SENTIMENT ANALYSIS HOW COVID-19 CHANGED CONSUMERS DAILY LIVES

Shuchita Mudgil ,II year M.tech [1]

Kalaniketan Polyetchnic College Jabalpur

Prof,Ashok Verma [2], HOD (CSE)

Gyan ganga Institute of Technology and Science 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 tool. We try to classify the polarity of the lifestyle wherever it is either positive, negative o neutral. The application of opinion mining and sentiment analysis (OMSA) is the calculative study of users views in the era of big data have been used a useful way in categorizing the opinion into different sentiment and in general evaluating the mood of the public. 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

Keywords : Machine learning ,Opinion mining ,Sentiment analysis, Bigdata , rapid miner tool

1. INTRODUCTION:

Sentiment analysis is the automated process of determining whether a text expresses a positive, negative, or neutral opinion about a product or topic. Sentiment analysis is the process of using natural language processing, text analysis, and statistics to analyze consumer sentiment. To analyze and understand the sentiment of the consumers—what people are saying, how they're saying it, and what they mean. Consumer sentiment can be found in tweets, comments, reviews, or other places where people refer to your product. Sentiment Analysis is the domain of understanding these emotions with software, and it's a must-understand for developers and business leaders in a modern workplace.

Over the past few months ,Covid-19 completely changed our lives. At the same time as the environment began to cure ,people where locked up at home due to induced lockdown. Major life style changes were seen. Societies were inhabitants were taught social interaction was prohibited from it. Citizens begin to adapt, and now, a new way of living has emerge. We have realize it, there have been so many changes in our everyday lives because of corona virus, that would have never happened.

Opinion Mining also refers to NLP (Natural Language Processing), biometrics, text analysis and computational linguistics in order to detect , extract and refer subjective information. Sentiment analysis basically aims to identify the attitude of a writer with respect to a topic or the complete polarity to a document. The behavior may be a valuation or judgmental or affective state of the author or the emotional communication or interlocutor. It is the calculative study of users opinions, views, behavior and emotions toward an object. Sentiment mining helps to gather positive, negative or neutral information about a product. Then, the highly counted opinions about a product are passed to the user. For promoting marketing, big companies and business magnets are making use of this opinion mining.

Using given studies by Behdenna, et al [1], sentiment analysis is being performed at three levels i.e.:

- **Document level analysis**: The task at this level is to determine the overall opinion of the document. Sentiment analysis at document level assumes that each document expresses opinions on a single entity.

- **Sentence level analysis**: The task at this level is to determine if each sentence has expressed an opinion. This level distinguishes the objective sentences expressing factual information and subjective sentences expressing opinions. In this case, treatments are twofold; firstly identify if the sentence has expressed or not an opinion, then assess the polarity of opinion. But the main difficulty comes from the fact that objective sentences can be carrying opinion.

- **Aspect level analysis**: This level performs a finer analysis and requires the use of natural language processing. In this level, opinion is characterized by a polarity and a target of opinion. In this case, treatments are twofold: first identify the entity and aspects of the entity in question, and then assess the opinion on each aspect

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 [16]

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 know 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 calculate 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. [16]

The sentiment analysis can be divided into different categories as shown in Figure 1
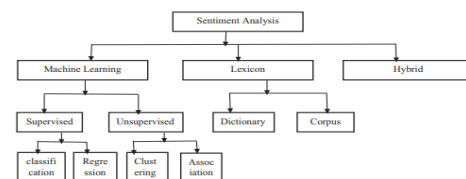
B.) Overview:
Nowadays and since the last two decades, digital data is generated on a massive scale, this phenomenon is known as Big Data (BD). This phenomenon supposes a change in the way of managing and drawing conclusions from data. Moreover, techniques and methods used in artificial intelligence 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 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]
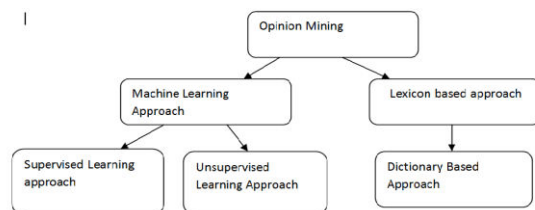
Fig 2.1 Opinion Mining Techniques

