MULTIMODAL SENTIMENT ANALYSIS

Akshada Kothavale[1], Khushi Sahoo[2], Kedar Awale[3], Dnyanesh Kunkulol[4] , Priyanka Khalate[5]

[1,2,3,4]Student, Dept. of Computer Engineering, S.K.N.C.O.E. Vadgaon, Pune, Maharashtra, India

([5]Professor, Dept. of Computer Engineering, S.K.N.C.O.E. Vadgaon, Pune, Maharashtra, India)

Abstract - ***In this article, we provide a novel story of multimodal sentiment analysis that includes the collection of sentiments from Online videos using a model that incorporates audio, visual, and textual modalities as knowledge resources. We employed feature- of new neural network architectures, including as autoencoder networks, convolutional neural networks (CNNs), deep belief networks (DBNs), and memory-enhanced neural network models such as long short-term memory (LSTM) models, have lately acquired prominence.***

Keywords — Sentiment Analysis, Multimodal Sentiment Analysis, Multimodal Fusion, Human Computer Interaction

I. MOTIVATION

While making judgements, it is critical that we consider the perspectives of people near to us. This group was formerly modest, with only a few trustworthy friends and family members. But, with the introduction of the Internet, we have seen people express their opinions on blogs and forums. People searching for information about a certain entity now actively read these (product, movie etc.). As a result, there are several points of view available on the internet. Gathering client thoughts on a certain entity is critical from the customer's standpoint. Because there is so much information accessible, it is difficult for customers to go through it all to understand the consensus users. As a result, a system that distinguishes between positive and negative ratings is required. Additionally, classifying these articles based on their emotion would provide readers with a rapid assessment of the general view of an entity. Customers now have a venue to communicate their brand experiences and ideas, whether favorable or unfavorable, on any product or service, owing to the emergence of Web 2.0 platforms such as blogs, discussion forums, and so on. According to Pang and Lee (2008), these customer voices have a substantial influence on how other customers view a brand, how they choose to make purchases, and how they advocate for their own brands

II. INTRODUCTION

Individualism and sentiment analysis are self-regulating perceptions of one's own brain condition (opinions, emotions, behaviors). Distinct perception is well defined in determining whether or not the data is subjective. In this case, sentiment analysis divides data into positive, negative, and neutral categories and thereby finds the data's sentiment polarity. Until far, the majority of sentiment analysis research has focused on NLP [natural language processing]. The current dataset and sentiment analysis sources are confined to text-based sentiment analysis. People are increasingly using social media applications on a large scale to express their opinions, thanks to the introduction of multi-media platforms. As a result, mining views and identifying feelings from many modalities is critical. Possible link to the diversity of input Perhaps connected to the range of input modalities are modalities. Multimodal real-time media analysis is only now gaining popularity and attention. The timbre of the human voice, fluctuations in the subject's face expression over time, eye and lip movement, and dynamic multimodal analysis are all significantly more potent than static multimodal analysis. This paper proposes a method for automated real-time sentiment analysis that is useful in retail. It employs microscopic neural network-based modules to dynamically anticipate the emotional content of an incoming video stream into three categories: positive, neutral, and negative. Despite the fact that people might exhibit a wide range of emotions in a particular situation, we believe that this does not provide a clear indicator to a corporation tracking customer behavior satisfaction. The owner or analyst may desire to know if the client's surprise was positive or negative, for example, if the system indicates that the customer expressed surprise. To do so, we first determine whether a face is visible in the video stream, and if so, we classify the mouth as open or closed. When the mouth is closed, the facial image is used purely for sentiment analysis. Instead, a windowed Fourier transform is utilized to generate a spectrogram from the spoken input. By sending this spectrogram and the face image to different CNN modules, a learned representation of the audiovisual input is generated. At the conclusion, a SoftMax classifier is utilized to categorize the fused representation. An exponentially weighted moving average is used to calculate sentiment accumulation over time and provide the final forecast for a certain time frame. that some forms of more granular emotion corporations utilized the information obtained or mined to explore new possibilities and better target their customers. A large portion of the information era is built on people's views, and firms undertake surveys and opinion polls to learn about their negative and neutral clients It is proposed to employ six emotions: anger, pleasure, contempt, sorrow, fear, and surprise. The advantage of this strategy over the previous one is that it is easier to determine the author's specific feeling, which allows companies to make correct and educated judgements. In our study, we utilize a video as the text source, and we extract sentences from different moments in the film to assess the overall sentiment. R was largely utilized to analyze tweets; text extraction from videos was not considered. These studies may modify the intended message since words can have a range of context-specific meanings that can only be detected when facial expressions and voice are taken into consideration. A decision tree classification is utilized in audio analysis. The dataset is first separated into smaller subsets for categorization, and then a tree is constructed using a regression model. Data collection, training, testing (the backend), and front-end are the four complicated aspects that comprise Multimodal Sentimental Analysis. Unlike training and testing, which must wait until data collection is complete, data collection and the front end may work independently. Pre-processing and data processing are both aspects of data collection. Data pre-processing include splitting created and acquired datasets into training and testing datasets, as well as cleaning such datasets. This dataset slicing for training and testing will be done in a 70:30 ratio.

