Scientific Research Publishing

Journal of Computer and Communications, 2023, 11, 183-196
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Emotion Deduction from Social Media Text Data Using Machine Learning Algorithm

Thambusamy Velmurugan[1], Baskaran Jayapradha[2]

[1]PG and Research Department of Computer Science, Dwaraka Doss Govardhan Doss Vaishnav College, Chennai, India
[2]PG and Research Department of Computer Science, Dr. Ambedkar Government Arts College, Chennai, India
Email: velmurugan_dgvc@yahoo.co.in, jayapradha.bhaskar@gmail.com

How to cite this paper: Velmurugan, T. and Jayapradha, B. (2023) Emotion Deduction from Social Media Text Data Using Machine Learning Algorithm. *Journal of Computer and Communications*, **11**, 183-196.
https://doi.org/10.4236/jcc.2023.1111010

Received: October 8, 2023
Accepted: November 27, 2023
Published: November 30, 2023

Abstract

Emotion represents the feeling of an individual in a given situation. There are various ways to express the emotions of an individual. It can be categorized into verbal expressions, written expressions, facial expressions and gestures. Among these various ways of expressing the emotion, the written method is a challenging task to extract the emotions, as the data is in the form of textual dat. Finding the different kinds of emotions is also a tedious task as it requires a lot of pre preparations of the textual data taken for the research. This research work is carried out to analyse and extract the emotions hidden in text data. The text data taken for the analysis is from the social media dataset. Using the raw text data directly from the social media will not serve the purpose. Therefore, the text data has to be pre-processed and then utilised for further processing. Pre-processing makes the text data more efficient and would infer valuable insights of the emotions hidden in it. The preprocessing steps also help to manage the text data for identifying the emotions conveyed in the text. This work proposes to deduct the emotions taken from the social media text data by applying the machine learning algorithm. Finally, the usefulness of the emotions is suggested for various stake holders, to find the attitude of individuals at that moment, the data is produced.

Keywords

Data Pre-Processing, Machine Learning Algorithms, Emotion Deduction, Sentiment Analysis

1. Introduction

A significant amount of text data has accumulated as a result of the post-COVID spike in social media. This textual information reservoir has the capacity to forecast a number of significant variables and provide insightful information. In order to gain important insights into people's attitudes and sentiments, researchers are actively involved in the analysis of social media text data to extract emotions. Extracting the emotion involved in text data is a challenging task as it could bring out different understandings for the same text by different people and also depends on how the text data is read. The dataset taken for this research work consists of social media text data. It consists of comments of people in text format. The primary aim of this research would be to extract the emotion hidden in the text data.

Social media has taken a vast shape after the COVID-19 pandemic and has reshaped the way we interact, work, and communicate [1]. With social distancing measures in place, people turned to digital platforms more than ever to stay connected, informed, and entertained. Social media, in particular, witnessed unprecedented growth during this period. With the growth of text-based date in this period [2], the analysis of text-based data became an essential tool for understanding human emotions, sentiments, and behaviours. This article explores the surge in social media usage and the pivotal role that text-based data analysis plays an essential role in estimating human emotions in today's digital age.

Emotion is a mental state that is in line with feelings and thoughts, usually with regard to a particular thing. Emotion is a behaviour that expresses personal significance or opinion about how we connect with other people or about a particular occurrence. Due to different misconceptions about how they see the text data, humans are unable to comprehend its essence. This research uses machine learning technique to extract the information contained in the text data [3]. Emotion extraction from text is a natural language processing (NLP) task that involves identifying and categorizing the emotional content expressed in written or textual data. The goal is to determine the emotions, sentiments, or affective states conveyed by the author of the text. This technology has gained significant importance in various fields, including marketing, customer service, mental health, and social media analysis, as it provides valuable insights into how people feel and react in different contexts.

This paper is organized in the following way. Section 2 gives the literature survey of some of the papers related to this research. Section 3 describes the material and the methods used for the research. It has the dataset description taken for the research and the techniques used for pre-processing and also the result of applying the pre-processing techniques to the text data. In addition, it elaborates on the RoBERTa method for extracting the emotions from text data. Section 4 consists of the Experimental results and their discussions and the last section, Section 5 concludes the research with some of its findings in this research work.

