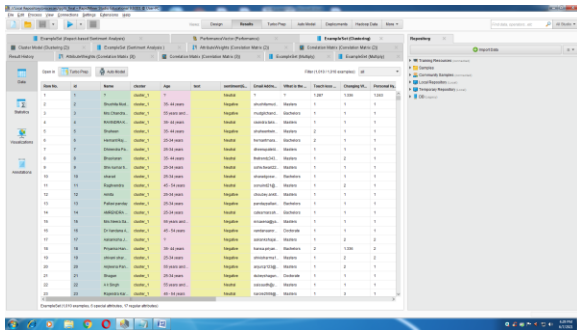
is set to true, these start points are either the positions of k randomly chosen Examples from the input Example Set, or they are decided by the k-means++ heuristic. All Examples are assigned to the cluster that is closest to them (nearest is defined by the measure type). The cluster centroids are then recalculated by averaging over all examples of a single cluster. The previous stages are repeated for fresh centroids until they no longer move or the maximum number of optimization steps is reached. The operation is done as many times as possible, with a different set of starting positions each time. The set of clusters with the shortest sum of squared distances between all Examples and their respective centroids is delivered. This output is now fed into the X-means operator. **X-Means** is a clustering algorithm that uses a heuristic to calculate the correct number of centroids. It starts with a small number of centroids and then iterates to see if adding more makes sense based on the data. Now the input of both K-means and X-means is given to **Try multiple**, this operator can be used to try different processing variants for a given input.

The output of try operator id given as input to **set role operator**, An Attribute's role explains how other Operators interact with it. Regular is the default role; additional roles are categorized as special. The output of set role operator is given to apply model, the goal is to get a prediction on unseen data or to transform data by applying a preprocessing model. The Example Set upon which the model is applied, has to be compatible with the Attributes of the model. This means, that the Example Set has the same number, order, type and role of Attributes as the Example Set used to generate the model. Now the performance operator is a very simple operator. It takes an Example Set as input and returns a performance vector that has the count of attributes in the given Example Set and given as input to apply model to get a prediction on unseen data to estimate the statistical performance of a learning model.

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

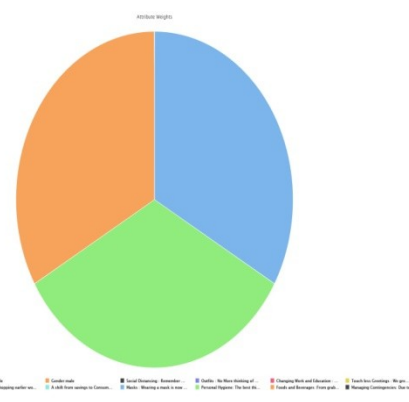
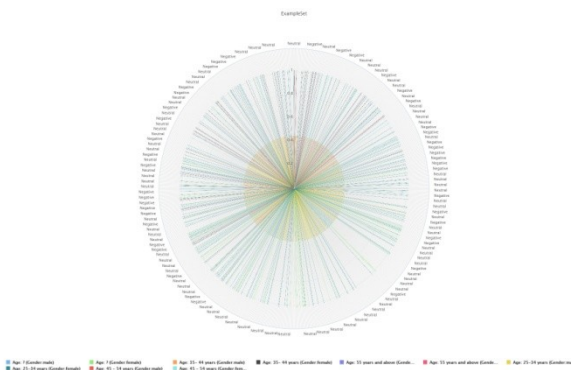
Shuchita Mudgil

Kalaniketn Polyetchnic College Jabalpur

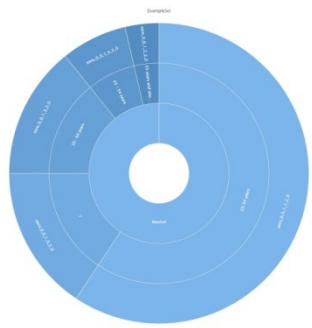


The test result port delivers only an Example Set if the test set results port of the inner Testing sub process is connected.

The example set port returns the same Example Set which has been given as input to the **Correlation matrix operator**



Now after the cross validation process is completed the port delivers the prediction model trained on the whole Example Set.



Index	Nominal value	Absolute count	Fraction
1	Neutral	651	0.645
2	Negative	358	0.355

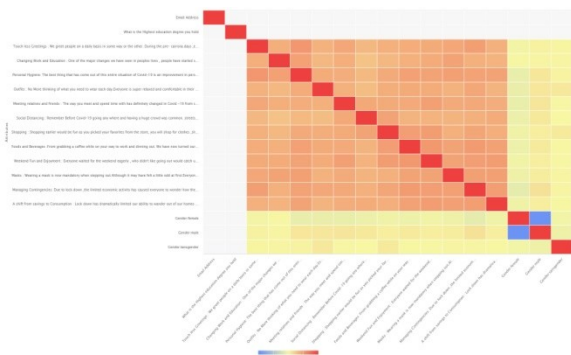
The performance output ports of the Cross Validation Operator deliver the average of the performances over the number of folds iterations.

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

Shuchita Mudgil

Kalaniketan Polytechnic College Jabalpur

The **Correlation matrix** is calculated for sentiments which determine correlation between all Attributes and it can produce a weights vector based on these correlations. Correlation is a statistical technique that can show whether and how strongly pairs of Attributes are related.