III. LITERATURE SURVEY

Sentiment analysis (also known as opinion mining) is the process of analyzing and extracting sentiment from textual data, typically in the form of online reviews, social media posts, and customer feedback. Sentiment analysis systems have become increasingly popular in recent years due to the explosion of user-generated content on the web and the growing importance of customer feedback in the business world. In this literature survey, we will explore some of the recent research and development in the field of sentiment analysis systems. The article "A Review of Sentiment Analysis Techniques and Applications" was written by Xia Owen Ding and colleagues (2018) This paper provides a comprehensive survey of various sentiment analysis techniques and applications, including lexicon-based approaches, machine learning-based approaches, and deep learning-based approaches. The authors also discuss the challenges and future directions of sentiment analysis. Sanjay Chawla and colleagues published "A Survey of Deep Learning for Text-based Sentiment Analysis." The use of deep learning algorithms for sentiment analysis is the main topic of this study. The authors give an overview of deep learning models for sentiment analysis, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks. By F. Martin et al., "Sentiment Analysis: A Review and Comparative Study of Online Services" (2018)This paper compares various web-based sentiment analysis services, including both lexicon-based and machine learning-based approaches. The authors evaluate the services based on their accuracy, speed, and ease of use. "A Survey of Sentiment Analysis in Social Media" by Dong Nguyen. (2013) This paper provides a comprehensive survey of sentiment analysis in social media, including challenges such as noisy data, sarcasm, and ambiguity. The authors also discuss the use of sentiment analysis in various applications, such as predicting election outcomes and stock prices. Rasha Ismail and colleagues published "A Survey of Sentiment Analysis Research in Arabic" (2018) This paper focuses on sentiment analysis research in the Arabic language. The authors provide an overview of the challenges of sentiment analysis in Arabic, including the lack of resources and the complexity of the language, and review various approaches to Arabic sentiment analysis. By M. Sudhakar et al., "Sentiment Analysis of Customer Reviews Using Machine Learning Approaches" (2019) This paper presents a machine learning-based approach to sentiment analysis of customer reviews. The authors compare various machine learning algorithms, including Support Vector Machines (SVM), Random Forests (RF), and Naive Bayes (NB), and evaluate their performance on a dataset of customer reviews.

IV. DATA SETS

A. Textual Analysis

The goal of an alternative to topic detection in the realm of sentiment analysis was to extract evaluative meaning. Deep learning may have contributed to the most encouraging gain in text sentiment. Deep learning may have contributed to the most encouraging gain in text sentiment. Deep learning may make use of extremely large datasets to record word embeddings that are relevant for emotion analysis and provide organically expanded vocabulary. Continued research indicated that extrapolating word sentiment consistent variables depending on word embeddings still requires substantial work, despite the creation of word classes relying on deep learning algorithms achieving results extraordinarily close to those of human annotators. Hyperlink to our text sentimental analysis https://www.kaggle.com/datasets/konradb/text-recognition-total-text-dataset