2. Literature Survey

The significance of extracting emotions from text by using Natural Language Processing (NLP) has kindled the research interests of many researchers in this domain. While it's impractical to analyse deep into every research study comprehensively in this section, some of the related works on the related area has been discussed in this section. The main innovation in a study by Kantrowitz [4] is the recommendation to use a dictionary-based stemmer, which is effectively a perfect stemmer to analyse its impact on data retrieval. Its performance can be selectively changed in terms of coverage and accuracy. The system designers can more accurately evaluate the relative trade-offs between desired levels and increase stemming accuracy by using this stemmer.

Another research work by Sridevi *et al.*, titled "Impact of Preprocessing on Twitter Based Covid-19 Vaccination Text Data by Classification Techniques "in [5], takes up Twitter dataset and performs pre-processing on the data. It uses the classification algorithms LIBLINEAR and Bayes Net to determine the most effective techniques for data for preprocessing purposes. It is determined that pre-processed data results in greater performance and precision for the data analysis in contrast to the raw data.

The Sentiment Analysis and Emotion Detection subfield, with a focus on text-based emotion detection, is covered in detail in the article [6] by Acheampong. It begins by outlining the fundamental ideas of text-based emotion detection, emotion models, and emphasising the accessibility of big datasets necessary for the field of study. The article then describes the three main strategies frequently used in the creation of text-based emotion detection systems, outlining their advantages and disadvantages. The paper concludes by outlining current difficulties and prospective future study avenues for academics and researchers in the field of text-based data.

A research work titled, "Hierarchical Bi-LSTM based emotion analysis of textual data "by Mahto *et al.*, [7] suggests an improved deep neural network (EDNN) based on a hierarchical Bidirectional Long Short-Term Memory (Bi-LSTM) model for emotion analysis. The findings show that, in comparison to the current CNN-LSTM model, the suggested hierarchical Bi-LSTM technique achieves an average accuracy of 89% for emotion analysis. Another work carried out by Kumar *et al.* [8], has put forward the Emotion-Cause Pair Extraction (ECPE) technique to preprocess the text data at the clause level. To create sets of emotion and cause pairs for a document, it isolates cause clauses from emotion clauses, pairs them, and filters them. The BERT model receives its input from these pre-processed data. The classifier model performs at the cutting edge on a benchmark corpus for emotion analysis. The ECPE-BERT emotion classifier beats previous models on English sentences, obtaining a remarkable accuracy of 98%.

An article by Rashid *et al.* in [9], the researchers describe the Aimens system, which analyses textual dialogue to identify emotions. The Long Short-Term Memory (LSTM) model, which is based on deep learning, is employed by the system to identify emotions like happiness, sadness, and anger in context-sensitive speech. The system's primary input is a mixture of word2vec and doc2vec embeddings. The output findings exhibit significant f-score changes from the baseline model, where the Aimens system score is 0.7185. In the research article titled "An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network" by Shrivastava *et al.*, in [10], the authors offer a framework built upon Deep Neural Networks (DNN) for handling the problem of emotion identification inside multimodal text data. A TV show's transcript was used to create a brand-new dataset that was carefully curated for the emotion recognition test. In order to extract pertinent characteristics from the text dataset, a CNN model with an attention mechanism was trained using the obtained information. The effectiveness of the suggested model was assessed and contrasted with benchmark models like LSTM and Random Forest classifiers.

3. Methods and Materials

There are various methods that are used for text pre-processing and emotion prediction from text data. This article is categorized into two stages as Data pre-processing and emotion extraction. The various methods that are used for pre-processing is detailed along with the dataset that is taken for this research a then the emotion extraction is applied to the pre-processed text.

3.1. Description of the Dataset

The data set taken for this research work is from a text-based, social media dataset consisting of text in the form of a sentence. This sentence expresses the current emotion of an individual such as joy, sad, fear, anger, surprise and so on. The emotion of the induvial can never be predicted from the text easily. The purpose of this research work is to predict the emotion of an individual from the text data that is taken for the study after pre-processing by applying the machine learning model.

3.2. Dataset before Pre-Processing

The chosen dataset for this work simply consists of induvial expressions in the form of sentences. It consists of only one attribute. The attribute is in the form of a sentence by a person expressed in direct speech. The text data which is taken for this research work is an uncleaned data and need to be pre-processed for the effective application of the machine learning algorithm.

Table 1 consists of the dataset used for the research before pre-processing. The text data is a combination of words, punctuations and many other textual representations. The objective is to eliminate the unnecessary words and symbols which are expressed along with the root word and to predict the emotion from the text taken for the research by using the machine learning algorithm.

