Sentiment Classification in Internet Memes: A Multimodal Study

Saish Vasudev Mhatre
Indiana University
svmhatre@iu.edu

Gandhar Ravindra Pansare
Indiana University
gpansar@iu.edu

Tanmayee Tajane
Indiana University
tatajane@iu.edu

project-B565-Fall2023/project_gpansar_svmhatre_tatajane

Abstract

The internet has given rise to a unique form of expression: memes. These humorous and frequently sarcastic pictures and text mixes have turned into a huge part of our existence on the web. Memes can convey a large number of feelings, from humor to anger to political commentary. Analyzing the sentiments behind these memes is not only a fascinating study in itself but also holds the potential for applications in various fields, including social media analysis, NLP, image/text classification, text mining, machine learning, deep learning, and sentiment analysis. This project investigates and comprehends the opinions behind memes through a far-reaching examination of a diverse dataset comprising 6992 meme images. The primary objectives of the project include sentiment classification and majority voting. The project will provide a holistic evaluation of classifier performance by presenting key metrics such as confusion matrices, accuracy, recall, precision, and F1-measure in the first part and accuracy vs loss in the second part. By analyzing both image and text data in the context of memes, our project adopts a multimodal approach to sentiment analysis, offering a deeper insight into the emotions and messages conveyed through this unique form of online communication. The results of this study can improve how we might interpret online sentiment dynamics and the manners by which people use humor and sarcasm to communicate their feelings and opinions in a digital age. Furthermore, the project's outcomes may have applications in content moderation, trend analysis, and brand sentiment monitoring, making it a valuable endeavor with relevance in various domains.

In summary, this project embarks on an exciting journey to uncover the sentiments and emotions behind internet memes, offering a rich dataset for analysis and a toolkit of machine learning techniques to decode the laughter, satire, and underlying messages of online meme culture.

Keywords: Sentiment analysis, memes, data mining, expression, NLP, and machine learning.

1. Introduction

Memes are an integral component of Internet communication platforms, which has made them an effective means of communicating thoughts, feelings, and responses. Memes are a great resource for learning multimodal sentiment analysis since they combine text, graphics, and occasionally even videos. Memes provide a special chance to investigate the complex interaction between words and visuals in expressing feelings and emotions because of the way that they juxtapose textual information and visual features. This project's main goal is to use social data mining, natural language processing, and machine learning to determine sentiment from meme image datasets. We hope to close the sentiment analysis gap that standard textual approaches frequently fall short of by doing this. We seek to identify and comprehend user sentiments, opinions, and humor through meme analysis, which can be useful for a variety of purposes such as content suggestion, social media monitoring, and market research.

Through the advent of this project, we will explain the project's objectives and methods in order to clarify how we intend to create a system that can reliably determine the sentiment contained in memes, as well as the possible advantages and uses of this cutting-edge instrument.

2. Previous work

An overview of the previous work on similar titles of the project brings forth that an approach that uses a combination of Vision transformers, Word-based transformers, sentence transformers, and Bidirectional LSTM giving a testing accuracy of 62.77[6]. For data preprocessing[1], the images are transformed into equal sizes of $150 \times 150 \times 3$ followed by the Keras image preprocessing function to make them suitable before driving into the CNN models. Another way implemented previously[2]uses Machine Learning methods and a process involving text detection and text extraction which is later supplied to a BERT language model along with image segmentation. For the text classification part, Pimpalkar et al. use the VGG16 convolution neural network model. Furthermore, an optimized LSTM model[5] was effectively used for Sentiment Analysis involving textual data(Amazon reviews, book reviews, and TripAdvisor dataset) where the text pre-processing was carried out using word tokenization and stemming followed by segregation. If only text data had been considered for sentiment analysis, a very good model[3] was formed based on the DL method and the GloVe word embedding approach, learns the features using a CNN layer and then coordinating all of them into a Multi-Layered Bi-Directional Long-Short-Term Memory (MBiLSTM) to capture long-range embedded circumstances. After this brief overview, the most impressive method followed a multimodal[4] involving VGG19 and BERT language model which gave an accuracy of 67.12 on a dataset of 10,115 images.

3. Methods

We implemented two models for dataset processing. The first one was solely based on text data extracted from the images. This model used the Hugging Face sentiment analysis architecture. The second model was 'multimodal' and used RoBERTa architecture with ResNet50 and VGG16 for constructing a neural network.