B. Audio Analysis

Nevertheless, it is a similarly new field to target opinion exclusively from verbal utterances. Specifically focusing on the acoustic aspect of language, it is sometimes quite difficult to distinguish between opinion and feeling. Focusing on pitch-related provisions, Maire see discovered that, even without text-based signals, pitch provides information about feeling. The focus of several further works is on sensation analysis only from the text-based material as it appears in the conversation. The method suggested by Costa Pereira et al., for instance, takes a vocally articulated inquiry and retrieves report. Here is the hyperlink to our audio dataset https://zenodo.org/record/1188976#.XA48aC17Q1J

C. Video Analysis

Although there have been related lines of research in vision-based emotion recognition for a long, e.g., directed sentiment analysis by computer vision is still an active area of study. The main research tasks in "visual opinion analysis" revolve with showing, identifying, and using sentiment communicated by facial or precise signals of feeling linked to visual sight and sound. Wangen al. explored descriptor connections coordinated into 12 adjective-modifier word sets more than 100 photos remarked on by 42 individuals. This research is among the first in visual opinion assessment. Here is the hyperlink to our dataset. https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data.

V. METHODOLOGY

The methodology is a relevant structure for research. We have multiple sections that cover the Architecture, Working, and Tools Used in the project. Major papers on the topic are being published more often, and sentiment analysis has become one of the research hotspots in the field of natural language processing. The web application for sentiment analysis's system architecture and fundamental operation are discussed in the companion document. The programmer is made to deliver feelings by texts, sounds, and visuals.

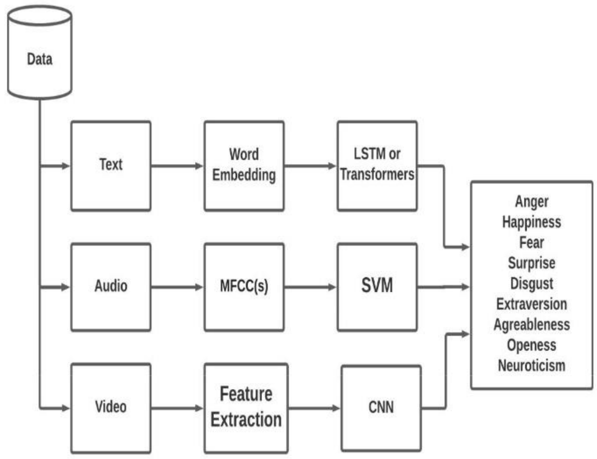
Embedding a word – Word embeddings are dense vectors with a significantly reduced dimensionality. The distance and direction of the vectors, on the other hand, show the semantic associations between words. The Word Embeddings module, which is able to detect the setting of a word in an expression, is given the text data that will be projected for sentimental analysis. relationship with other words, similarity in syntactic and semantic structure.

MFCC – Mel-frequency cepstrum, or MFCC in solid preparation, is a representation of the instantaneous force span of a sound in light of a direct cosine variation of a log power range on a nonlinear mel size of recurrence. Mel-frequency Cepstral Coefficients (MFCC), and synthetically increased MFCC coefficients are compared to classify emotions. dictionally, feature extraction is one of its uses. instance, the distortion in recurrence in the case of sound pressure might take into consideration better sound representation following factors are used to determine MFCCs transform is a symbol. windows with three sides to cover the forces from the range mentioned above, or windows with cosine to encircle the forces. Make a frequency for each meal frequency. identification so an algorithm with a productivity higher than the present algorithm is mandatory. Hybrid technique- based prototype need huge data sets to trail the data file and this procedure also sometimes doesn't categorize the data file so there is a higher risk factor of combining with the unassociated details which will lead to affect the precision ofthe news.

Transformer – Transformer is a catchy technique that teaches the contextual connections between words (or little words) in a text. Transformer combines two separate operating modes: a text input that reads text input and a video that creates a functioning forecast. Only the encoding technique is needed because BERT's main function is to model the language. The graphic that follows shows transformers, also known as sequence-to-sequence architecture. An example of a neural network is sequence-to-sequence architecture. which creates a new grouping by changing a certain sequence of elements, such as the words in a phrase.

VI. PROPOSED SYSTEM

The web application for sentiment analysis's system architecture fundamental operation are discussed in the companion document. With the use of text, voice, and video, the system may convey feelings. A list of emotions that can be predicted using any of the three models is shown at the system's endpoints, along with individual probability for each feeling.



