IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12743

Sentiment Analysis of Consumer Post Covid Utilizing Quick Mining Approaches

Amit Kashyap[1], Sushma Kushwaha[2]

Research Scholar[1], Department of Computer Science & Engineering,

Swami Vivekanand College of Science and Technology[1]

Assistant Prof , Department of Computer Science & Engineering,

Swami Vivekanand College of Science and Technology[2]

Abstract: Consumers and families are challenged by the Post-COVID-19 to maintain a healthy lifestyle, as unhealthful behaviors raise the mortality risk. In this investigation, we look at how the prevalent Corona virus has affected a wide range of consumer attitudes, convictions, and behavior. A variety of client data has been gathered using sentiment analysis. Additionally, utilizing a variety of quick mining methods, current breakthroughs in machine learning algorithms have enhanced sentiment analysis estimates on lifestyles. While machine learning automates the creation of logical models, a perspective's semantic orientation determines whether it is positive, negative, or neutral. This research focuses on the sentiment analysis of lifestyles utilizing quick mining approaches, classifying their polarity as good, negative, or neutral. To estimate attitudes, machine learning employs methods such as Support Vector Regression and K-means clustering.

Keywords*:* Machine Learning, Big Data, Sentiment Analysis, Support vector Regression (SVR), K-means

INTRODUCTION

People's lives and workplaces have changed significantly as a result of COVID-19, and social involvement has dropped. Sentiment analysis calculates, identifies, and categorizes reader views in text to assist determine the correctness of such opinions. To extract subjective information, it makes use of text analysis, computational linguistics, biometrics, and natural language processing. Sentiment analysis reveals the emotional communication and effect of a reader, revealing whether they have a Positive, Negative, or neutral impression of a product.

According to Shuchita et al.[1][2], aspect-level analysis, sentence-level analysis, and document-level analysis are all included in sentiment mining. AI has influenced new analytical approaches like sentiment analysis and opinion mining, while big data has had a huge impact on data management and decision-making. Managing this volume becomes more difficult as internet access rises, but machine learning algorithms can assist people and businesses in managing and profiting from user data.

Machine learning is a field of technology that teaches computers to carry out a variety of tasks, including predictions, recommendations, guesses, etc., based on previous experience or historical data. Through the use of projected data and prior experience, machine learning trains computers to act like people.

LITERATURE REVIEW

The following three categories serve as the main divisions of machine learning techniques [3][4][5]

1. Supervised learning

It is possible to use supervised learning when a machine has sample data, or input and output data with accurate labels. Using a few labels and tags, the model's accuracy is checked to ensure that it uses the proper labels. Utilizing our prior knowledge and labeled samples, supervised learning techniques enable us to anticipate future events. Prior to introducing an inferred function that forecasts output values, it first analyzes the known training dataset. Additionally, it foresees mistakes during the entire learning process and uses algorithms to remedy them.

Let's use an example where we have a collection of pictures with the tag "dog." These dog photos are used to train a machine learning algorithm so that it can quickly determine if an image is of a dog or not.

2. Unsupervised Learning

Unsupervised learning involves training a computer with only a small subset of input samples or labels, with no knowledge of the final product. As a result, a machine may not always produce the right results compared to supervised

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023
DOI: 10.17148/IJARCCE.2023.12743

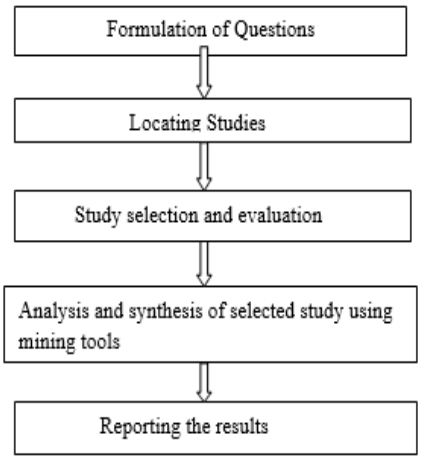
learning because the training data is neither categorized nor labeled. Though less common in real-world business contexts, unsupervised learning aids in data exploration and can be used to infer latent structures from unlabeled data.
Example: Assume that a machine has been educated with a set of papers belonging to several categories (Type A, B, and C), and that we now need to classify them according to their appropriateness. The machine can classify these datasets into types A, B, and C because it is simply given input samples or no output, so it is not crucial whether the classification is correct or not.