3.3. Dataset after Pre-Processing

Raw data must be transformed into legible and defined sets, in order for researchers to conduct data mining, analyse the data, and process it for various activities. It is a must to correctly preprocess their data as a variety of inputs they utilise to gather raw data might have an impact on the data's quality. Preprocessing data is crucial because raw data may be formatted inconsistently or incompletely. Preprocessing raw data effectively can increase its accuracy, which can raise project quality and reliability. The various stages that are involved in the process of preprocessing of text data in this research are lowercasing, punctuation removal, stop word removal, tokenization, stemming and lemmatization [11]. These steps help the researchers effectively to interpret the underlying emotion in the text involved from the dataset taken. By pre-processing researchers would be able to uncover valuable insights, detect patterns and predict user behaviour and understand the emotional content of any individual easily.

Table 1. Dataset before pre-processing.

COMMENTS
I AM FEELING QUITE SAD AND SORRY FOR MYSELF BUT I WILL SNAP OUT OF IT SOON!!!
I feel like I am still looking at a blank canvas blank pieces of paper
I feel like a faithful servant
I am just feeling cranky and blue
I can have for a treat or if I'm feeling festive!!
I start to feel more appreciative of what god has done for me
I am feeling more confident that we will be able to take care of this baby
I feel incredibly lucky just to be able to talk to her
I feel less keen about the army every day

The pre-processing stages of the taken dataset are as follows. Initially all the text that are involved in the research are converted to lower case. There might be punctuations in the text data which would be unnecessary. This step is to remove the punctuations in the text data. Then the stop words like in, to, is are removed. Secondly the entire sentence is broken into tokens. Thirdly the tokenized words are stemmed (Running is changed as Run) and lastly the words are lemmatized which draws the essence of the sentence without changing the meaning of the words.

Table 2 depicts the pre-processed text data taken from **Table 1**. The pre-processed text may change the actual spelling of the word that is involved but the meaning of it would be the same and would contribute to a greater extent in the process of analysing the given dataset. The resultant data is subject to various techniques of machine learning for further identification of emotions. These pre-processing methods play an essential role to extract the exact content needed from the social media data [12].

The research also produces another file that compares the word count between the original text and the pre-processed text.

Table 2. Dataset after pre-processing.

COMMENTS PRE-PROCESSED
[“feel”, “quit”, “sad”, “sorri”, “snap”, “soon”]
[“feel”, “like”, “still”, “looking”, “blank”, “canvas”, “blank”, “pieces”, “paper”]
[“feel”, “like”, “faith”, “servant”]
[“feel”, “cranki”, “blue”]
[“treat”, “feel”, “festiv”]
[“start”, “feel”, “appreci”, “god”, “done”]
[“feel”, “confid”, “abl”, “take”, “care”, “babi”]
[“feel”, “incred”, “lucki”, “abl”, “talk”]
[“feel”, “le”, “keen”, “armi”, “everi”, “day”]

Table 3 illustrates the reduction in the number of words throughout the different stages of pre-processing for the research. It is discovered that the total word count in the original text has been reduced to approximately one-third of its original size after completing the pre-processing steps.

Table 3. A comparative analysis (number of words).

S. No	Number of words in the original Text	Number of words in the Pre-Processed Text
1	25	13
2	16	8
3	23	11
4	9	5
5	13	6
6	26	8
7	8	5
8	18	11
9	24	11
10	43	22

Table 3 shows the comparison of the number of words that has been in the original dataset and the number of words produced after pre-processing. It is clearly evident that the number of words in each case of the comments gets reduced to one-third of the words in the original data which is taken for the research. This makes the further work of predictions on the chosen dataset easier and simple. It would be easy for applying any of the algorithms on the dataset for further analysis of predicting the emotion.

The data from various social network platforms are useful for the real-world applications. But it contains a large number of unprocessed words. Therefore, it becomes mandatory to remove unwanted words from the original dataset taken for the research. It also becomes essential to use some of the preprocessing techniques to make the text dataset shrink which would make it easy for applying any of the machine learning algorithm for extracting the emotion. A number of techniques are available in the field of sentiment analysis for pre-processing and text reduction like stop words removal, stemming and many more.

Figure 1 demonstrates the graph comparison of the number of words on the text data taken for the research before and after pre-processing. It shows that there is almost one-third reduction in the number of words before and after pre-processing. This makes the further work on the dataset easier as there is a considerable reduction in the number of words before and after pre-processing.

