MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
A Constituent Institution of Manipal University

Department of Computer Science and Engineering

Artificial Intelligence and Machine Learning

A mini project report on

Sentiment Recognition using Machine Learning

By

Moulik Singh

210962130

Rohit Sahay

210962066

Kumar Raunak

210962172

towards qualitative assessment for the course

Artificial Intelligence – CSE 2271

Sentiment Recognition using Machine Learning

Moulik Singh, 210962130, **moulik.singh@learner.manipal.edu**

Rohit Sahay, 210962066, **rohit.sahay1@learner.manipal.edu**

Kumar Raunak, 210962172, **kumar.raunak@learner.manipal.edu**

Abstract :

Sentiment analysis is a crucial area of research that aims to improve the communication between humans and machines by enabling machines to understand and analyse human emotions, opinions, and attitudes. It involves the use of natural language processing, machine learning, and statistical techniques to identify, extract, and classify subjective information from source material, such as social media posts, customer reviews, and feedback forms. The primary objective of this paper is to develop a machine learning model that can accurately predict sentiment by following a structured approach. The first step is data collection, which involves gathering relevant text data from various sources. The next step is text preparation, which involves cleaning and preprocessing the text data to remove noise and irrelevant information.The third step is sentiment detection, which involves using machine learning algorithms to identify the polarity of the text data (positive, negative, or neutral). The fourth step is sentiment classification, which involves grouping the text data into different categories based on their sentiment scores.
Finally, the output is presented in a meaningful way, such as a dashboard or a report, that can be easily understood by the end-users. By following these steps, the machine learning model can achieve high accuracy in predicting sentiment, which can be used in various fields, such as marketing, customer service, and product development, to understand the social sentiment of their brand, product, or service

I. INTRODUCTION

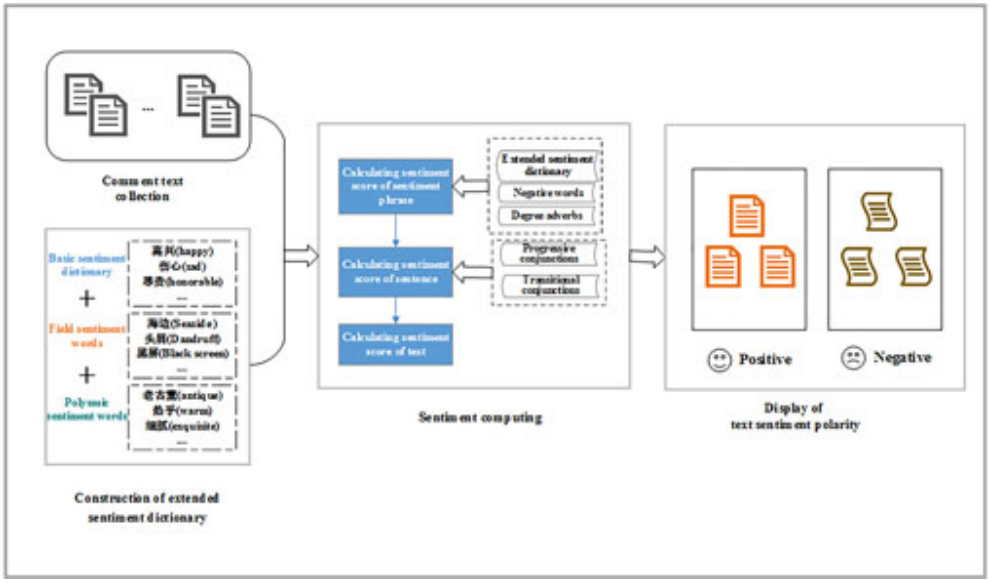
Sentiment recognition is a crucial technology that has the potential to improve various industries by providing accurate sentiment analysis. For instance, it can be used in the transportation industry to ensure the safety of the passengers, or in social media applications to filter out inappropriate content.