6. Discussion and Conclusion:

The major goal of this experiment was to determine an individual's attitudes during a covid19 circumstance. The polarity of unsupervised data was collected, and clustering was performed using k-means and the X means algorithm. The entire data obtained from Google forms responses was 1010 examples, which were imported into Rapid miner, which is one of the top programs with advanced features such as data exploration, sampling, replacement, partitioning, Bayesian modeling, clustering, and modeling assessment. The result of using the rapid miner tool with two different machine learning algorithms gives accuracy to consumer sentiments. The data set was 1010 example set with 6 special attributes and 22 regular attributes, so the sentiment calculated for nominal values is maximum neutral with 651 absolute count and negative absolute count, and the rest is positive.

More research has been done on the evaluation or evaluation of the various approaches of opinion mining and sentiment analysis in terms of applicability. The data sets retrieved from users' application databases thus include an aspect of

human application, despite the fact that these correspond to the evaluation of the methodologies used. The financial, healthcare, hospitality and tourism industries, and marketing-related activities continue to lead the applications. It's also worth noting that the use of opinion mining and sentiment analysis in politics and administration is still on the rise. Etc..

References:

[1] CONCEPT OF SENTIMENT ANALYSIS USING MACHINE LEARNING TECHNIQUES , International Journal of Engineering Applied Sciences and Technology, 2021 Vol. 6, Issue 1, ISSN No. 2455-2143, Pages 141-147 Published Online May 2021 in IJEAST (<http://www.ijeast.com>)

[2] SENTIMENT ANALYSIS HOW COVID-19 CHANGED CONSUMERS DAILY LIVES , International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930 <http://ijsrem.com/current-issues/>

[3] Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review Shanshan Yi1 · Xiaofang Liu2 Received: 11 March 2020 / Accepted: 13 May 2020 / Published online: 11 June 2020

[4] Document Level Sentiment Analysis: A survey S. Behdenna, F. Barigou and G. Belalem Department of Computer Science, Faculty of Sciences, University of Oran 1 Ahmed Ben Bella, PB 1524 El M'Naouer, Oran, Algeria (31000), published :2018-03-14,EAI , <http://dx.doi.org/10.4108/eai.14-3-2018.154339>

[5] A Survey on Sentiment Analysis for Big Data Swati Sharma, Mamta Bansal, Ankur Kaushik.vol no 6,issue no .06, june 2017,International journal of Advanced Research in Science and Engineering

[6] F. R. Lucini et al., "Text mining approach to predict hospital admissions using early medical records from the emergency department," Int. J. Med. Inf., vol. 100, pp. 1_8, Apr. 2017.

Unsupervised data Sentiment Analysis using K-means and the X-means Algorithm

Shuchita Mudgil

Kalaniketan Polytechnic College Jabalpur

- [7] Z. Khan and T. Vorley, "Big data text analytics: An enabler of knowledge management," *J. Knowl. Manage.*, vol. 21, no. 1, pp. 18_34, 2017.
- [8] T. T. Thet, J.-C. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *J. Inf. Sci.*, vol. 36, no. 6, pp. 823_848, 2010.
- [9] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2003, pp. 129_136.
- [10] R. Piryani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000_2015," *Inf. Process. Manage.*, vol. 53, no. 1, pp. 122_150, 2017.
- [11] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," *IEEE Comput. Intell. Mag.*, vol. 9, no. 2, pp. 48_57, May 2014.
- [12] Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges, SHAHID SHAYAA et al, ,IEEE Access ,Received April 24, 2018, accepted June 18, 2018, date of publication June 28, 2018, date of current version July 30, 2018.
- [13] BigData Concepts and Architecture *Journal of Computer Science and Information Systems*, Volume 1. Issue 4, August 2020 ISSN 2535-1451
- [14] A Survey on Sentiment Analysis for Big Data Swati Sharma¹, Mamta Bansal², Ankur Kaushik³ *IJARSE* ,Vol no.6 Issue no.06, june 2017
- [15] D. Denyer and D. Tran_eld, "Producing a systematic review," in *The Sage Handbook of Organizational Research Methods*. Thousand Oaks, CA, USA: Sage, 2009.
- [16] D. Tran_eld, D. Denyer, and P. Smart, "Towards a methodology for developing evidence-informed management knowledge by means of systematic review," *Brit. J. Manage.*, vol. 14, no. 3, pp. 207_222, 2003.
- [17] R. J. Light and D. B. Pillemer, *Summing Up: The Science of Reviewing Research*. Cambridge, MA, USA: Harvard Univ. Press, 1984.
- [18] D. M. Rousseau, J. Manning, and D. Denyer, "11 evidence in management and organizational science: Assembling the_eld's full weight of scientific knowledge through syntheses," *Acad. Manage. Ann.*, vol. 2, no. 1, pp. 475_515, 2008.
- [19] B. Usharani, "Analysis of Supervised and Unsupervised Learning Classifiers for Online Sentiment Analysis "Asian Journal of Computer Science Engineering 2018;3(4):17-21
- [20] Moilanen K, Pulman S. Sentiment Composition. The Oxford Computational Linguistics Group. Proceedings of RANLP; 2007. P
- [21] <https://docs.rapidminer.com/latest/studio/getting-started/>
- [22] Evaluation of Sentiment Data using Classifier Model in Rapid Miner Tool Devipriya B, Kalpana Y. *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019