➢ 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.* [13] 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.
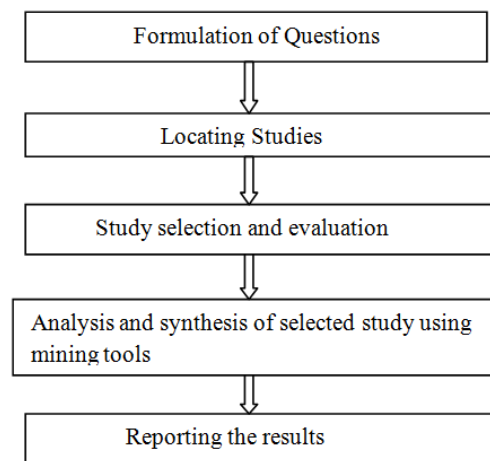

Fig :Research methodology of systematic literature review

1.QUESTION FORMULATION :

A deep and insightful literature review should start with the development of a clear understanding about your objectives [12]. Therefore, to ascertain this, we clearly formulated and considered research questions to evade doubts in our study [13]. 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.
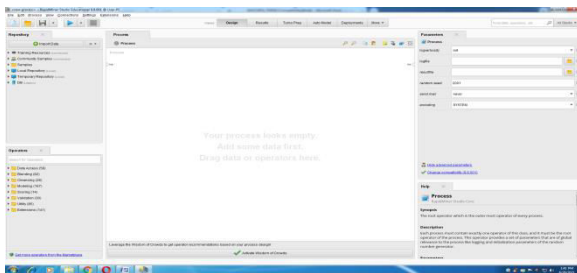
Rapid Miner studio is visual workflow designer, which is used for predictive analysis. This combines machine learning and big data science for the better analysis. Understanding of the rapid miner studio is important to design the process diagram. The IDE (Integrated Development Environment) has repository, 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 contribute 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[27] [28].
Explicit Features of the Rapid Miner tool is explained in the below :[27][28]

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, Bivariate 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 verse 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.

Meaning Cloud

The Sentiment Analysis tool is one of the products of meaning cloud. Sentiment Analysis Application Programming Interface (API) analyzes the context by identifying individual phrases and evaluating the relationship between them. Some of the features are global sentiment, detection of irony, difference in sentiment, agreement and disagreement between analyzed text contents

Unstructured text hides huge amount of important information, but it is hard to process it automatically. Meaning Cloud's extension for Rapid Miner enable you to give it a structure, extract its meaning and combine it with other data sources in your favorite text analytics platform.

Meaning Cloud's extension for Rapid Miner provides new operators with advanced text analytics functionalities:[29][28]

- **Topic Extraction**: extracts from the text names of people, organizations, brands or places, abstract concepts, and amounts.
- **Text Classification**: categorizes a text according to predefined taxonomies.
- **Sentiment Analysis and Aspect-based Sentiment Analysis**: detects the positive/negative/neutral polarity expressed in the text both at a global and at aspect-level.
- **Lemmatization**: extracts a list of the lemmas of the words found in the text.
- **Deep Categorization**: assigns one or more categories to a text, using a very detailed rule-based language.

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 [28]. 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 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 :
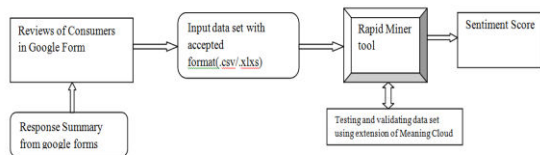

Methodology Diagram for Proposed Framework

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.

. Before that the Pre-processing methods are used to the acceptable data-set. Pre-processing includes the transform the given case into acceptable case, Text Vectorization, process documents, and cross validation is done .Some of the extension like meaning cloud for sentiment analysis and aspect based sentiment analysis are also helpful to build the model. The data model is a construction of the algorithm(**Support Vector Clustering** (SVC) algorithm for cross validation of dataset and clustering using the *k-means* algorithm)with specified data set. Finally, **Extract aggregates operator** calculates descriptive features (e.g. sum, mean, min, max, ...) of the distribution of the values of one or more time series . The calculated features are

provided as an Example Set at the *features* output port of the operator,based on the output of the operator the sentiment score of Sentiment Analysis (also known as Opinion Mining) & Aspect Based Sentiment Analysis and can be scored, the **Correlation matrix** is calculated for sentiments which determines correlation between all Attributes and it can produce a weights vector based on these correlations