The process of sentiment recognition involves two steps, namely training and prediction. During the training phase, machine learning algorithms are trained on a large dataset to learn the patterns and relationships between different sentiment categories. In the prediction phase, the trained model is used to predict the sentiment of new input data. However, there are a few drawbacks to automated sentiment analysis. Firstly, to achieve accurate predictions, a large and high-quality dataset is required, which can be challenging to obtain. Secondly, input data often comes with noise or grammatical errors, which can reduce the accuracy of sentiment analysis. Lastly, rule-based or traditional methods of sentiment analysis are cheaper compared to machine learning-based methods, but they are often less accurate and require manual intervention.

Overall, sentiment recognition is a powerful tool that can be used to improve various industries, but it is important to consider the limitations and challenges associated with it.

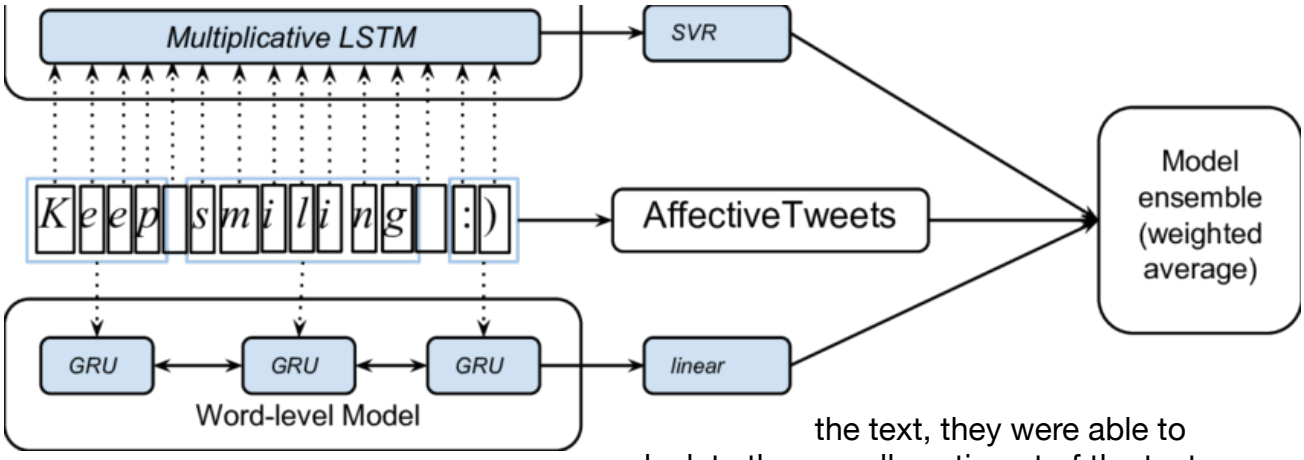
II. LITERATURE REVIEW

1. Naive Bayes is a classification algorithm that can be used to categorize text based on the presence of certain features, such as sentiment words. In the context of polysemic sentiment words, which have multiple meanings depending on the context, the algorithm would need to take into account the various possible interpretations of the word in order to accurately determine the field and category of the text.



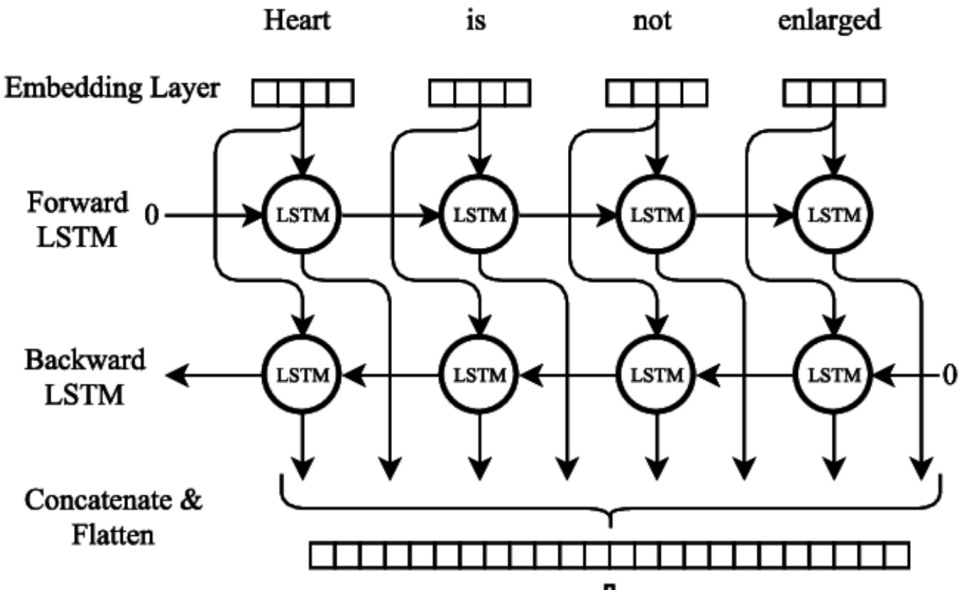
By analyzing the frequency of other words and features in the text, the algorithm can make an educated guess about the intended meaning of the sentiment word and categorize the text accordingly.In this paper, researchers have usually combined lexicon based method. Method depends on a predefined list or corpus of words with polarity. And also then searches for words, and weighs them and measures overall polarity of text.