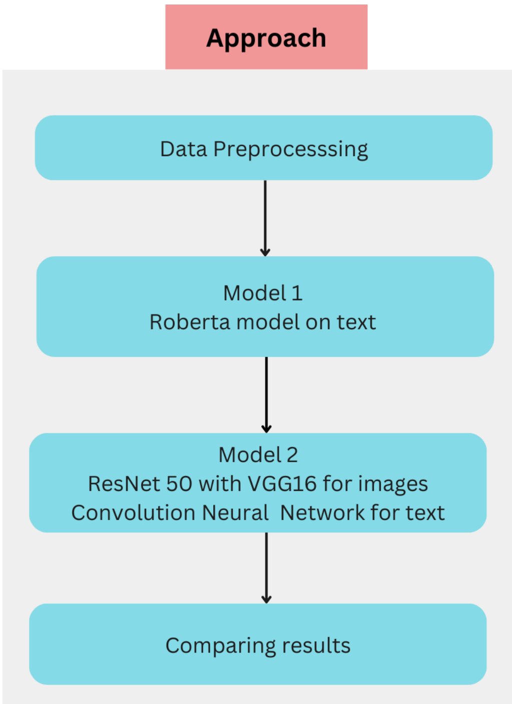

Fig. 1: Approach of the project

3.1 Data Pre-processing

Since the dataset comprises meme images, the OCR detected text, and the corrected text found in the images, we had to take certain measures to ensure that the data was suitable to feed the model. We handled the missing data by identifying and removing rows that contained NaN values. Further, we encoded the sentiment labels by mapping the text labels to numeric categories. Next, we performed image pre-processing by resizing the images to 100x100 pixels and normalizing RGB pixel values. We also processed the text data by converting all of it to lowercase, removing punctuations and numbers, and reducing the vocabulary corpus size.

3.2 Text Pre-processing

The cleaned text is then passed through the RoBERTa model with some changes to the model architecture. The text is passed through a bidirectional LSTM layer with 256 neurons followed by a convolution layer of 128 neurons. Before passing it to these layers we are embedding text input with a 50% dropout rate.

After this, we reduce dimensions using the pooling method to keep the most prominent features. Finally, the text is passed through a Dense layer consisting of 2048 neurons with a dropout rate of 50%.

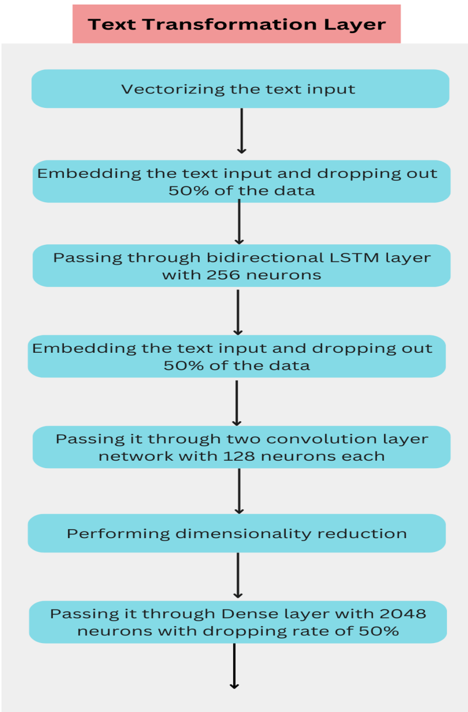

Fig. 2: Text pre-processing

3.3 Image Pre-processing

In the image transformation layer, we used two separate keras models, Resnet50 and VGG16, and then combined their functionalities. For the Resnet50 model, we first resized the image to 100*100*3 dimensions followed by data augmentation and pre-processing layer before passing it to our Resnet50 model.

Later, we passed it through 2048 and 512 neurons convolution and dense layer respectively. This generates our final output of the Resnet50 model.

For VGG16, we passed the images to the model, and to match the output of the Resnet model, we used a dense layer of 512 neurons. We have concatenated the outputs from these two models and passed this to the pooling layer with a 20% dropout rate to keep prominent features. This generates the final output of our Image transformation layer.