5. Experiment and Performance Analysis :[29]

Rapid Miner is a great tool which has filled with text processing capabilities and third party Application Programming Interface (API) also easy to connect to this tool. The below figure explains the process diagram for the sentiment analysis how covid-19 changed consumers daily lives
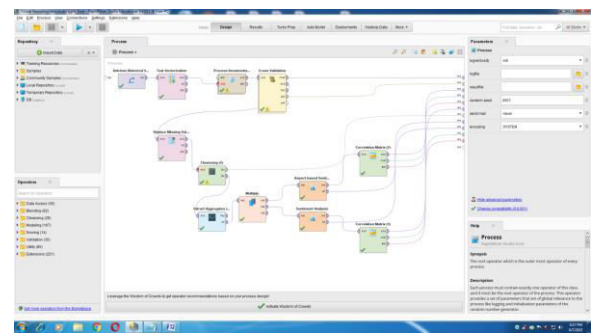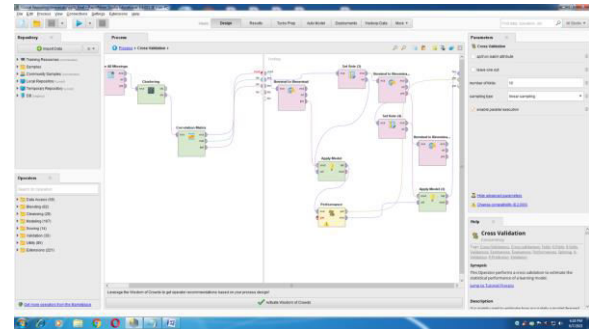

Rapid Miner sentiment analysis is starts from **Retrieve Operator** which import the excel sheet need to analysis. This Operator can access stored information in the Repository and load them into the Process. Retrieving data this way also provides the Meta data of the Rapid Miner Object. The Example set is having 1010 examples with 3 special attributes and 14 regular attributes

The next step is performing the **text vectorization** then extracting/ genertaing the word vectors from string attribute in **process documents from data** for performing a cross validation to estimate the statistical performance of a learning model.
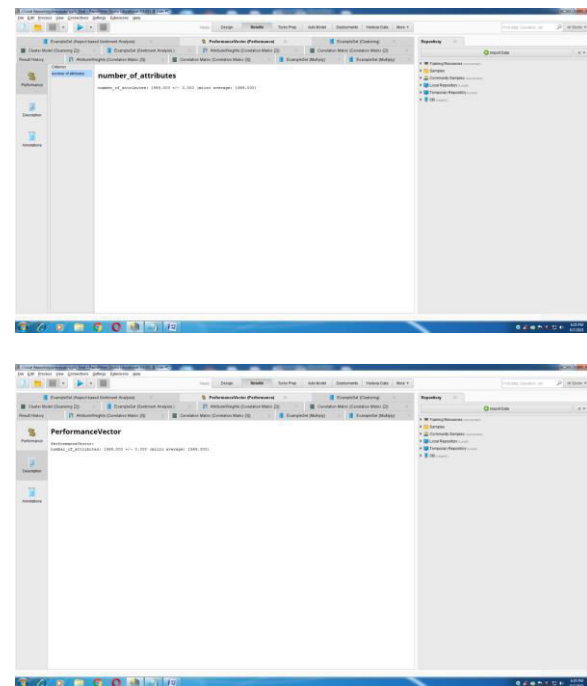
The **Cross Validation Operator** is a nested Operator. It has two sub processes: a Training sub process and a Testing sub process. The Training sub process is used for training a model. The trained model is then applied in the Testing subprocess. The performance of the model is measured during the Testing phase. The input dataset is partitioned