3. Reinforcement Learning

A machine learning technique that relies on feedback is reinforcement learning. Agents (computer programs) are required to investigate their surroundings, take actions, and then receive feedback and rewards as a result of their behaviors. They receive rewards for both good and bad deeds, with positive rewards going to good deeds. The maximization of positive rewards is the objective of a reinforcement learning agent. The agent can only learn from its experience because there isn't any labeled data available.
Machine learning is becoming increasingly important in today's digital age, as it allows computers to process vast amounts of data quickly and efficiently. With increased memory handling capabilities and computational powers, machine learning has applications in various sectors, such as healthcare, fraud detection, sentiment analysis, and e-commerce. For instance, OLA and Uber use machine learning models to predict surge pricing and provide real-time pricing based on factors like demand, availability, weather, and rush hours. This approach helps drivers manage demand and minimize surge prices, ensuring efficient and timely service[6][7]

METHODOLOGY

This work attempts to understand opinion mining and sentiment analysis techniques in text analytics with a focus on socioeconomic conditions. It employs a systematic examination of the literature that comprises a clearly stated problem, relevant publications, evaluations, and summaries. For ongoing study to yield new insights, current research must be combined.[8]
Combining data from the body of existing literature to produce unique insights in ongoing inquiry is as important as conducting original research. Systematic reviews are emphasized by Rousseau et al. [9] as being thorough, open, objective, and accurate in their analysis. The objectivity and repeatability of outcomes are the goals of these reviews.
Finding, evaluating, synthesizing, coming to conclusions, and reporting findings from sentiment analysis and opinion mining literature are all steps in the five-step investigative approach.[10][11] A five-step investigative Processisshown:



A. Formulation of The Question

To remove uncertainty, a good literature review needs goal-setting and the consideration of research issues. Finding the applicability of sentiment analysis and opinion mining approaches for particular individuals or groups is the focus of this study. [2][3][9]

B. Studies On Locations

It is vital to find appropriate academic journal articles in order to compile a list of all the papers that are relevant to our study issues. I've built a simple database using a few of thoughtfully chosen questions. The main components of the study are sentiment analysis and opinion mining, so we used alternate strings to find relevant literature.[12][13].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12743

However, a method was employed in this study to highlight both the good and negative emotions.[14][15].Data science is carried out using Rapid Miner 10.01, an analytics and AI tool. Text mining is also included in this application.

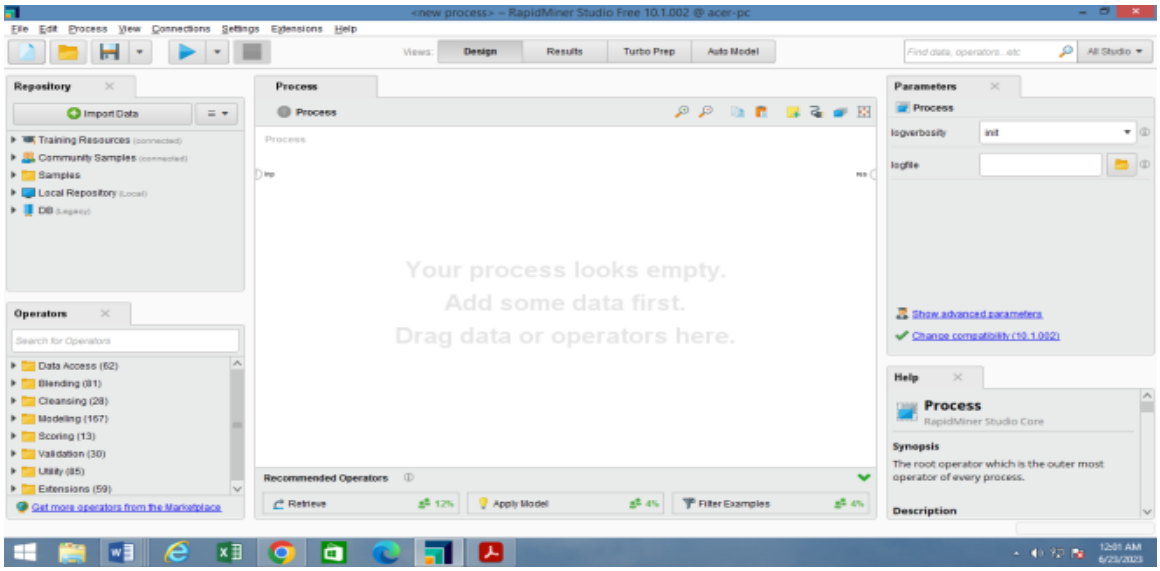


Fig 1. Visual Workflow for Rapid Miner Studio

Rapid Miner Studio is a powerful visual workflow builder for predictive analytics that combines big data science with machine learning for better analysis. It allows users to create and implement analytics processes, presentations, and models, transforming uncluttered data into valuable information. Rapid Miner supports data access and management using SAS, ARFF, Stata, and URLs. It offers data exploration tools, data prep, Turbo, and more. Meaning Cloud's Sentiment Analysis tool simplifies unstructured text processing by analyzing context, identifying specific phrases and offering topic extraction, text classification, aspect-based sentiment analysis, lemmatization, and deep categorization using rule-based language.[16][17][18]

C. Selection and Evaluation

We choose the articles from peer-reviewed journals for quality control and systematic procedures, concentrating on sentiment analysis and opinion mining. To ensure objectivity and enrich the research, an impartial reader reviewed the publications. Research questions were used to guide the analysis and synthesis of articles. For a few papers, taxonomy was made that included datasets, methods, applications, and key difficulties. [2][3][17].