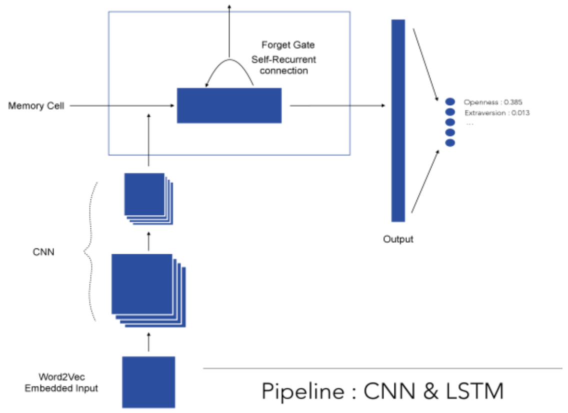
A. Text Analysis

SVM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.

CNN/LSTM

A CNN-LSTM is a model architecture that has a CNN model for the input and an LSTM model to process input time steps processed by the CNN model.



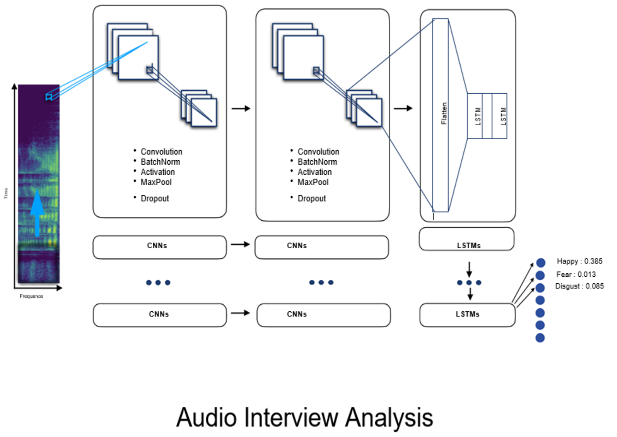
B. Audio Analysis

SVM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.

Time Distributed CNNs

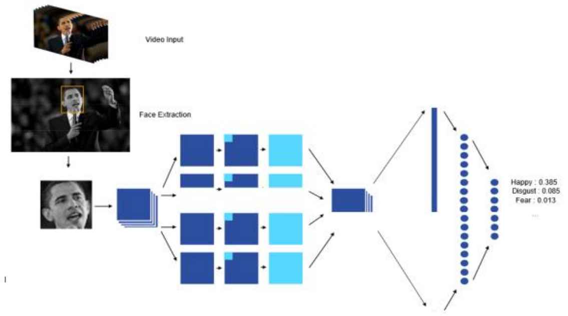
Time Distributed layer is very useful to work with time series data or video frames. It allows to use a layer for each input.



C. Video Analysis

Xception Model

Xception is a convolutional neural network that is 71 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.



Video Interview Analysis

VII. ALGORITHMS AND MODELS

BERT - BERT is an open source machine learning framework for natural language processing (NLP). BERT is intended to assist computers in understanding the meaning of ambiguous words in text by leveraging surrounding material to build context. The BERT framework was pre-trained using Wikipedia text and may be fine-tuned with question and answer datasets. BERT, which stands for Bidirectional Encoder Representations from Transformers, is based on Transformers, a deep learning model in which every output element is linked to every input element and the weightings between them are dynamically determined depending on their relationship. The use of bidirectional training of Transformer, a prominent attention model, to language modelling is BERT's major technological breakthrough. Previously, researchers looked at a text sequence from left to right or a combination of left-to-right and right-to-left training. The findings of the research reveal that bidirectionally trained language models have a better grasp of language context and flow than single-direction language models. The researchers describe a unique approach called Masked LM (MLM) in the publication, which permits bidirectional training in models that were previously unachievable.