into k subsets of equal size. Of the k subsets, a single subset is retained as the test data set (i.e. input of the Testing sub process). The remaining $k - 1$ subsets are used as training data set (i.e. input of the Training sub process). The cross validation process is then repeated k times, with each of the k subsets used exactly once as the test data. The k results from the k iterations are averaged (or otherwise combined) to produce a single estimation. The value k can be adjusted using the *number of folds* parameter.
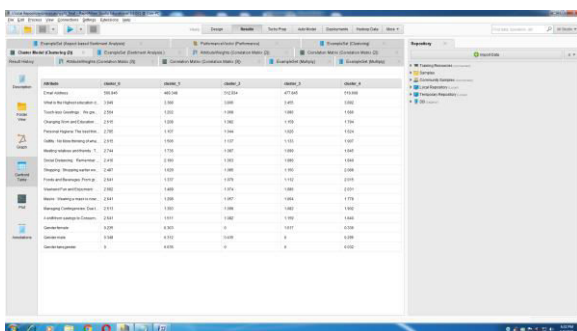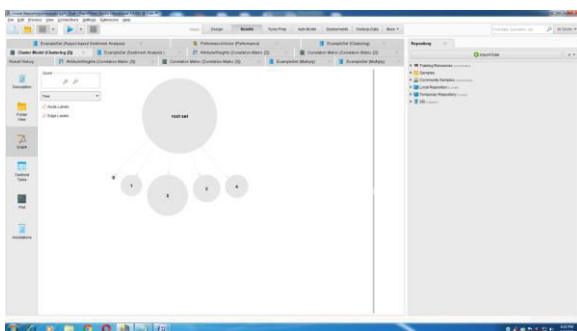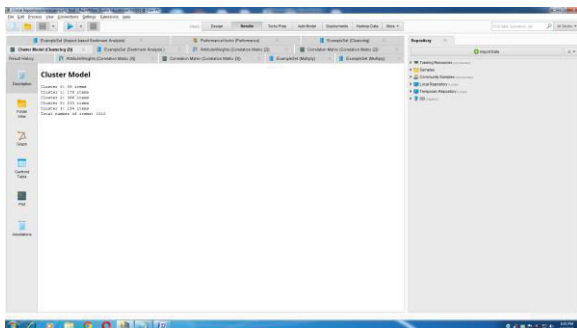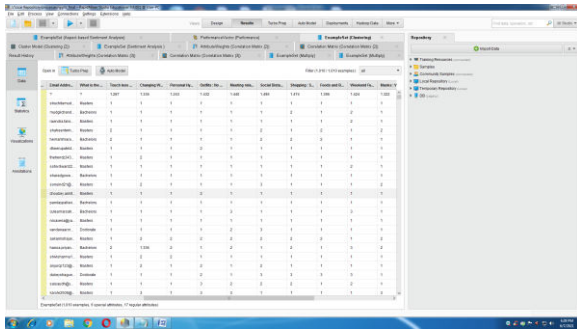

The evaluation of the performance of a model on independent test sets yields a good estimation of the performance on unseen data sets. It also shows if 'over fitting' occurs. This means that the model represents the testing data very well, but it does not generalize well for new data. Thus, the performance can be much worse on test data.

Cross validation Process then perform the clustering using support vectors:

In this meth **Support Vector Clustering** (SVC) algorithm data points are mapped from data space to a high dimensional feature space using a Gaussian kernel. In feature space the smallest sphere that encloses the image of the data is searched. This sphere is mapped back to data space, where it forms a set of contours which enclose the data points. These contours are interpreted as cluster boundaries. Points enclosed by each separate contour are associated with the same cluster. As the width parameter of the Gaussian kernel is decreased, the number of disconnected contours in data space increases, leading to an increasing number of clusters. Since the contours can be interpreted as delineating the support of the underlying probability distribution, this algorithm can be viewed as one identifying valleys in this probability distribution.
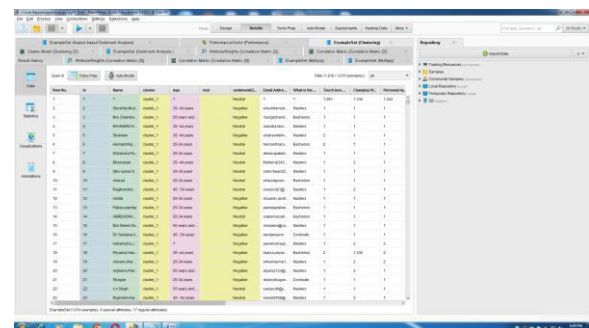
Clustering is concerned with grouping together objects that are similar to each other and dissimilar to the objects belonging to other clusters. It is a technique for extracting information from unlabeled data and can be very useful in many different scenarios e.g. in a marketing application we may be interested in finding clusters of customers with similar buying behavior.

Result of Cluster data set :

After clustering of objects is done **Correlation Matrix Operator** determines 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

Secondly Operator performs clustering using the *k-means* algorithm. Clustering groups Examples together which are similar to each other. As no *Label* Attribute is necessary, Clustering can be used on unlabelled data and is an algorithm of unsupervised machine learning.

The **k-means algorithm** determines a set of *k* clusters and assignes each Examples to exact one cluster. The clusters consist of similar Examples. The similarity between Examples is based on a distance measure between them.

A **cluster in the k-means algorithm** is determined by the position of the center in the n-dimensional space of the n Attributes of the ExampleSet. This position is called centroid. It can, but do not have to be the position of an Example of the ExampleSets.