DATASETS

1010 responses were gathered for a 12-question survey that was used to analyze data from Google articles using Google forms.

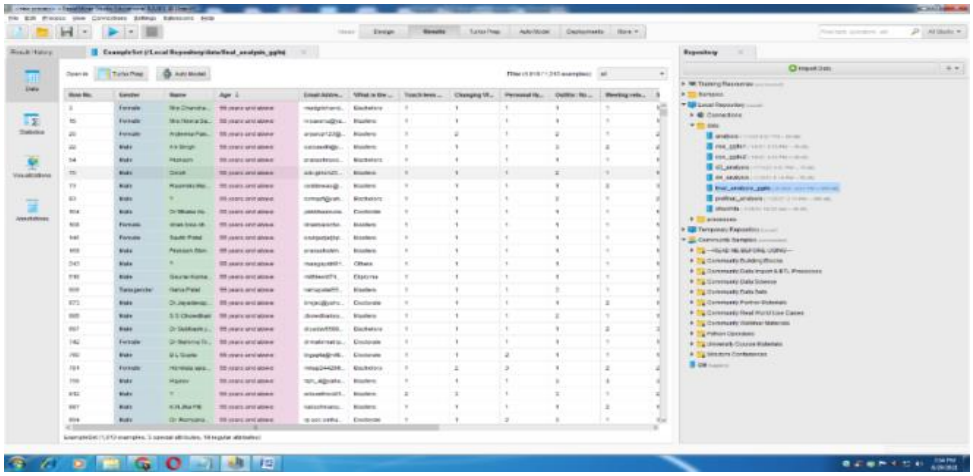


Fig 2. Data Set Analysis

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12743

A. Method and Proposed framework

The rapid mining approach in statistical text mining involves loading data, preprocessing it, creating models, and applying them to predict outcomes using a term-by-document matrix. Data is collected through Google forms, reviewed, and downloaded, and pre-processed using techniques like text vectorization and polynomial to binomial classification. Extensions like meaning clouds and logistic regression algorithms are used to build models. Extract Aggregates produces descriptive distribution aspects of time series values. Opinion Mining Aspect Based Sentiment Analysis calculates sentiment scores, correlation matrix, and weights vectors based on operator output.

B. Proposed Technique

Sentiment research converts ambiguous online reviews into explicit evaluations based on scores, revealing how post Covid-19 has affected customers' daily life. This method is essential for gathering insightful data and examining consumer opinions. Combining polynomial to binomial classification and sub processing, a multilingual sentiment analysis model is built using Meaning Cloud. The tool recognizes sentiment, polarity, information and viewpoints, and allows users to create their own sentiment models.

Experiment and Performance Analysis

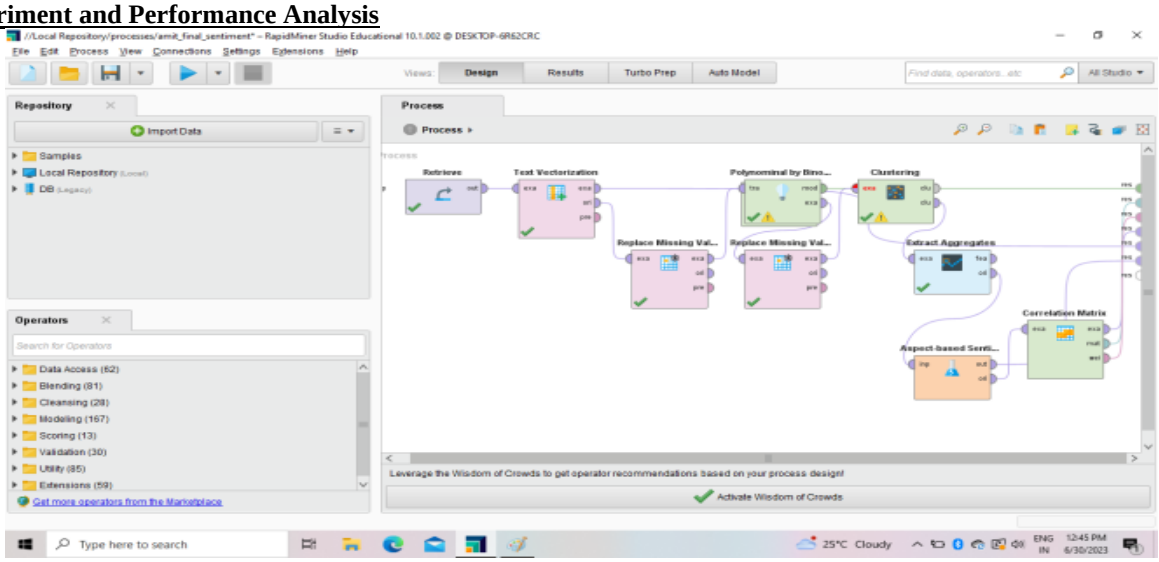


Fig 3. Methodology model for proposed frame work

The proposed framework uses Rapid Miner's Retrieve Operator and Meaning Cloud's solution for multilingual sentiment analysis of clustered texts. The model imports an Excel sheet, loads data from the repository, and performs text vectorization. The results show 651 neutral and 359 negative sentiments in the sample. The methodology model utilizes information to carry out the analysis.