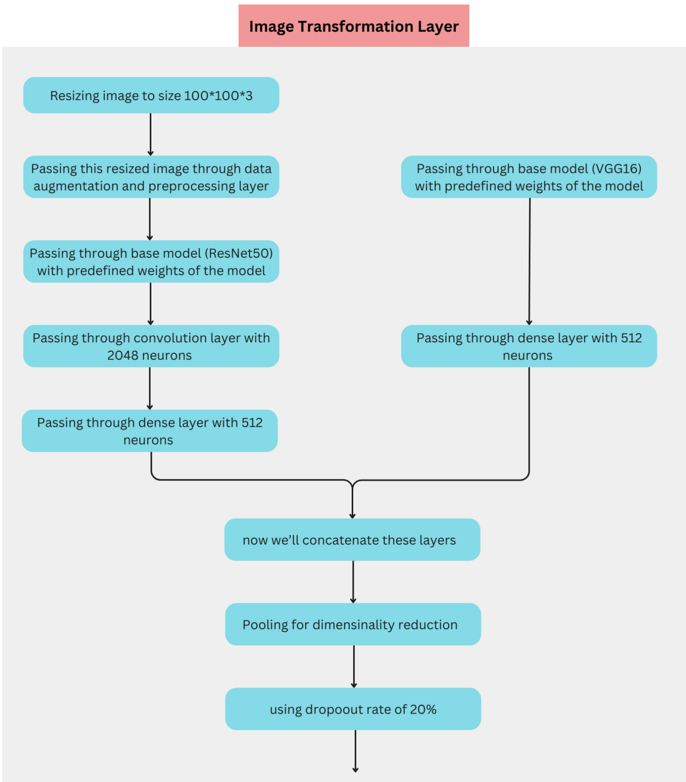

Fig. 3: Image Pre-processing

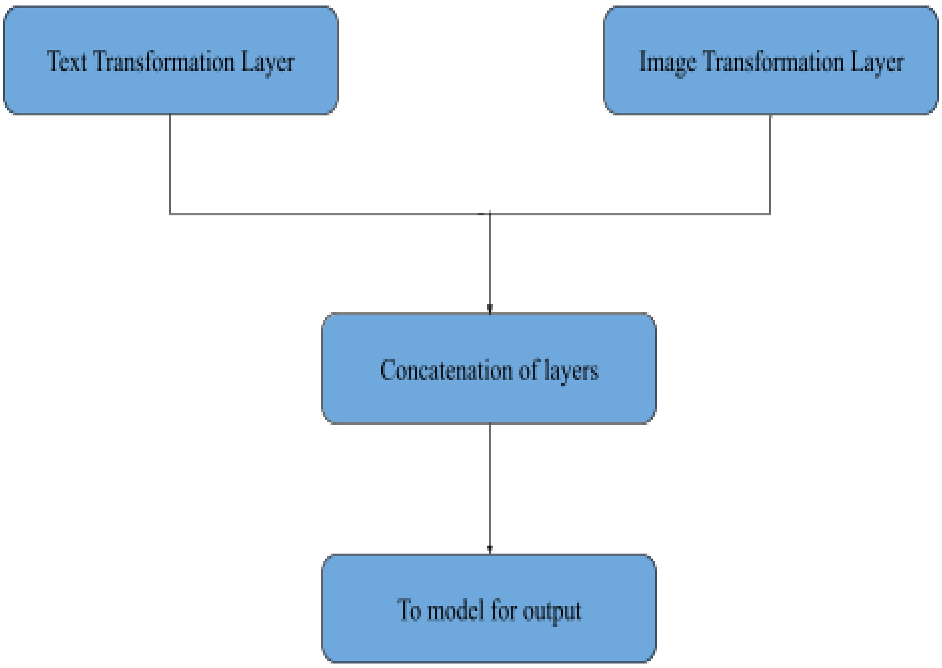

Fig. 4: Combination Model

4. Results

The results that we obtained from this model are as follows:

The text model constructed using RoBERTa architecture gave an accuracy of 25% after processing the text extracted from the meme images. The sentiment analysis model applied to meme text content using the Hugging Face sentiment analysis model exhibits the following training performance: Accuracy (0.2597): The model accurately predicts the sentiment of meme text content for approximately 25.97% of instances in the training dataset. Precision (0.3705): When the model predicts a positive sentiment for memes, it is correct around 37.05% of the time. Recall (0.2597): The model successfully identifies approximately 25.97% of memes with positive sentiment out of the total actual positive instances. F1 Score (0.2683): The F1 Score, representing the balance between precision and recall, is approximately 26.83%.

Whereas the multimodal model which uses a combination of ResNet50 and VGG16 gave an increased accuracy of 45%. Fig. 5 and Fig. 6 display the accuracy and loss experienced by the model for the 5 separate categories, “humor”, “offensive”, “motivational”, “ sarcasm” and overall accuracy.