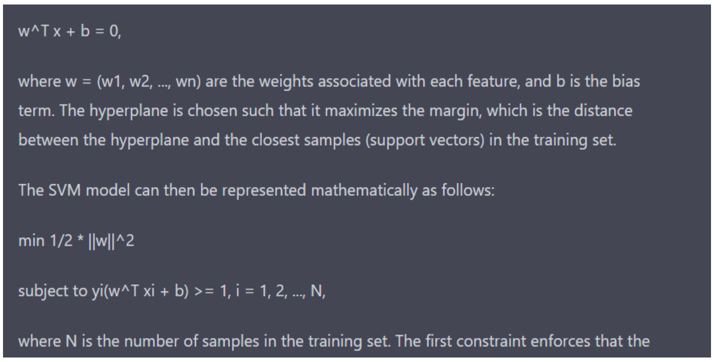
LSTM - LSTM is an abbreviation for long short-term memory networks, which are utilised in Deep Learning. It is a kind of recurrent neural networks (RNNs) that may learn long-term dependencies, particularly in sequence prediction tasks. Apart from single data points such as photos, LSTM contains feedback connections, which means it can process the complete sequence of data. This is used in speech recognition, machine translation, and other areas. LSTM is a kind of RNN that performs exceptionally well on a wide range of tasks. A memory cell known as a 'cell state' that maintains its state throughout time plays the fundamental function in an LSTM model. The horizontal line that goes through the top of the figure below represents the cell state. It may be seen as a conveyor belt on which information just passes, unmodified. In LSTM, information may be added to or withdrawn from the cell state, which is controlled by gates. These gates allow information to flow into and out of the cell. It includes a pointwise multiplication operation as well as a sigmoid neural net layer to help the mechanism.

SVM - Support Vector Machine, or SVM, is a prominent Supervised Learning technique that is used for both classification and regression issues. Nevertheless, it is mostly utilised in Machine Learning for Classification difficulties. The SVM algorithm's purpose is to find the optimum line or decision boundary for categorising n-dimensional space so that we may simply place fresh data points in the proper category in the future. A hyperplane is the optimal choice boundary. SVM selects the extreme points/vectors that aid in the creation of the hyperplane. These extreme examples are referred to as support vectors, and the method is known as the Support Vector Machine. Consider the figure below, in which two distinct categories are categorised. a decision line or hyperplane SVMs are classified into two types: Linear SVM is used for linearly separable data, which implies that if a dataset can be categorised into two classes using a single straight line, it is considered linearly separable data, and the classifier employed is the Linear SVM classifier. Non-linear SVM is used for non-linearly separated data, which implies that if a dataset cannot be categorised using a straight line, it is considered non-linear data, and the classifier employed is the Non-linear SVM classifier.

TRANSFORMERS - A transformer is a deep learning model that uses the self-attention mechanism to weight the relevance of each element of the input data differently. It is largely utilised in natural language processing (NLP) and computer vision (CV) Transformers, like recurrent neural networks (RNNs), are meant to analyse sequential input data, such as natural language, with applications such as translation and text summarization. Transformers, on the other hand, process the full input at once, unlike RNNs. The context is provided by the attention mechanism at any place in the input sequence. If the incoming data is a natural language sentence, for example, the transformer does not have to process one word at a time. This allows for more parallelization than RNNs, resulting in shorter training durations. Transformers were launched in 2017 by a team at Google Brain and are becoming the model of choice for NLP problems, displacing RNN models such as long short-term memory (LSTM). The added training parallelization enables training on bigger datasets. This resulted in the development of pretrained systems such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), which were trained with large language datasets such as the Wikipedia Corpus and Common Crawl and can be fine-tuned for specific tasks.

VIII. MATHEMATICAL MODEL AND ACCURACY

Text Interview Analysis – For text sentiment analysis, the features of the text are first extracted and used to represent the text. For example, a feature could be the frequency of certain words or phrases. These features are then used as inputs to the SVM model. Let's consider a binary classification problem, where the text documents are represented by feature vectors x = (x1, x2, ..., xn) and the sentiment label y is either positive (y = 1) or negative (y = -1). The goal is to find a hyperplane that separates the positive and negative samples where N is the number of samples in the training set. The first constraint enforces that the samples are correctly separated by the hyperplane, and the second constraint maximizes the margin. The optimization problem can be solved using a quadratic programming algorithm. Once the SVM model is trained, it can be used to predict the sentiment of new text documents by computing the value of w^T x + b. If the value is positive, the text is classified as positive, and if the value is negative, the text is classified as negative.



w^T x + b = 0,

where w = (w1, w2, ..., wn) are the weights associated with each feature, and b is the bias term. The hyperplane is chosen such that it maximizes the margin, which is the distance between the hyperplane and the closest samples (support vectors) in the training set.