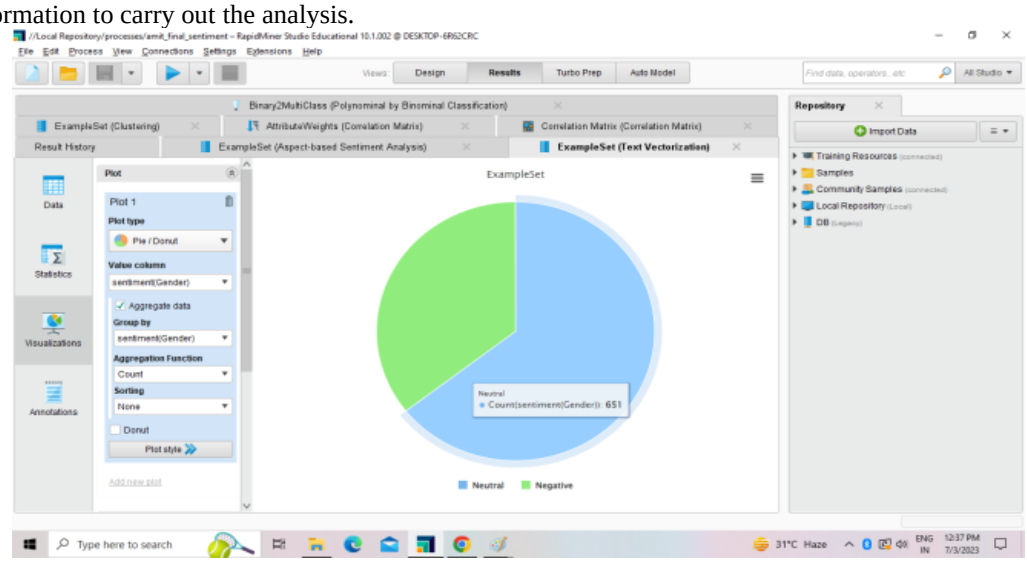


Fig 4. Text Vectorization

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12743

Data is transformed using a nested operator, converting polynomials to binomial classification. Numeric values are converted, and regression analysis uses the support vector machine (SVM) technique. The logistic regression algorithm identifies the hyper plane best fitting data points, while the SVR technique focuses on the hyper plane that traverses the most data points within a predetermined margin.

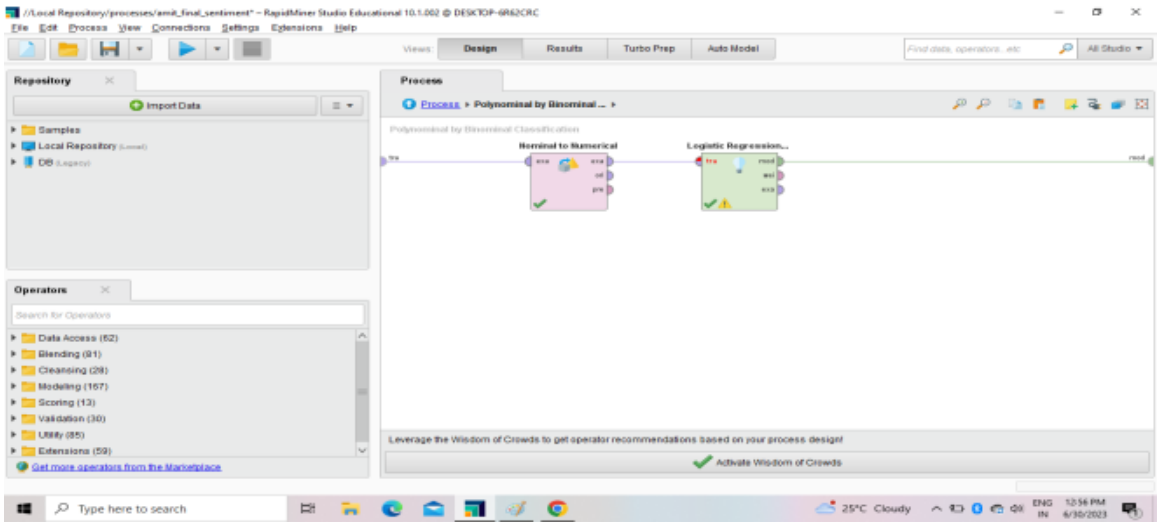


Fig 5. Sub process model for polynomial to binomial conversion

SVR uses kernel function to handle non-linear relationships, reducing prediction error and making it effective for complex assignments requiring regression in input-target variables.