The k-means algorithm starts with *k* points which are treated as the centroid of *k* potential clusters. These start points are either the position of *k* randomly drawn Examples of the input ExampleSet, or are determined by the k-means++ heuristic if *determine good start values* is set to true.

All Examples are assigned to their nearest cluster (nearest is defined by the *measure type*). Next the centroids of the clusters are recalculated by averaging over all Examples of one cluster. The previous steps are repeated for the new centroids until the the centroids no longer move or *max optimization steps* is reached. Be aware that it is not ensured that the k-means algorithm converges if the measure type is not based on Euclidean Distance calculation (cause the recalculation of the centroids by averaging is assuming Euclidean space). The procedure is repeated *max runs* times with each time a different set of start points. The set of clusters is delivered which has the minimal sum of squared distances of all Examples to their corresponding centroids.

Next, **Extract aggregates operator** calculates descriptive features (e.g. sum, mean, min, max, ...) of the distribution of the values of one or more time series. The calculated features are provided as an data Set at the *features* output port of the operator.

Depending on the parameter *add time series name* the data Set will have one example with attributes for all combination of time series and features, or n examples, one example per time series. The features be calculated can be

selected individually. In combination with the Process Windows operator, this operator can be used to calculate features of windows of time series as a preparation for a general machine learning problem.

By default invalid values (missing, positive infinity and negative infinity) are included in the calculation of the aggregated values. See the description of the parameters on how the calculation of the individual features handle invalid values. Select the parameter *ignore invalid values* to change this and ignore invalid values. This operator works only on numerical time series.