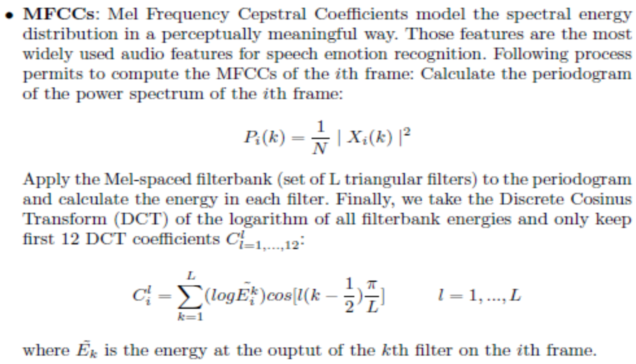
The SVM model can then be represented mathematically as follows:

min 1/2 * ||w||^2

subject to yi(w^T xi + b) >= 1, i = 1, 2, ..., N,

where N is the number of samples in the training set. The first constraint enforces that the

Audio Interview Analysis –



- **MFCCs**: Mel Frequency Cepstral Coefficients model the spectral energy distribution in a perceptually meaningful way. Those features are the most widely used audio features for speech emotion recognition. Following process permits to compute the MFCCs of the ith frame: Calculate the periodogram of the power spectrum of the ith frame:

$$P_i(k) = \frac{1}{N} \mid X_i(k) \mid^2$$

Apply the Mel-spaced filterbank (set of L triangular filters) to the periodogram and calculate the energy in each filter. Finally, we take the Discrete Cosinus Transform (DCT) of the logarithm of all filterbank energies and only keep first 12 DCT coefficients $C^l_{l=1,...,12}$:

$$C^l_i = \sum_{k=1}^{L} (log\bar{E}^k_i) cos[l(k - \frac{1}{2})\frac{\pi}{L}] \qquad l = 1, ..., L$$

where $\bar{E}_k$ is the energy at the ouptut of the kth filter on the ith frame.

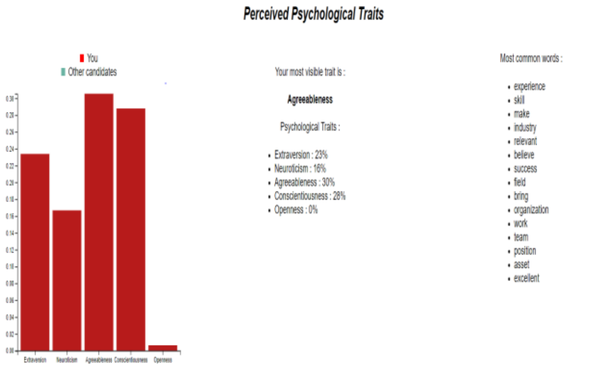
Video Interview Analysis - CNNs are special types of neural networks for processing data with grid-like topology. The ingredients of a convolution neural network are the following: the convolution layer, the activation layer (applying an activation function), the pooling layer the fully connected layer, similar to a dense neural network The order f the layers can be switched : ReLU(MaxPool(Conv(X))) = MaxPool(ReLU(Conv(X)))