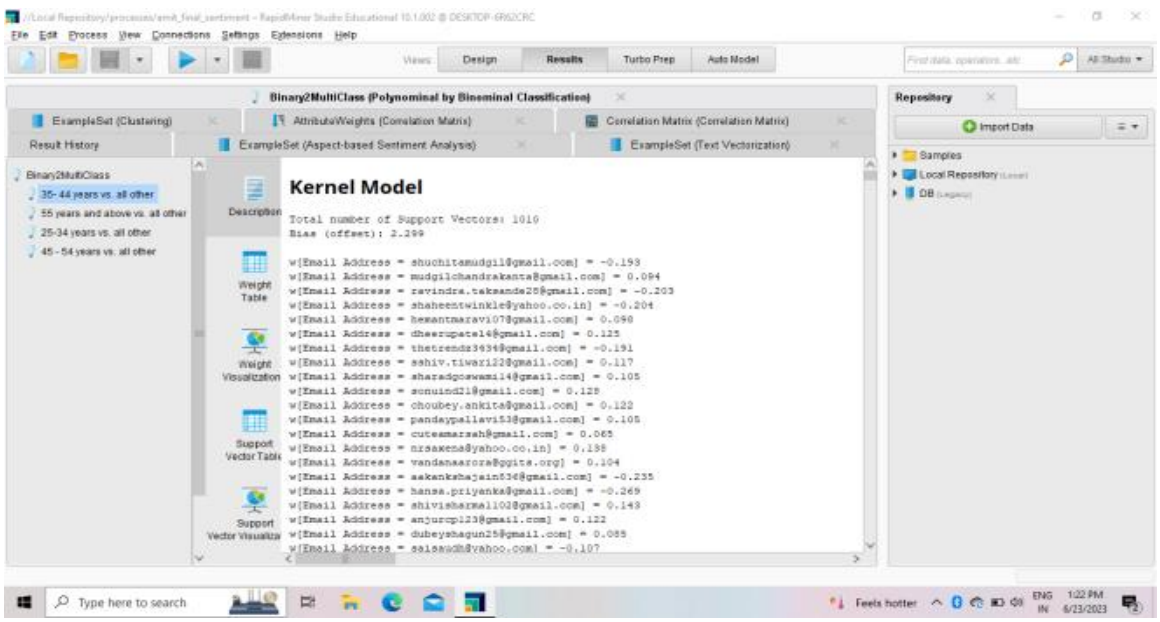


Fig 6. Kernal data Model

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12743

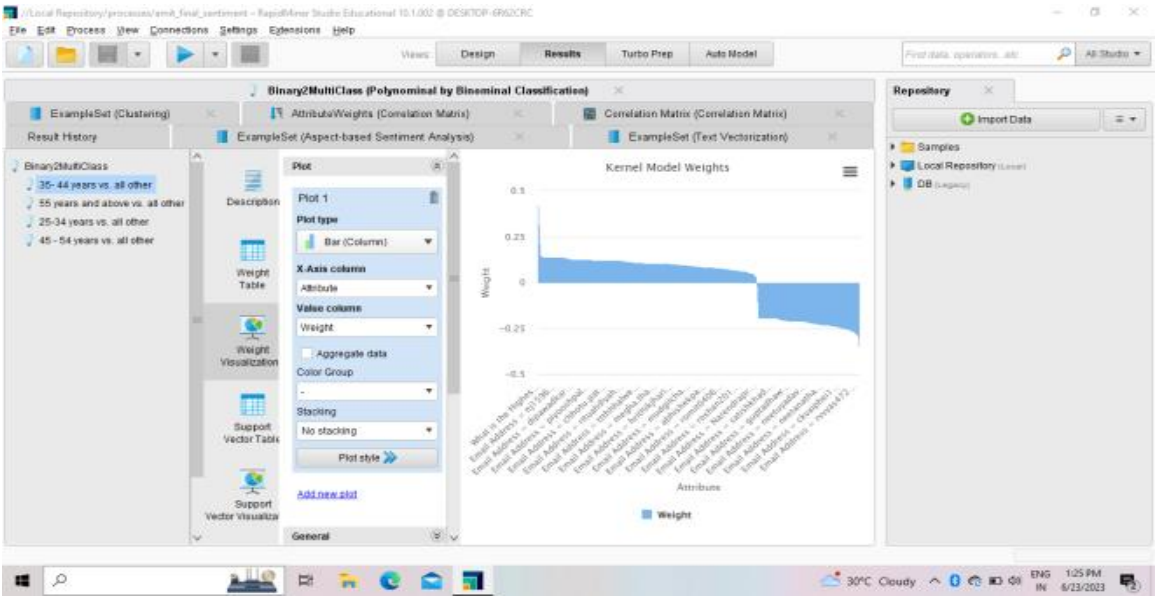


Fig 7. Polynomial to binomial classification result

Clustering uses polynomial to binomial classification for efficient data extraction in marketing applications.