3.4. The RoBERTa Model

RoBERTa ("Robustly Optimized BERT Approach") is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model, which was developed by researchers at Facebook AI. Like BERT, RoBERTa is a transformer-based language model that uses self-attention to process input sequences and generate contextualized representations of words in a sentence [13]. The model trains a larger dataset and with a more efficient training process than BERT. Additionally, in order to aid the model in learning more resilient and broadly applicable word representations, the model employs a dynamic masking strategy during training [14].

RoBERTa also performs better than BERT and other cutting-edge models on a range of Natural Language Processing tasks, such as text classification, question answering, and language translation. It is considered as a well-liked option for both academic study and commercial applications. It also has served as the foundation model for numerous other effective NLP models [15]. Additionally, RoBERTa employs a method known as "No-Mask-Left-Behind" (NMLB), which guarantees that every token is masked at least once during training, in contrast

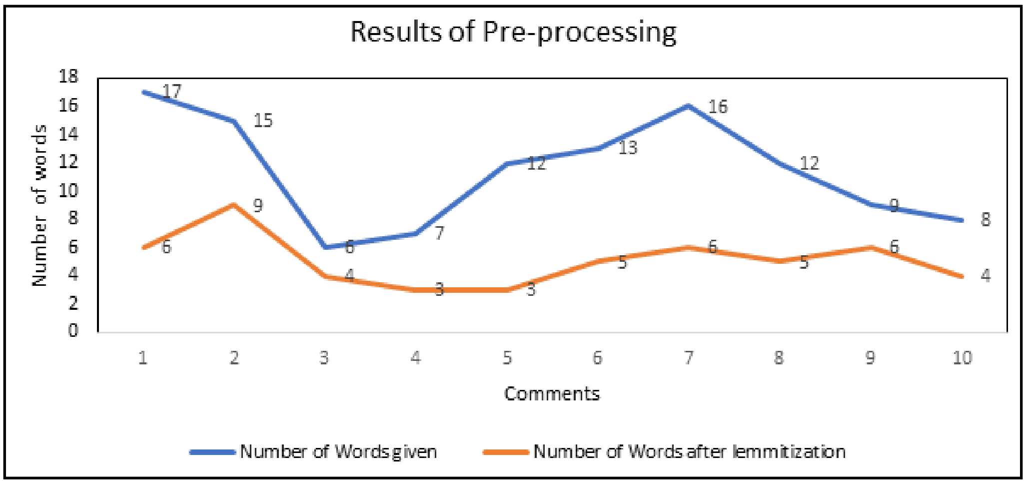

Figure 1. Comparison of the number of words.

to BERT, which employs a method known as "Masked Language Modelling" (MLM), which masks only 15% of the tokens [16]. As a result, the input sentence is represented more accurately and the sentiment analysis on the text data is more precise. The model is a potent tool particularly for NLP tasks.

3.5. Architecture of RoBERTa Model

The architecture of the RoBERTa model is the same as that of the BERT model. It is a reimplementation of BERT with some significant hyperparameter changes and minor embedding. Except for the output layers, the identical architectures are used in both pre-training and fine-tuning in BERT.

All parameters are modified during the process of fine-tuning. The RoBERTa, does not use the next-sentence pretraining target and is trained with substantially bigger mini-batches and learning rates. Furthermore, RoBERTa employs a different pretraining approach and replaces a byte-level BPE tokenizer with a character-level vocabulary. This makes the model more reliable and efficient for extracting emotion from text data.

RoBERTa is trained on a huge dataset spanning over 160 GB of uncompressed text [17]. RoBERTa's dataset includes (16 GB) of English Wikipedia and Books Corpus, which are used in BERT. The Web text corpus (38 GB), CommonCrawl News dataset (63 million articles, 76 GB), and Stories from Common Crawl (31 GB) were also included. This dataset, combined with 1024 Tesla V100 GPUs running for a day, was used to train RoBERTa. The Facebook team first converted BERT from Google's TensorFlow deep-learning framework to their framework, PyTorch, to build RoBERTa. RoBERTa was trained using

1) FULL-SENTENCES with no NSP loss;
2) dynamic masking;
3) big mini-batches;
4) a larger byte-level.

Figure 2 shows the architecture of the model. The left half of the Transformer architecture, is to map an input sequence to a sequence of continuous representations, which is then fed into a decoder. The decoder, on the right half of the architecture, receives the output of the encoder together with the decoder output at the previous time step to generate an output sequence.