IX. RESULTS AND ANAYLSIS

Sentiment analysis, also known as opinion mining, is a computational approach that involves using natural language processing (NLP) techniques to identify and extract subjective information from text data. This can include determining whether a particular piece of text expresses a positive, negative, or neutral sentiment, as well as more nuanced emotions such as happiness, sadness, anger, or fear. The results of a sentiment analysis system can vary depending on a range of factors, including the quality and size of the data set used for training the model, the specific NLP techniques and algorithms employed, and the accuracy and reliability of the annotation process used to label the data. However, when properly implemented and validated, sentiment analysis can provide a powerful tool for understanding public opinion and sentiment on a wide range of topics, from product reviews to social media conversations to political discourse. One common approach to sentiment analysis is to use machine learning algorithms such as support vector machines (SVMs) or deep neural networks to classify text data based on pre-labeled examples. These models are typically trained on large data sets of annotated text, which have been manually labeled by humans with positive, negative, or neutral sentiment scores .One challenge in sentiment analysis is the need to handle ambiguity and sarcasm, which can make it difficult for machines to accurately classify sentiment based solely on the literal meaning of the text. To address this issue, some researchers have explored more advanced techniques such as deep learning models that can capture more subtle and complex patterns in language use. Overall, sentiment analysis has proven to be a valuable tool for a wide range of applications, including marketing research, customer feedback analysis, political polling, and brand management. However, it is important to carefully evaluate the accuracy and reliability of sentiment analysis systems before relying on them for decision-making purposes, as errors and biases can lead to incorrect or misleading conclusions.

OUTPUTS :

TEXT :



Your most visible trait is :

Agreeableness

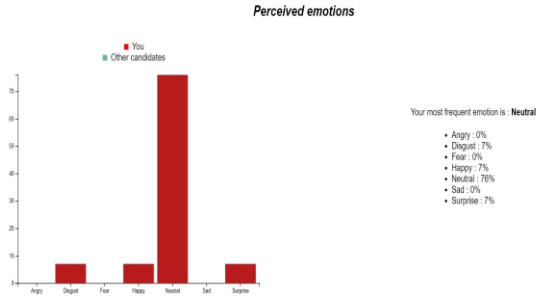
Psychological Traits :

- Extraversion : 23%
- Neuroticism : 16%
- Agreeableness : 30%
- Conscientiousness : 28%
- Openness : 0%

Most common words :

- experience
- skill
- make
- industry
- relevant
- believe
- success
- field
- bring
- organization
- work
- team
- position
- asset
- excellent

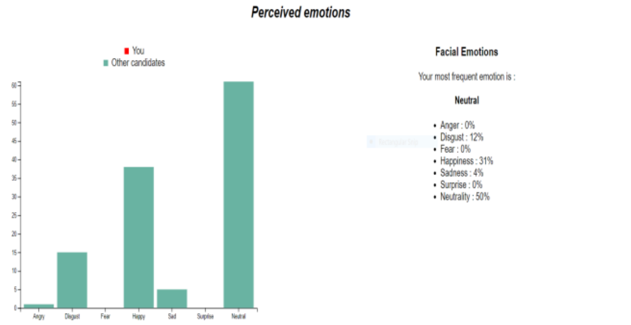
AUDIO :



Your most frequent emotion is : **Neutral**

- Angry : 0%
- Disgust : 7%
- Fear : 0%
- Happy : 7%
- Neutral : 76%
- Sad : 0%
- Surprise : 7%

VIDEO :

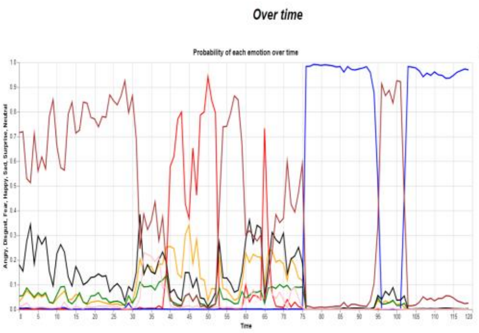


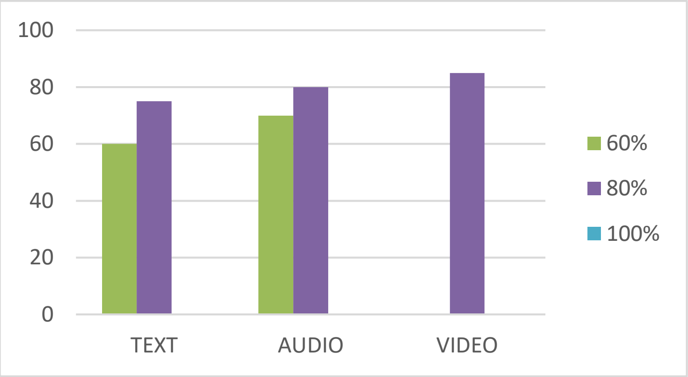
Facial Emotions

Your most frequent emotion is :

Neutral

- Anger : 0%
- Disgust : 12%
- Fear : 0%
- Happiness : 31%
- Sadness : 4%
- Surprise : 0%
- Neutrality : 50%





In this bar graph chart the first bar represents the accuracy of the existing system and the second graph presents the accuracy of the proposed system.