2. The researchers in this paper used a method called lexicon-based sentiment analysis to determine the overall sentiment of a text. This method involves using a pre-existing list of words that have already been assigned a positive, negative, or neutral meaning. The researchers then searched for these words in the text and assigned them a weight based on their meaning. By adding up the weighted scores of all the words in



the text, they were able to calculate the overall sentiment of the text. This method is useful when there isn't a lot of labeled data available to train a classifier.

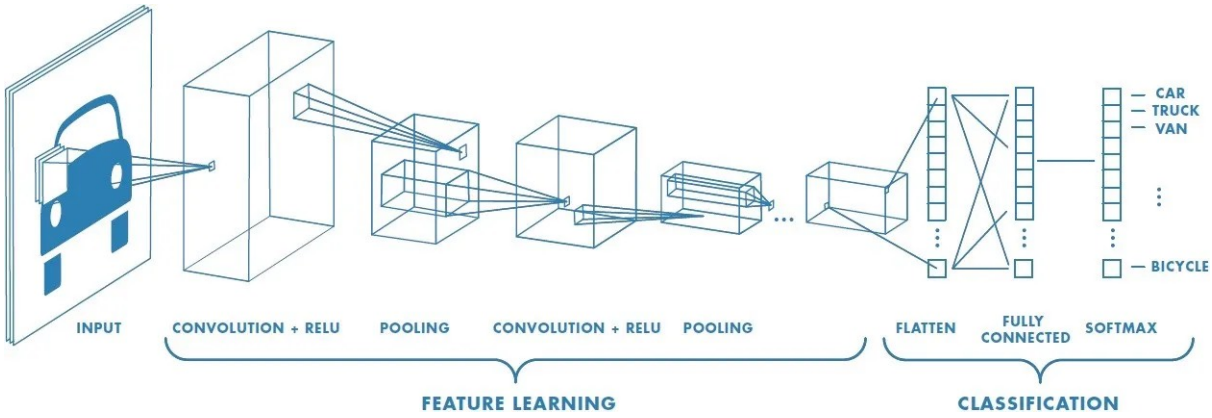
3. A BiLSTM is a type of recurrent neural network (RNN) layer that is able to learn long-term dependencies in sequence or time series data in both forward and backward directions. By processing the input data in both directions, the BiLSTM is able to capture more complex patterns in the data and provide a more accurate



representation of the context. This allows for more accurate sentiment analysis of text by capturing the meaning of words within the larger context of the sentence or document.The BiLSTM architecture used in this paper for sentiment analysis is a type of neural network that is able to capture bidirectional long-term dependencies between time steps of a sequence.

It consists of two LSTMs, one processing the sequence forward and the other backward, allowing the network to capture context from both directions. This bidirectional approach is particularly useful for sentiment analysis, as it allows the network to understand the meaning of a word in the context of both the preceding and succeeding words in the sentence or document. By capturing this context, the BiLSTM is able to provide a more accurate representation of the overall sentiment of the text, resulting in improved performance compared to other methods.