RoBERTa, like BERT, is pre-trained on a huge corpus of text, but it has more thorough training data and training objectives. RoBERTa, like BERT, is pre-trained on a huge corpus of text, but it has more thorough training data and training objectives. To learn language representations, it is trained on enormous amounts of text data from the internet. Its features include,

1) Language Model (MLM): It employs an MLM technique in which some words in a sentence are masked out and the model is trained to predict these masked words. This bi-directional context learning assists it in understanding linguistic nuances [18].

2) Larger Training Datasets: RoBERTa learns a richer comprehension of

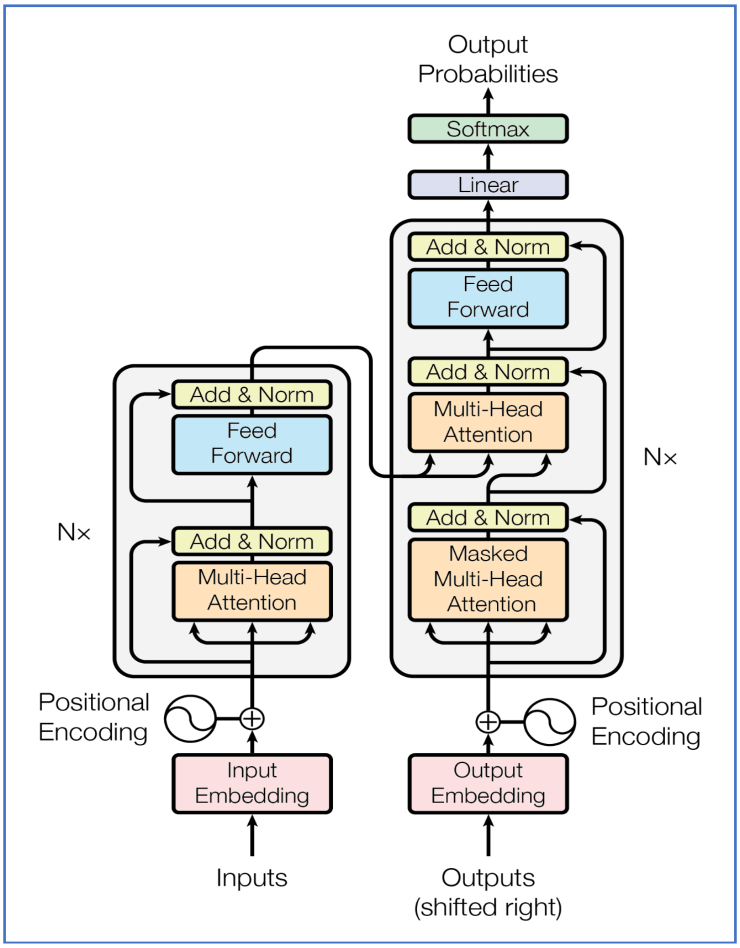

Figure 2. The transformer architecture.

language by using more training data and longer sequences than BERT [19].

3) No Next Sentence Prediction (NSP): Unlike BERT, RoBERTa does not use the Next Sentence Prediction task during pre-training instead it focuses entirely on MLM. This modification has been shown to boost performance in downstream NLP tasks [20].

4) Hyperparameter Tuning: Roberta optimises hyperparameters and training procedures in order to improve the model performance [21].

4. Experimental Results

The dataset that is involved for this research work consists of individual persons comments taken from a social media post. These comments are all in a direct speech where the individual expresses his or her emotion at that moment of time. These comments depict their mood and their thoughts at that particular time while expressing their emotions. These emotional text data in the form of comments help us to know about their current emotion at that particular time, however it does not serve as a measure to know all about the persons on the whole. The given dataset consists of data in an uncleaned format. In order to process the data for better understanding and to withdraw insights from it by applying data mining algorithms on it, the gathered data should be cleaned and pre-processed. The dataset taken here consists of missing words, wrong spellings, punctuations, conjunctions, prepositions and many more. This noisy information accumulated along with the data should be removed for better understanding of the textual data before any algorithm is applied on it for analysis.

After the process of pre-processing on the dataset, which is involved in this research, the text is subject to a process of fine grinding by the RoBerta model. As it is a fine grinded model, the model extracts around 28 emotions on the whole. This is comparatively a larger number of emotion when compared to the traditional methods which would extract only a fewer emotion from the text data.