Lastly after determining the Sentiment Analysis (also known as Opinion Mining) & Aspect Based Sentiment Analysis ,consists of the application of natural language processing, text analytics and computational linguistics to identify and extract subjective information from various types of content. It identifies the positive, negative, neutral polarity in collected unlabelled data
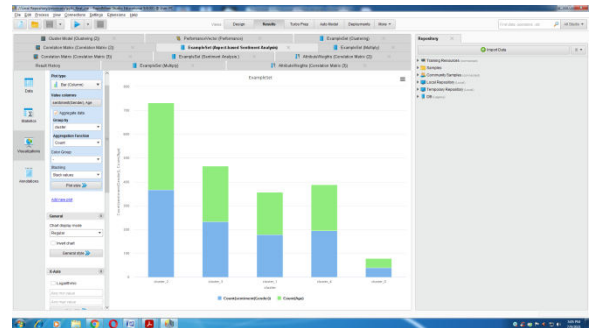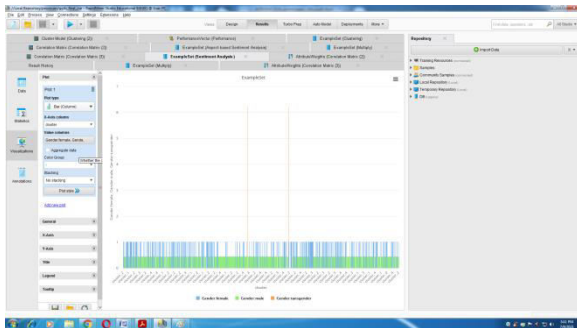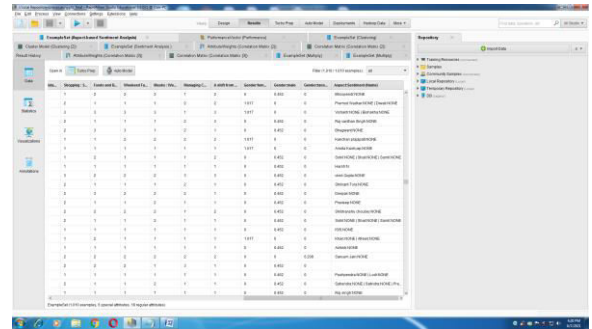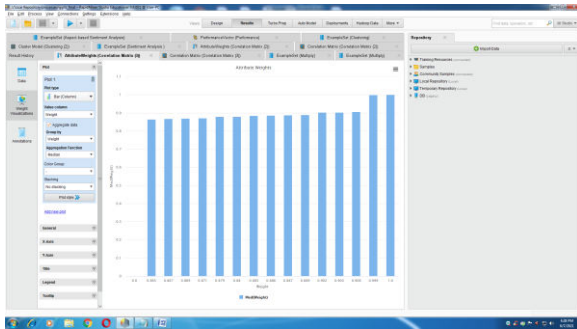
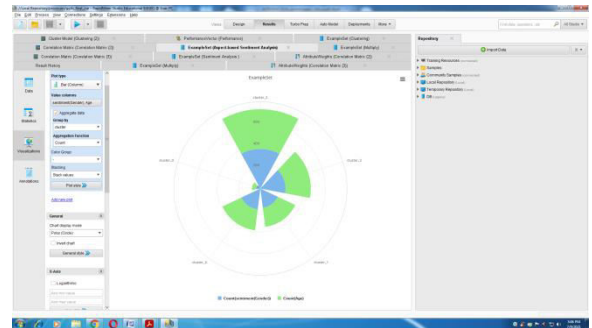**Result obtained of Sentiment Analysis**:
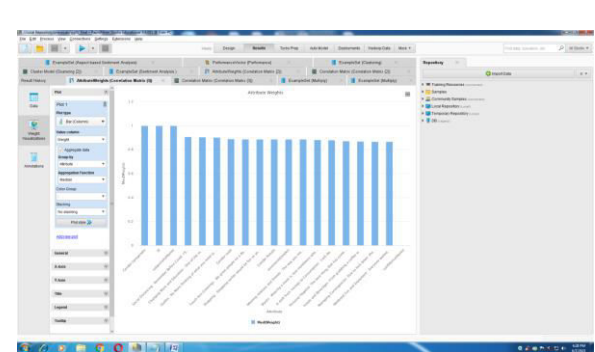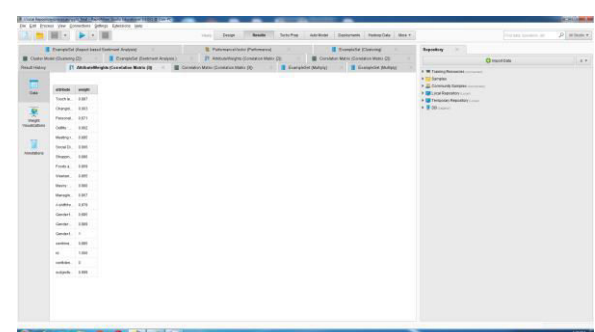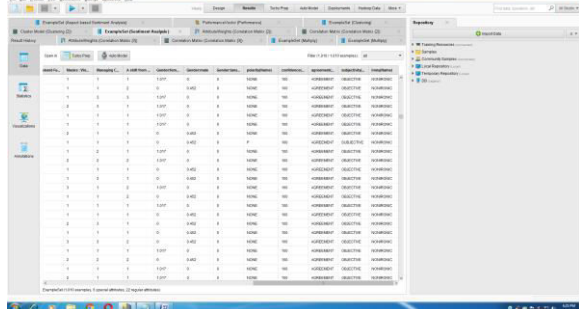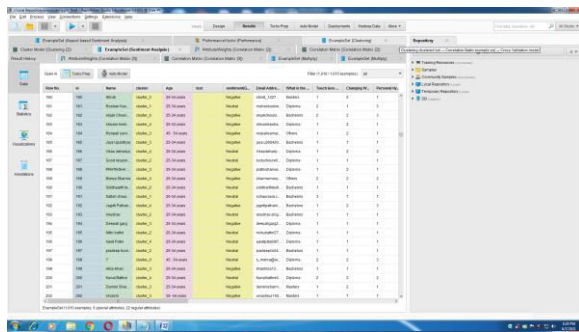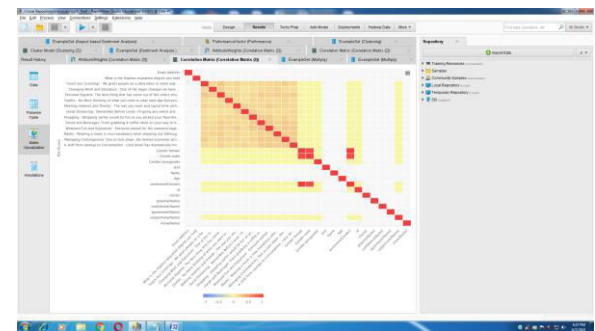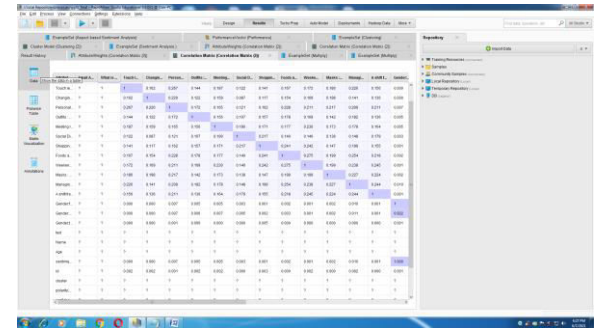
Correlation matrix(sentiment analysis):