Cluster data set results

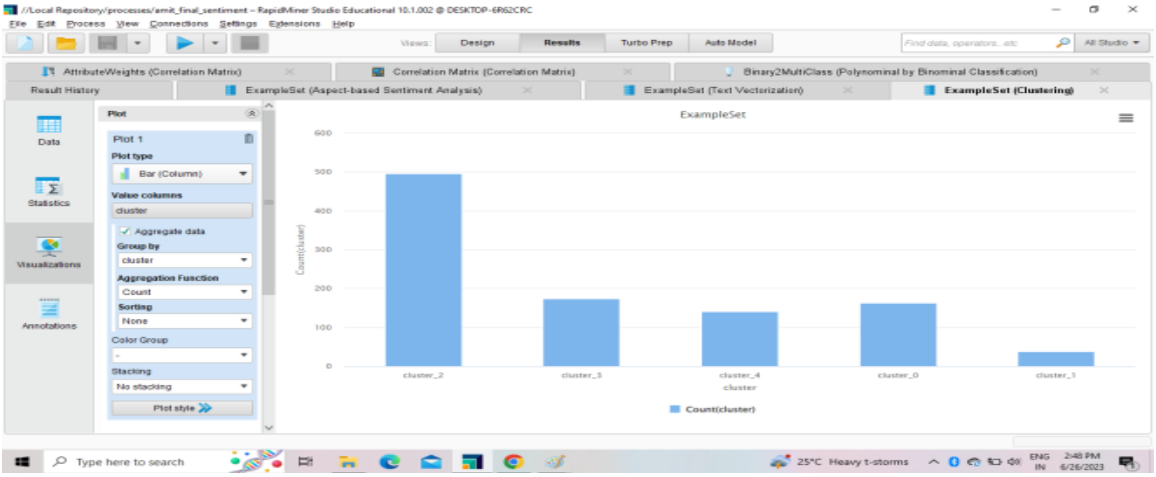


Fig 8. Cluster data set results Polynomial to Binomial Classification

The Correlation Matrix After the objects have been clustered, the operator analyzes the correlation between their attributes and builds a weight vector based on those correlations. Correlation exists. The operator uses the k-means algorithm to cluster the data after that. Examples that are connected to one another constitute a cluster. Clustering is a machine learning technique that can be applied to unlabeled data because it does not require a Label Attribute.

The k-means algorithm places each sample in a single cluster out of a total of k. The clusters are made up of comparable cases. The center position in the n-dimensional space of the n characteristics determines the K-means cluster. The centroid is the name given to this region. It may or may not be the location of an Example Sets example. The centroid of each of the initial k potential clusters is represented by a set of k points in the k-means algorithm. If the "determine good start values" option is set to true, the k-means++ heuristic or the positions of these nodes will be applied to discover these start points. The k-means++ heuristic or random examples from the input Example Set are used to locate the center of the initial k potential clusters with k points. By averaging over cases, the k-means algorithm determines the cluster centroids, finishing when the centroids stop moving or when the maximum number of optimization steps have been performed. Maybe it won't converge.

The time series data produced by Extract Aggregates include total, mean, min, and max descriptive distributions. This operator has the ability to ignore faulty values and is suitable for machine learning. To extract subjective information from multiple content types, computational linguistics, text analytics, and natural language processing are combined.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12743