The result obtained is shown in **Table 4** which extracts the emotion involved in the text data taken for the research. Column 1 is "comments" which is the original uncleaned text data. Column 2 is the result of pre-processing on the text data. After the data is subject to the pre-processing pipeline, it is transformed as a cleaned data eliminating the unnecessary words in the text, without changing the actual meaning of the text. This would be efficient in extracting the emotion

Table 4. Emotion extraction from text data.

Comments	pre_prcessed_comments	Emotions
I feel bitchy but not defeated yet	["feel", "bitchi", "defeat", "yet"]	Anger
I was dribbling on mums coffee table looking out of the window and feeling very happy	["dribbl", "mum", "coffe", "tabl", "look", "window", "feel", "happi"]	joy
I woke up often got up around am feeling pukey radiation and groggy	["woke", "often", "got", "around", "feel", "pukey", "radiat", "groggi"]	neutral
I was feeling sentimental	["feel", "sentiment"]	sadness
I walked out of there an hour and fifteen minutes later feeling like I had been beaten with a stick and then placed on the rack and stretched	["walk", "hour", "fifteen", "minut", "later", "feel", "like", "beaten", "stick", "place", "rack", "stretch"]	sadness
I never stop feeling thankful as to compare with others I considered myself lucky because I did not encounter ruthless pirates and I did not have to witness the slaughter of others	["never", "stop", "feel", "thank", "compar", "other", "consid", "lucki", "encount", "ruthless", "pirat", "wit", "slaughter", "other"]	gratitude
I didn't feel abused and quite honestly it made my day a little better	["feel", "abus", "quit", "honestli", "made", "day", "littl", "better"]	Joy
I know what it feels like he stressed glaring down at her as she squeezed more soap onto her sponge	["know", "feel", "like", "stress", "glare", "squeez", "soap", "onto", "spong"]	Neutral
I also loved that you could really feel the desperation in these sequences and I especially liked the emotion between knight and squire as they've been together in a similar fashion to batman and robin for a long time now	["also", "love", "could", "realli", "feel", "desper", "sequenc", "especi", "like", "emot", "knight", "squir", "theyv", "togeth", "similar", "fashion", "batman", "robin", "long", "time"]	Love
I had lunch with an old friend and it was nice but in general I'm not feeling energetic	["lunch", "old", "friend", "nice", "gener", "im", "feel", "energet"]	Joy

from text data taken for this research. Column 3 is the result of applying the fine grind model, RoBERTa on the text data to extract the emotion hidden in the text. The result obtained may serve as a great key in disclosing the emotion hidden in an individual just by analysing the textual content of the person. This would serve as an essential factor to understand the person and their attitude. It would also be of great value for various stakeholders in various fields as in today's scenario it becomes very crucial and necessary to understand the attitude of an individual for various reasons.

Table 5 contains a sample of the various emotions extracted from text data of the dataset taken for this research. It shows around 15 emotions in column 1, which is obtained after the process of extracting the emotion by the RoBERTa model. Column 2 shows the count of the number of emotions in each of the extracted emotion.

The research uncovers a number of emotions from text data. The traditional models and algorithms would be able to produce only fewer emotions which would be not sufficient to know the attitude of the person who has commented through text data. The result of the research would be of immense help for various stakeholders to know and understand the attitude of the individual. This would not only serve as a factor to know about the person but also understand the person's inner feeling or emotion at that particular moment of commenting through text data.

Table 5. Count of the emotions extracted.

Emotions	Count
Remorse	2
Neutral	4
Approval	1
Annoyance	2
Joy	5
Admiration	1
Caring	1
Realization	2
Embarrassment	3
Anger	1
Sadness	3
Gratitude	1
Love	2
Optimism	1
Disappointment	1

Figure 3 depicts a sample of the various emotions extracted from the dataset taken for this research. A sample of the emotion extracted from the result is shown. It shows various emotions like joy, anger, care, sadness and many more.