X. CONCLUSION

In conclusion, sentiment analysis is an important area of research that has applications in various domains, such as social media analysis, marketing, and customer service. The purpose of sentiment analysis is to automatically identify the sentiment expressed in text data, whether it is positive, negative, or neutral. Sentiment analysis systems can use different methods and techniques, such as rule-based methods, machine learning, and deep learning. Each method has its advantages and limitations, and the choice of method will depend on the specific application and data set. Overall, sentiment analysis systems have shown promising results in identifying the sentiment expressed in text data. However, there is still room for improvement, particularly in areas such as identifying sarcasm, irony, and ambiguity in text data. Despite the challenges, sentiment analysis has the potential to provide valuable insights into the opinions and attitudes of people towards different topics and issues. As such, sentiment analysis is likely to remain an active area of research and development, with new techniques and applications being developed in the future.

XI. REFERENCES

[1] L.-P. Morency, R. Mihalcea, P. Doshi, Towards multimodal sentiment analysis: harvesting opinions from the web, In: Proceedings of the 13th International Conference on Multimodal Interfaces, ACM, Alicante, Spain, 2011, pp. 169–176.

[2] E. Cambria, N. Howard, J. Hsu, A. Hussain, Sentic blending: scalable multimodal fusion for continuous interpretation of semantics and sentics, In: IEEE SSCI, Singapore, 2013, pp. 108–117.

[3] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1) (2006) 489–501.

[4] G.-B. Huang, E. Cambria, K.-A. Toh, B. Widrow, Z. Xu, New trends of learning in computational intelligence, IEEE Comput. Intelli. Mag. 10 (2) (2015) 16–17.

[5] G. Huang, G.-B. Huang, S. Song, K. You, Trends in extreme learning machines: a review, Neural Netw. 61 (2015) 32–48.

[6] S. Decherchi, P. Gastaldo, R. Zunino, E. Cambria, J. Redi, Circular-ELM for the reduced-reference assessment of perceived image quality, Neurocomputing 102 (2013) 78–89.

[7] E. Cambria, P. Gastaldo, F. Bisio, R. Zunino, An ELM-based model for affective analogical reasoning, Neurocomputing 149 (2015) 443–455.

[8] E. Principi, S. Squartini, E. Cambria, F. Piazza, Acoustic template-matching for automatic emergency state detection: an ELM based algorithm, Neurocomputing 149 (2015) 426–434.

[9] S. Poria, E. Cambria, A. Hussain, G.-B. Huang, Towards an intelligent framework for multimodal affective data analysis, Neural Netw. 63 (2015) 104–116.

[10] H. Qi, X. Wang, S.S. Iyengar, K. Chakrabarty, Multisensor data fusion in distributed sensor networks using mobile agents, In: Proceedings of 5th International Conference on Information Fusion, 2001, pp. 11–16

[11] Ekman, Paul, Friesen, Wallace V, O'Sullivan, Maureen, Chan, Anthony, Diacoyanni-Tarlatzis, Irene, Heider, Karl, Krause, Rainer, LeCompte, William Ayhan and Pitcairn, Tom and Ricci-Bitti, Pio E and others, Universals and cultural differences in the judgments of facial expressions of emotion. J. person. soc. psychol. 53 (4) (1987) 712–717.

[12] D. Matsumoto, More evidence for the universality of a contempt expression, Motiv. Emot. 16 (4) (1992) 363–3968.

[13] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, Palo Alto, 1978.

[14] W.V. Friesen, P. Ekman, Emfacs-7: Emotional Facial Action Coding System, Unpublished manuscript, University of California at San Francisco 2.

[15] A. Lanitis, C.J. Taylor, T.F. Cootes, A unified approach to coding and interpreting face images, In: Fifth International Conference on Computer Vision, 1995. Proceedings .