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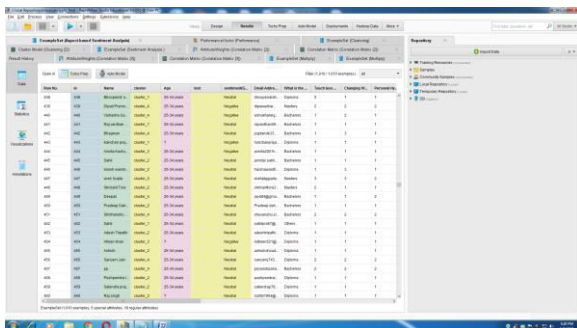
produce a weights vector based on these correlations. Correlation is a statistical technique that can show whether and how strongly pairs of Attributes are related.

Index	Nominal value	Absolute count	Fraction
1	Neutral	651	0.645
2	Negative	359	0.355

Result obtained of Aspect based Sentiment Analysis :

Index	Nominal value	Absolute count	Fraction
1		112	0.111
2	Subham:NONE | Shub...	5	0.005
3	Ashish:NONE	4	0.004
4	Harsh:N	4	0.004
5	Satyam:NONE	4	0.004
6	Shivani:NONE	4	0.004
7	Singh:NONE	4	0.004
8	Vivek:NONE	4	0.004

Correlation matrix(Aspect based sentiment analysis)

The **Correlation matrix** is calculated for Aspect based sentiments which determines correlation between all

Attributes and it can produce a weights vector based on these   correlations..
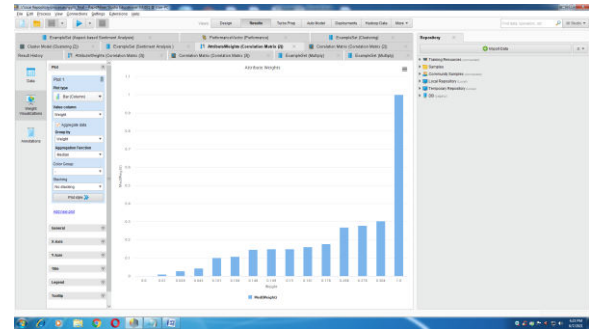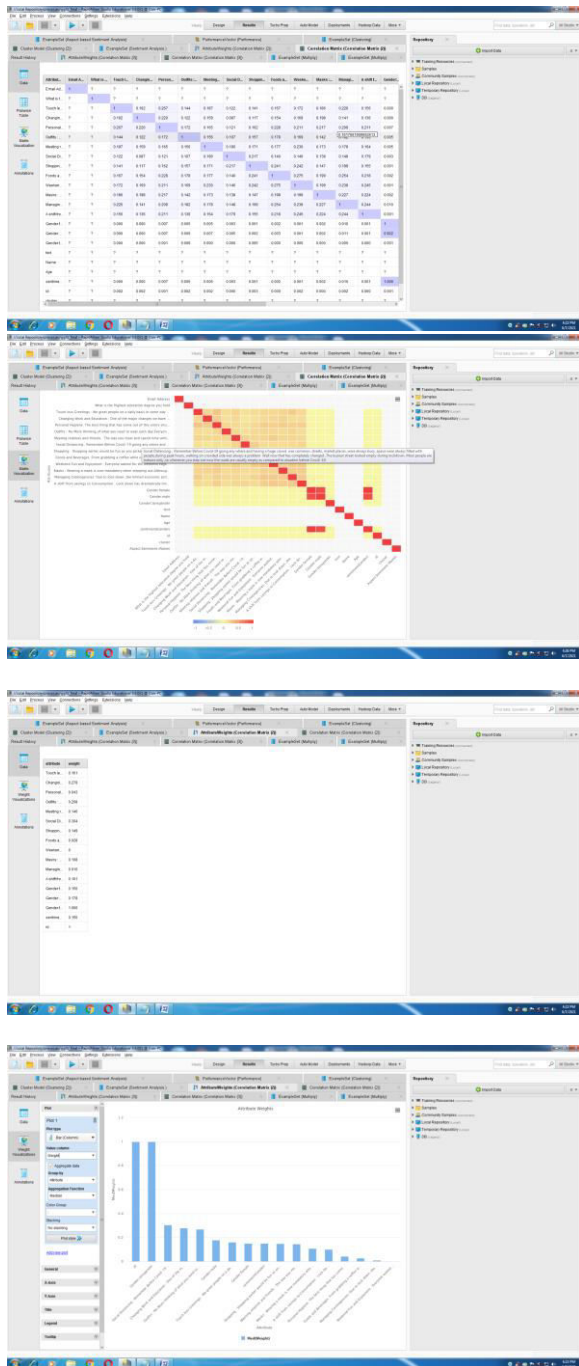
6. Discussion and Conclusion:

The main aim of this experiment was to find the sentiments of the individual during the situation of covid19 , the polarity of unsupervised data collected and clustering is done by using k-means algorithm is calculated .The total data collected from the response of google forms is 1010 examples which was imported in the Rapid miner is one of the best tools which have the advance features like data exploration, sampling, replacing, partitioning, Bayesian modeling, clustering and modeling evaluation. Using the rapid miner tool with two different machine learning algorithm the result gives accuracy to the sentiments of consumers where the data set was of 1010 example set with6 special attributes and 22 regular attributes, so the sentiment calculated for nominal values is maximum neutral with 651 absolute count and 359 with negative absolute count and rest is positive . Cluster 2 is found with highest aspect based sentiment polarity both gender and age wise. Name wise Aspect based sentiment polarity with 112 absolute count.

In terms of applications, more research has been done on the evaluation or evaluation of the various methods of opinion mining and sentiment analysis. Although these refer to the assessment of the techniques used, the data sets extracted from users' application databases thus include an aspect of human application. Marketing-related activities still lead the applications followed by the financial, healthcare, and hospitality and tourism industries. It is further noted that applications of opinion mining and sentiment analysis for politics and government views are still increasing. Etc..

Refrences:

[1A] 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

[1B] 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

[2] 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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges,SHAHID SHAYAA et al, ,IEEE Acess ,Received April 24, 2018, accepted June 18, 2018, date of publication June 28, 2018, date of current version July 30, 2018.

[10] BigData Concepts and Architecture Journal of Computer Science and Information Systems, Volume 1. Issue 4, August 2020 ISSN 2535-1451

[11] A Survey on Sentiment Analysis for Big Data Swati Sharma1, Mamta Bansal2, Ankur Kaushik3 IJARSE ,Vol no.6 Issue no.06, june 2017

[12] D. Denyer and D. Tran_eld, ``Producing a systematic review,'' in The Sage Handbook of Organizational Research Methods. Thousand Oaks, CA, USA: Sage, 2009.

[13] 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.

[14] R. J. Light and D. B. Pillemer, Summing Up: The Science of Reviewing Research. Cambridge, MA, USA: Harvard Univ. Press, 1984.

[15] 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.

[16] B. Usharani ,"Analysis of Supervised and Unsupervised Learning Classifiers for Online Sentiment Analysis "Asian Journal of Computer Science Engineering 2018;3(4):17-21

[17]. Tsytsarau M, Palpanas T. Survey on mining subjective data on the web. Data Min Knowl
 Discov 2012;24:478-514.

[18]. Liu B. Sentiment Analysis and Opinion Mining. San Rafael: Morgan and Claypool Publishers; 2012. p. 1-168.

[19]. Pang B, Lee L. Opinion mining and sentiment anlaysis. Found Trends Inf Retr 2008;2:1-135.

[20]. Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis.
 IEEE Intell Syst 2013;28:15-21.

[21]. Feldman R. Techniques and applications for sentiment analysis. Commun ACM 2013;56:82-9.

[22]. Montoyo A, Barco PM, Balahur A. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. Decis Support Syst 2012;53:675-9.

[23]. Li N, Wu DD. Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decis Support Syst 2010;48:354-68

[24] Balahur A. Sentiment Analysis in Social Media Texts. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; 2013. p. 120-8

[25]. Nakagawa T, Inui K, Kurohashi S. Dependency Tree Based Sentiment Classification Using CRFs with Hidden Varaibles, Human Language Technologies: The 2010 Annual Conference of the North America Chapter of ACL; 2010. p. 786-94.

[26]. Moilanen K, Pulman S. Sentiment Composition. The Oxford Computational Linguistics Group. Proceedings of RANLP; 2007. P
[27]https://docs.rapidminer.com/latest/studio/getting-started/

[28] 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

[29]https://www.meaningcloud.com/developer/rapidminer-extension

[30] R. J. Light and D. B. Pillemer, Summing Up: The Science of Reviewing Research. Cambridge, MA, USA: Harvard Univ. Press, 1984