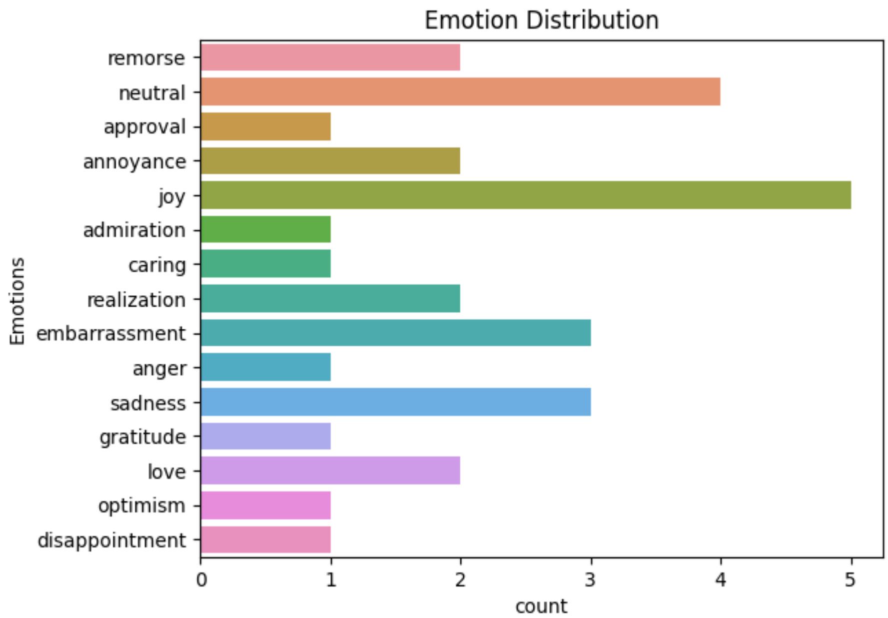

Figure 3. Emotion extracted from the dataset

It also gives the count of the number of people with the particular emotion. Unlike sentimental analysis which predicts whether the given text is positive, negative or neutral, this emotional extraction model, goes deeper and assesses the emotion of the person. The result obtained would serve as a boon for various stakeholders to know the attitude of a person.

5. Conclusion

Finding and analysing the emotions in social media text data is not an easy task. A crucial analysis is required to find the emotions hidden in text data. This research work is carried out to extract the hidden emotions from textual data and has yielded remarkable results, enabling us to identify approximately 28 distinct emotions within the text. These findings hold great promise for a wide range of applications. These applications include enhancing our understanding of an individual's personality and attitude. It also provides valuable insights for various stakeholders. Educators can utilize this information to better comprehend the attitudes of their students. This can enable more enhanced and effective teaching strategies to improve the understanding of the student's community. Parents can gain insight of the emotional states of their children, which would aid in the prevention of mental health problems, attitudes and suicidal thoughts in children. Additionally, interviewers can use this knowledge to gain a deeper understanding of the mental and emotional position of potential and eminent candidates. This would improve the selection and placement process. In essence, the current research contributes positively to countless factors to understand the human Psychology of a person and to understand their emotions.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Mason, A.N., Narcum, J. and Mason, K. (2021) Social Media Marketing Gains Importance after Covid-19. *Cogent Business & Management*, **8**, Article ID: 1870797. https://doi.org/10.1080/23311975.2020.1870797

[2] Feldkamp, J. (2021) The Rise of TikTok: The Evolution of a Social Media Platform during COVID-19. In: Hovestadt, C., Recker, J., Richter, J. and Werder, K., eds., *Digital Responses to Covid-19: Digital Innovation, Transformation, and Entrepreneurship during Pandemic Outbreaks*, Springer, Cham, 73-85. https://doi.org/10.1007/978-3-030-66611-8_6

[3] Chong, W.Y., Selvaretnam, B. and Soon, L.-K. (2014) Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets. 2014 4*th International Conference on Artificial Intelligence with Applications in Engineering and Technology*, Kota Kinabalu, Malaysia, 03-05 December 2014. https://doi.org/10.1109/ICAIET.2014.43

[4] Kantrowitz, M., Mohit, B. and Mittal, V. (2000) Stemming and Its Effects on TFIDF Ranking. *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, 24-28 July 2000, 357-359. https://doi.org/10.1145/345508.345650

[5] Sridevi, P.C. and Velmurugan, T. (2022) Impact of Preprocessing on Twitter Based Covid-19 Vaccination Text Data by Classification Techniques. 2022 *International Conference on Applied Artificial Intelligence and Computing* (*ICAAIC*), Salem, India, 9-11 May 2022. https://doi.org/10.1109/ICAAIC53929.2022.9792768

[6] Acheampong, F.A., Chen, W.Y. and Nunoo-Mensah, H. (2020) Text-Based Emotion Detection: Advances, Challenges and Opportunities. *Engineering Reports*, **2**, Article ID: e12189. https://doi.org/10.1002/eng2.12189

[7] Dashrath and Subhash Chandra Yadav (2022) Hierarchical Bi-LSTM Based Emotion Analysis of Textual Data. *Bulletin of the Polish Academy of Sciences, Technical Sciences*, **70**, Article No. e141001.