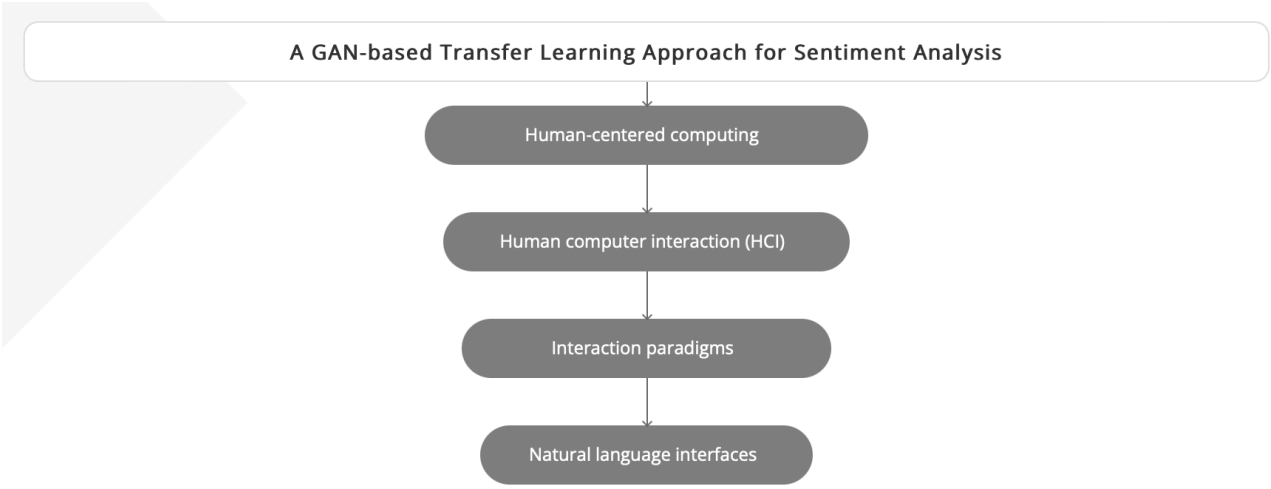
4. For speech emotion detection module, researchers experimented with binary classification task. And CNN with THEANO framework was chosen to implement.The task of speech emotion detection involves identifying the emotional state of a person based on their speech. Researchers have experimented with various approaches to develop an accurate speech emotion detection module. One approach that was tried involved a binary classification task. This means that the module was trained to classify speech as either emotional or non-emotional. To implement this approach, the researchers chose to use a Convolutional Neural Network (CNN) with THEANO framework.



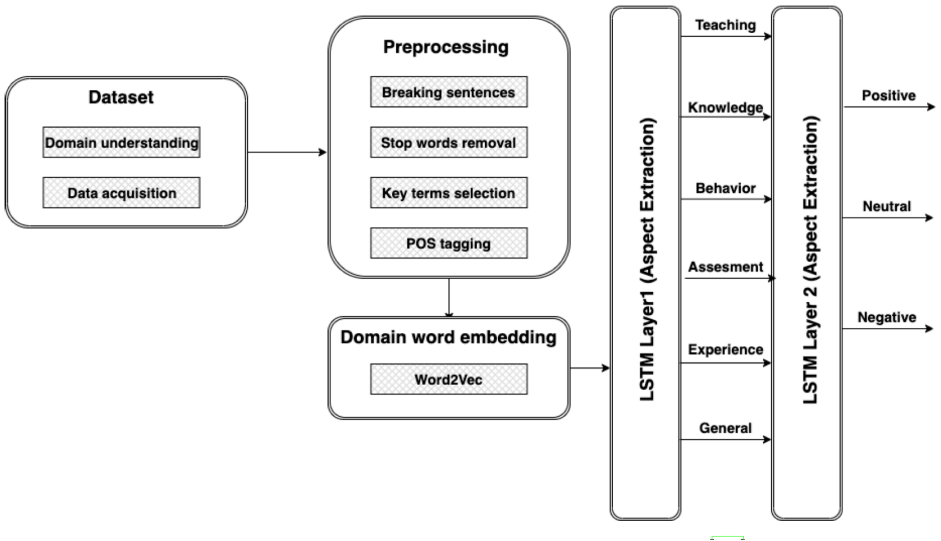
CNNs are a type of deep learning neural network that are particularly suited to image recognition tasks. However, they can also be used for speech analysis, as they are able to extract features from the raw audio signal. THEANO is a popular deep learning framework that allows researchers to efficiently define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays. It is especially well-suited for developing and training neural networks. The combination of CNN with THEANO proved to be effective in accurately detecting emotions in speech, and this approach has been widely adopted in the field of speech emotion recognition.