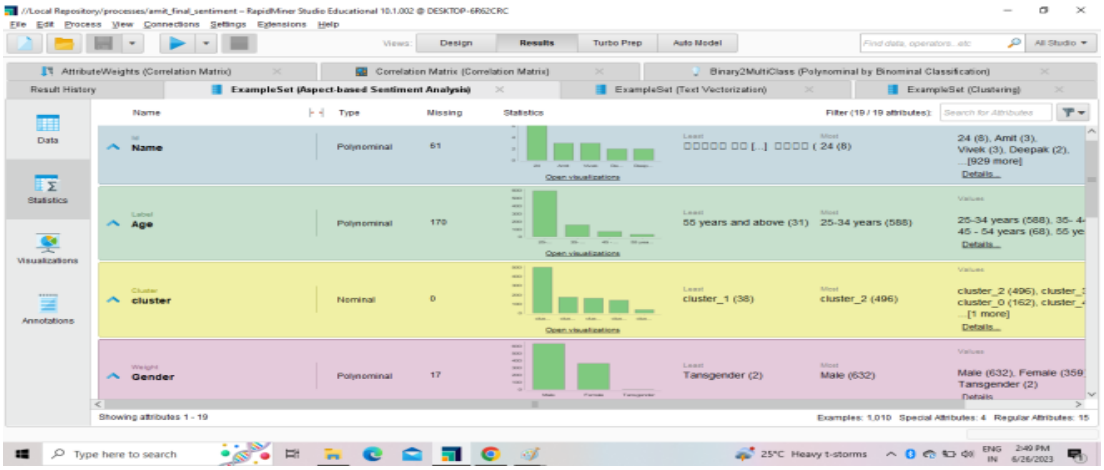


Fig 9a. Aspect-based sentiment analysis results

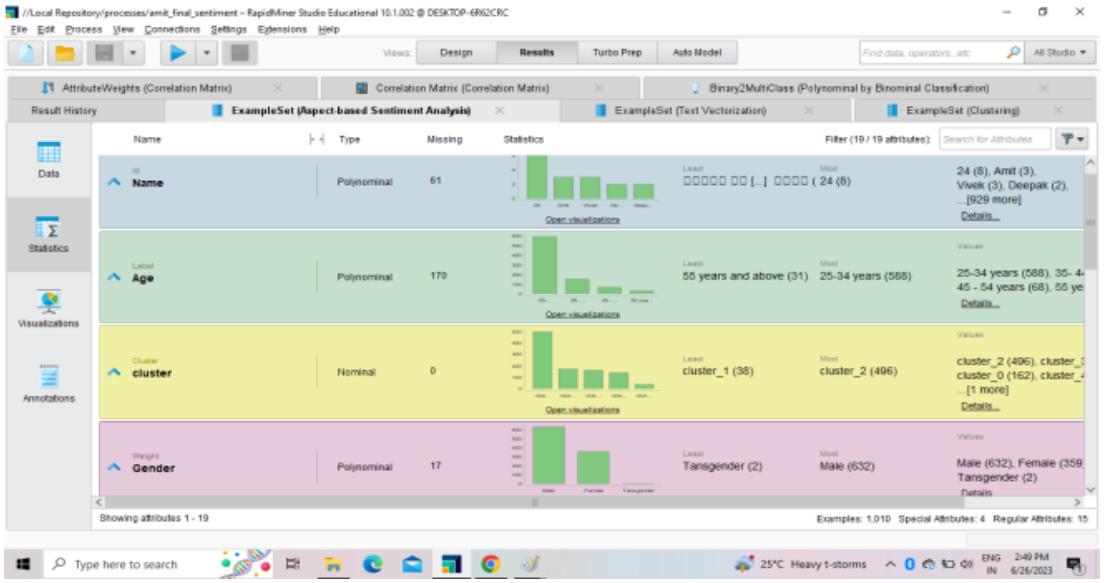


Fig 9b. Aspect-based sentiment analysis results

Correlation matrix for aspect sentiment analysis Correlation matrix calculates sentiment correlations, generating weights vectors based on these correlations, determining attribute relationships.