[8] Kumar, A. and Jain, A.K. (2022) Emotion Detection in Psychological Texts by Fine-Tuning BERT Using Emotion-Cause Pair Extraction. *International Journal of Speech Technology*, **25**, 727-743. https://doi.org/10.1007/s10772-022-09982-9

[9] Rashid, U., Iqbal, M.W., Skiandar, M.A., Raiz, M.Q., Naqvi, M.R. and Shahzad, S.K. (2020) Emotion Detection of Contextual Text Using Deep Learning. 2020 4*th International Symposium on Multidisciplinary Studies and Innovative Technologies* (*ISMSIT*), Istanbul, Turkey, 22-24 October 2020, 1-5. https://doi.org/10.1109/ISMSIT50672.2020.9255279

[10] Shrivastava, K., Kumar, S. and Jain, D.K. (2019) An Effective Approach for Emotion Detection in Multimedia Text Data Using Sequence Based Convolutional Neural Network. *Multimedia Tools and Applications*, **78**, 29607-29639. https://doi.org/10.1007/s11042-019-07813-9

[11] Jayapradha, B. and Velmurugan, T. (2003) Pre-Processing Emotional Text Data for

Sentiment Analysis. *International Conference on Information, System and Convergence Applications*. Tashkent, Uzbekistan, 3-6 July 2023, 350-360.

[12] Plisson, J., Lavrac, N. and Mladenic, D. (2004) A Rule Based Approach to Word lemmatization. *Proceedings of IS*, **3**, 83-86.

[13] Purachary, M. and Adilakshmi, T. (2003) Finetuned RoBERTa Architecture for MOOCS Evaluation using Adversarial Training. *Journal of Theoretical and Applied Information Technology*, **101**.

[14] Lin, T.-M., Chang, J.-Y. and Lee, L.-H. (2023) NCUEE-NLP at WASSA 2023 Shared Task 1: Empathy and Emotion Prediction Using Sentiment-Enhanced RoBERTa Transformers. *Proceedings of the* 13*th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, 548-552. https://doi.org/10.18653/v1/2023.wassa-1.49

[15] Adoma, A.F., Henry, N.-M. and Chen, W.Y.(2020) Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition. IEEE *International Computer Conference on Wavelet Active Media Technology and Information Processing*. https://doi.org/10.1109/ICCWAMTIP51612.2020.9317379

[16] Acheampong, F.A., Nunoo-Mensah, H. and Chen, W.Y. (2021) Transformer Models for Text-Based Emotion Detection: A Review of BERT-Based Approaches. *Artificial Intelligence Review*, 1-41. https://doi.org/10.1007/s10462-021-09958-2

[17] Ameer, I., Bölücü, N., Siddiqui, M.H.F., Can, B., Sidorov, G. and Gelbukh, A. (2023) Multi-Label Emotion Classification in Texts Using Transfer Learning.*Expert Systems with Applications,* **213**, 118534. https://doi.org/10.1016/j.eswa.2022.118534

[18] Zhou, Y., Xing, Y.Y., Huang, G.M., Guo, Q.K. and Deng, N.X. (2023) Multimodal Emotion Recognition Based on Multilevel Acoustic and Textual Information. *Proceedings of Fifth International Conference on Artificial Intelligence and Computer Science*, **12803**, 594-599. https://doi.org/10.1117/12.3009468

[19] Qin, X.Y., Wu, Z.Y., Zhang, T.T., Li, Y.R., Luan, J., Wang, B., Wang, L. and Cui, J.S. (2023) BERT-ERC: Fine-Tuning BERT Is Enough for Emotion Recognition in Conversation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 13492-13500. https://doi.org/10.1609/aaai.v37i11.26582

[20] Rajapaksha, P., Farahbakhsh, R. and Crespi, N. (2021) Bert, XLNet or Roberta: The Best Transfer Learning Model to Detect Clickbaits. *IEEE Access*, 9, 154704-154716. https://doi.org/10.1109/ACCESS.2021.3128742

[21] Xu, J.X. and Vinluan, A.A. (2023) Emotional Analysis and Prediction Based on Online Book User Comments. *Proceedings of Fifth International Conference on Artificial Intelligence and Computer Science,* **12803**, 157-164. https://doi.org/10.1117/12.3009554