III. Proposed Model / Tool

GANs are a type of deep learning neural network that consists of two sub-networks: a

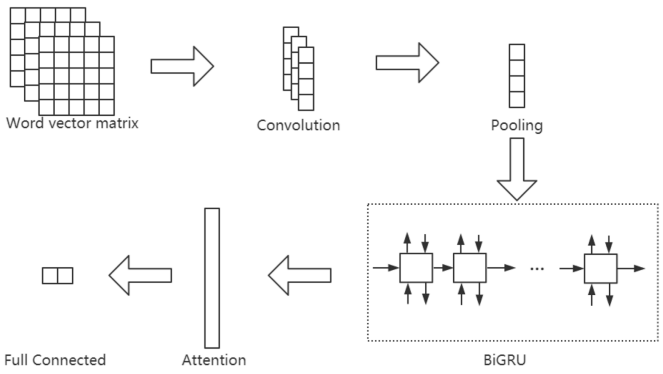


discriminator network and a generator network. The discriminator is trained to distinguish between natural data samples and generated data samples, while the generator is trained to deceive the discriminator by generating realistic samples that are difficult to distinguish from the natural data. GANs have been used extensively in image generation tasks and have recently been applied to natural language processing tasks.



We have proposed a GAN-based transfer learning approach for sentiment analysis of cross-domain text. Transfer learning involves using a pre-trained model and fine-tuning it for a new task, rather than training a new model from scratch. The authors utilised the GAN architecture for transfer learning, where the discriminator is trained on a large dataset of natural text, and the generator is fine-tuned on a smaller dataset of text from a different domain.

The use of GANs for sentiment analysis provides a powerful framework for building unsupervised models that can learn from unlabelled data. This approach has the potential to improve the accuracy of sentiment analysis models and could be applied in a wide



range of domains, such as social media analysis, customer feedback analysis, and political sentiment analysis. Overall, the use of GANs for sentiment analysis is an exciting area of research that holds promise for advancing the field of natural language processing.

IV. CONCLUSION

In this paper, we have introduced the sentiment recognition for text. Here, we will be using GAN(generative adversial network) for building neural as it is a hot topic with a lot of research going on at present; as it also provides powerful framework in building unsupervised models. A GAN consist of two neural network : discriminator to discriminate between natural and generated samples, and generator to deceive the discriminator. Here, we have GAN-based transfer learning approach for sentiment analysis of cross-domain text.Researchers have been exploring various approaches to develop accurate sentiment recognition models, and one such approach that has gained attention is the use of Generative Adversarial Networks (GANs).