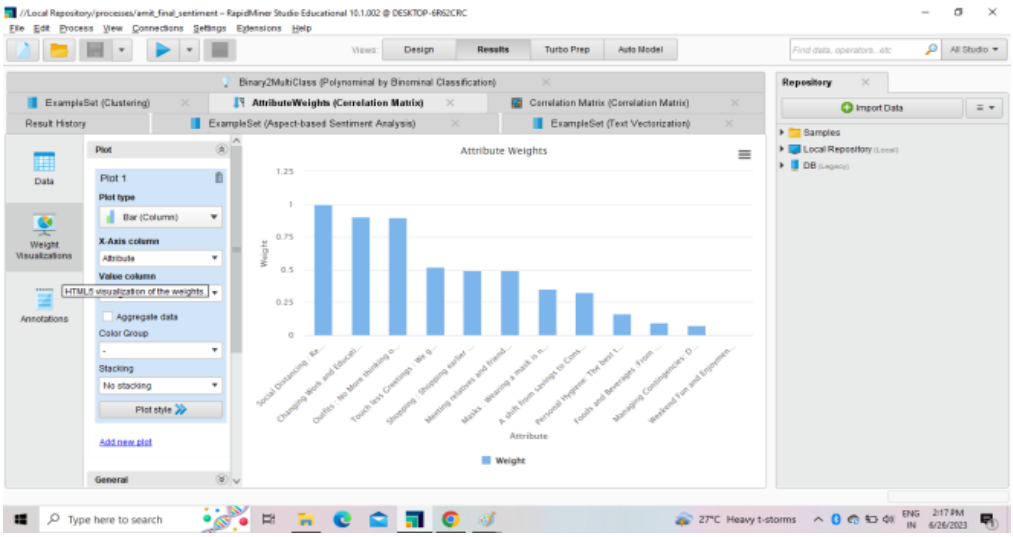


Fig 10a. Correlation matrix results

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12743

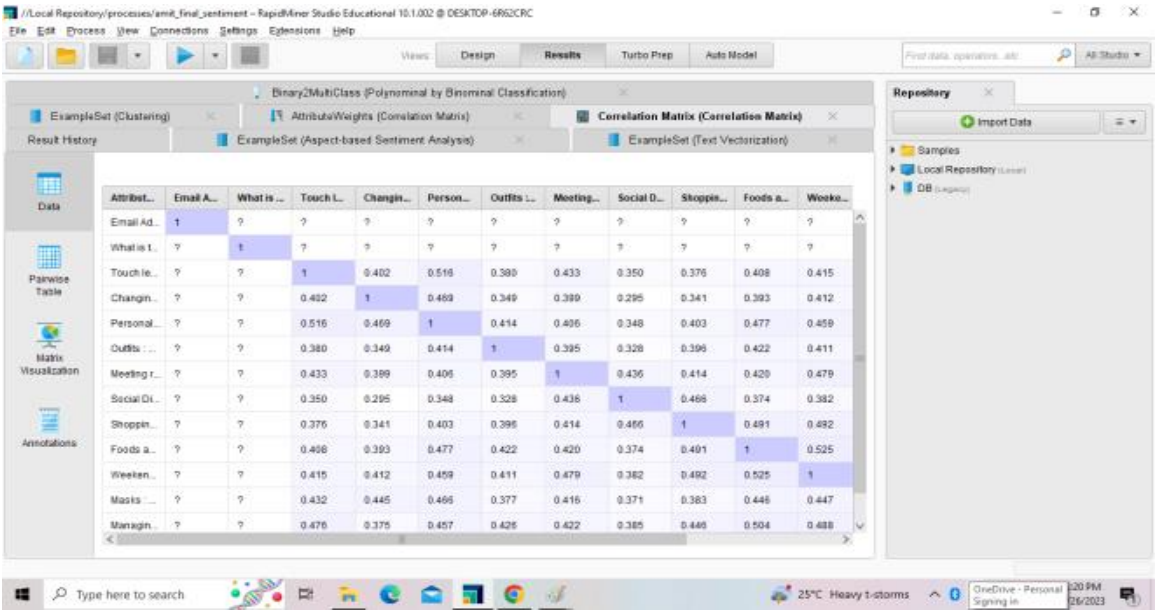


Fig 10b. Correlation matrix results

Fig 10c. Correlation matrix results

DISCUSSION AND CONCLUSION

This study used Rapid Miner's k-means method to analyze unsupervised data on consumer attitudes after COVID-19 pandemic. With 1010 samples and 6 unique features, the data showed mostly positive sentiments. The technology can be applied in various industries, such as filmmaking, financial services, medical care, and tourism, to gauge customer responses and address issues like mental health support, relocation, and internet safety.

FUTURE WORK

Technology can be applied to address various problems, such as mental health issues, education sector and international migration. The internet provides a secure haven for those facing difficulties, and the government's deployment of this technology could make it a safer place for everyone.

REFERENCES

1. Mudgil Shuchita, Concept of sentiment analysis using machine learning techniques , international journal of
2. 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)

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023
DOI: 10.17148/IJARCCE.2023.12743

3. Mudgil, Shuchita & Verma, Ashok. (2021). Sentiment Analysis How Covid-19 Changed Consumers Daily Lives. International Journal of Scientific Research & Management Studies. Volume 05. 13
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. 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. 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, 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, 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. 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.
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 2009. Thousand Oaks, CA, USA: Sage,.
13. B. Usharani ,"Analysis of Supervised and Unsupervised Learning Classifiers for Online Sentiment Analysis "Asian Journal of Computer Science Engineering 2018
14. Tsytsarau M, Palpanas T. Survey on mining subjective data on the web. Data Min KnowlDiscov 2012;
15. Liu B. Sentiment Analysis and Opinion Mining. San Rafael: Morgan and Claypool Publishers; 2012. p. 1-168.
16. Pang B, Lee L. Opinion mining and sentiment anlaysis. Found Trends InfRetr 2008
17. https://docs.rapidminer.com/latest/studio/getting-started/
18. 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
19. https://www.meaningcloud.com/developer/rapidminer-extension
20. Balahur A. Sentiment Analysis in Social Media Texts. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; 2013.
21. 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.
22. Moilanen K, Pulman S. Sentiment Composition. The Oxford Computational Linguistics Group. Proceedings of RANLP; 2007.
23. Feldman R. Techniques and applications for sentiment analysis. Commun ACM 2013;
24. 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
25. Li N, Wu DD. Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decis Support Syst 2010
26. R. J. Light and D. B. Pillemer, Summing Up: The Science of Reviewing Research. Cambridge, MA, USA: Harvard Univ. Press, 1984