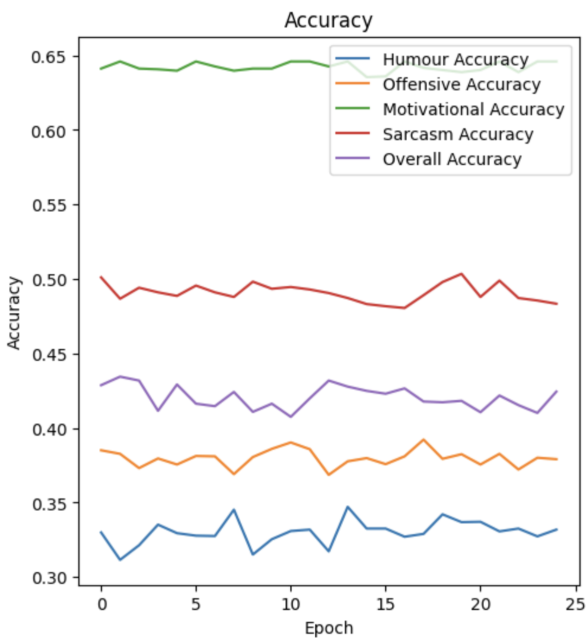

Fig. 5: Plot of accuracy

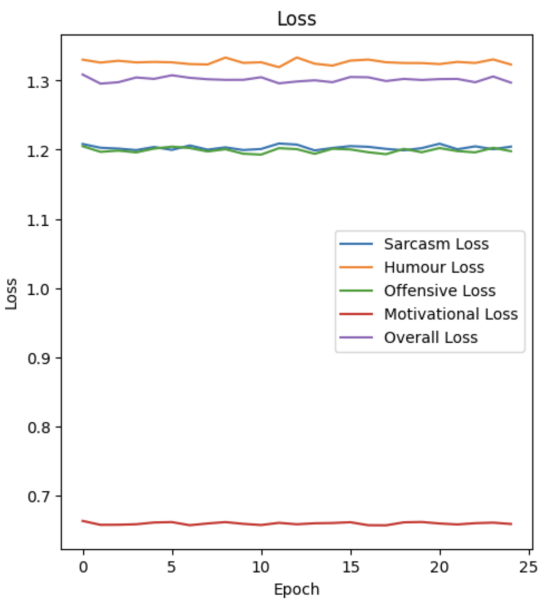

Fig. 6: Plot of loss

5. Discussion

We have used 5 separate categories to classify the memes. Our accuracy is heavily affected because of including 'sarcasm' and 'offensive' as our categories. The boundary between sarcasm and offense is very thin and that's where our model faces a major setback. Additionally, detecting sarcasm in a sentence is a big task for humans let alone a neural model, which further explains the low accuracy.

References

[1] Eftekhar Hossain, Omar Sharif, and Mohammed Moshiul Hoque. Memosen: A multimodal dataset for sentiment analysis of memes. In International Conference on Language Resources and Evaluation, 2022.

[2] Pimpalkar, Amit & Chaudhari, Ayush & Lilhare, Atul & Dighorikar, Nimish & Dakhole, Sagar & Asawa, Shreyash. (2022). Sentiment Identification from Image-Based Memes Using Machine Learning. International Journal of Innovations in Engineering and Science. 7. 89-96. 10.46335/IJIES.2022.7.8.16.

[3] Pimpalkar, Amit & R., Jeberson. (2022). MBiLSTMGloVe: Embedding GloVe knowledge into the corpus using a multi-layer BiLSTM deep learning model for social media sentiment analysis. Expert Systems with Applications. 203. 117581. 10.1016/j.eswa.2022.117581.

[4] Pranesh, Raj Ratn and Ambesh Shekhar. "MemeSem: A Multi-modal Framework for Sentimental Analysis of Meme via Transfer Learning." (2020).

[5] Jayakumar, Shobana & Malaisamy, Murali. (2021). An efficient sentiment analysis methodology based on long short-term memory networks. Complex & Intelligent Systems. 7. 10.1007/s40747-021-00436-4.

[6]N. V. Alluri and N. Dheeraj Krishna, "Multi-Modal Analysis of memes for Sentiment extraction," 2021 Sixth International Conference on Image Information Processing (ICIIP), Shimla, India, 2021, pp. 213-217, doi: 10.1109/ICIIP53038.2021.9702696.