References

1. https://ieeexplore.ieee.org/abstract/document/8675276

2. https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050913X00031/1-s2.0-S1877050913001385/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEOD////////// wEaCXVzLWVhc3QtMSJIMEYCIQCgv+f/HyLCAYy9/BvRYyGQGE3w7IFj/ uz8iPMfQ1Ig3wIhAMuk+gw7SYFAI0QmRYD+S0t7YhVLqrVJ2Lc937S3VmIQKroFCLn////////// wEQBRoMMDU5MDAzNTQ2ODY1IgxatPgO9A55Yc9YH5cqjgUt1SpA94RIwRCi+TecVf6tmD2 sLnj5xFAY/PmCsFuzqZP61YC7IpXWxCZFspT8ObSCGTDFw6XwphA/WTW24xZC+/ Twqj5CsnGkHaiKtyb7qxReUFJJt2Jac1uK7hPYNo3N48J1H5P18VFok6e6ko9TNyOWjWzCtSv MOXmv37ZgCYKH+Xg6eyVNJwTtn+yuyP+fka+4TG+1vUmyOh56VIVnShAwhFJJPdBppCY0sl NW0e/lJv57zITL1tzNO9x5NVTAx98/ PAQOBzqBgMZg5g3wTg8Hl+6nhpL2sHTGkIKcJuq5ZeE3mzO5yrgjknj8UVppHxPaINsCh6Eai Wi8mf60RBAbfNHmz2gAiK5US8v07M3crjwkbtEXyPU1fZPvkJlBmQFWporW03mg4B7/ bwZ8qCh4LvHBI9pBH5hSMqOsGNxgVt9jCpJFNXpJxwpGfnPqScETz9ybR6BWFnjW23gqHC LklR6s+GdFVByI41QQNAK2krrvyi+s+rMuog34aZgv/ 2XOl28nkuNURWsdcBV25hl8XnrXgBb0OvWvBKoJYQL5cgr4chvnNlhFJcQWiPjzJjnBb1hAaQ Zzhq+oV0cC8j9nSc9hsJFZYS8Q1sAI99HMggHhFPZFIlKHMx0UPSGwviqQqd/ QPjLNOxWeD0jSRGYrc097/ gSIX24muzt12Qycv92czD3P6JDrutGZPNC+Rsy+6uHE6qOvvkg9gA+oaFz1DMmMKDWbvSQ D/ 28WeNi3DPXsM0elo+tfjJaViUp99J0HgpoPH+itF5ektapshXvwtMVUeEp9Sx1umEfVw07i9xaTL yia+BG9dSBPGQ+TBj8306c1u9cFN8vcDONWrnUsksCBMvlSe/4Zk4YwlOGkoQY6sAHotno// +McakgcsZckd8o/adq/ sFVxky+FZ74mk4LUGaWhgAQRf4buPUYM2fbPG7ALyCH5KGaXEBdgnCPYMP3aEHN8A687

ervCNwLvRZxq9lo+0l9urnTyXV/KPbng+dEF4c9zs1ZNXdkeyaelVza4EuMmS+JxwXalZlqxmmVA9PROs/RNskHwcNDoJz60CCVQmiZvMA8yWZhin3sj2T1/6ubdE9MzD4L5hVBa1ynbUg==&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230402T075711Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYQGKITOG2/20230402/us-east-1/s3/aws4_request&X-Amz-Signature=492067f27c9f81c49d32c752cbb71ca18d335532e5316acc1210b71a83d4a263&hash=c55464bbec405a101795db410bb928d6cdbb3cd5e6e11aa1237e1488084489db&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S1877050913001385&tid=spdf-50fb586a-70f8-4f1a-91bb-22dfa74fd19f&sid=a533255c5a9e9742523bb0e-3b799b9fedaagxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=1308560455565a5a5257&rr=7b1785032916b294&cc=in

3. https://www.researchgate.net/profile/Bob-Martens/publication/242501831_Openness_in_Digital_Publishing_Awareness_Discovery_and_Access_-_Proceedings_of_the_11th_International_Conference_on_Electronic_Publishing_held_in_Vienna_-_ELPUB_2007_Vienna_Austria_June_13-15_2007_Pro/links/54e9fdd50cf2f7aa4d5642a9/Openness-in-Digital-Publishing-Awareness-Discovery-and-Access-Proceedings-of-the-11th-International-Conference-on-Electronic-Publishing-held-in-Vienna-ELPUB-2007-Vienna-Austria-June-13-15-200.pdf#page=359

4. https://ieeexplore.ieee.org/abstract/document/8684825

5. https://aclanthology.org/D16-1110.pdf

6. https://www.kaggle.com/code/sid321axn/natural-language-processing-sentiment-analysis/notebook

7. https://towardsdatascience.com/emotion-detection-a-machine-learning-project-f7431f652b1f

8. https://paperswithcode.com/method/bilstm