Smartphone Android App for Self-Monitoring Emotional and Mental Well-Being at the University using NLP

Jaywant Lugun [1, a)], Himanshu [2, b)], Khadija Abdulfatah Mussa Iddi [3, c)], Vegunta Yeseswini [4, d)] and Aman Deep [5, e)]

[1,2,3,4,5] *Department of Computer Science and Engineering,*

Lovely Professional University,

Jalandhar, 140011,

Punjab, India

[e)] *Corresponding author: aman.29605@lpu.co.in*

Contributing authors: [a)] *jaywantlugun2000@gmail.com* , [b)] *himanshu.11905095@lpu.co.in* , [c)] *kbtwins9899@hotmail.com* , [d)] *yeseswiniv@gmail.com*

Abstract. This paper details the process of creating an Android app that allows college students to keep track of their own mental and emotional health using natural language processing (NLP) tools. The app's stated goal is to equip students with a means of monitoring their own mental and emotional health and providing them with constructive criticism and advice for growth. The study's overarching question is this: how might NLP be used for the creation of a mobile app that helps college students keep tabs on their own mental and emotional health? The author adopted a mixed-methods strategy, drawing on the results of a literature review, interviews, and pilot research. The results show that the app prototype was well-received by the focus group members and that NLP approaches can effectively discern emotional and mental states from natural language inputs. More development and testing are necessary, the author concludes, because the software shows promise as a tool for boosting emotional and mental well-being among university students. Future research directions and potential applications of this discovery are discussed.

Keywords. sentiment analysis, deep learning, pre-processing, tokenization, model training, evaluation, deployment, RNN, LSTM, TensorFlow

INTRODUCTION

Mental health is an important aspect of total well-being and has a substantial influence on a person's level of life. Unfortunately, mental health disorders are widespread around the world, with one in every four people encountering a mental health problem at some point in their lives. Mental health issues can present in a variety of ways, including sadness, anxiety, and stress, and they can affect people of all ages, genders, and backgrounds.

Over the years, various interventions and approaches have been developed to address mental health issues. These include traditional mental health care, such as therapy and medication, as well as emerging technologies like digital mental health interventions. One such technology that has shown promise in improving mental health outcomes is sentiment analysis.

It is a form of NLP that involves the analysis of text data to determine the emotional state of individuals or groups. It can be used to analyze various types of text data, including social media posts, customer reviews, and other online content, to gain insights into the emotional state of individuals and communities. By analyzing the sentiment of text data, sentiment analysis can provide valuable insights into mental health issues, including identifying individuals who may be struggling with mental health issues, tracking mental health trends and patterns, and providing timely support and resources to those in need.

This research paper aims to explore the benefits of sentiment analysis in improving mental well-being. Specifically, the paper will focus on how sentiment analysis can be used in chat applications to provide immediate support and resources to users and detect warning signs of mental health issues. The paper will also examine how sentiment analysis can help to reduce the stigma around mental health and encourage users to seek help when needed.

Overall, the purpose of this research paper is to present a thorough review of the possible advantages of sentiment analysis for enhancing mental health. The goal of this research is to add to the growing body of knowledge about how people use technology in mental health treatment and to aid in the identification of new potential for improving mental health outcomes.

LITERATURE REVIEW

With an estimated 1 in 4 persons experiencing a mental health condition at some point in their lives, mental health concerns are a major public health concern worldwide. Over the years, various interventions and approaches have been developed to address mental health issues, including traditional mental health care, such as therapy and medication, as well as emerging technologies like digital mental health interventions. One such technology that has shown promise in improving mental health outcomes is sentiment analysis.

It is a type of natural language processing in which text data is analyzed to detect the emotional state of people or groups. It can be used to analyze various types of text data, including social media posts, customer reviews, and other online content, to gain insights into the emotional state of individuals and communities. By analyzing the sentiment of text data, sentiment analysis can provide valuable insights into mental health issues, including identifying individuals who may be struggling with mental health issues, tracking mental health trends and patterns, and providing timely support and resources to those in need.

Several research has looked into the application of sentiment analysis in the field of mental care. In one study released in the Journal of Medical Internet Research, for example, researchers utilized sentiment analysis to analyze social media data and identified those at risk of depression and anxiety[1]. The study found that sentiment analysis was a useful tool for detecting emotional distress in social media data and could help to identify individuals who may be in need of mental health support.

Similarly, a study published in the Journal of Medical Systems explored the use of sentiment analysis in online reviews of mental health services [2]. The study found that sentiment analysis could be used to detect negative sentiment in online reviews and provide valuable insights into the experiences of individuals accessing mental health services. This information could be used to improve the quality of mental health services and provide targeted support to individuals in need.

Another study published in the IEEE Journal of Biomedical and Health Informatics investigated the use of sentiment analysis in analyzing chat logs of therapy sessions [3]. The study found that sentiment analysis could be used to identify changes in emotional state over time, providing valuable insights into the effectiveness of therapy interventions.

Furthermore, sentiment analysis has been applied in chat apps to enhance mental health results. Researchers employed sentiment analysis to analyze chat discussions and give rapid help to people showing negative feelings, the Journal of Medical Internet Research recently published a research on this [4]. The study found that sentiment analysis could be used to detect warning signs of mental health issues and provide timely support and resources to users.

In addition, sentiment analysis has been used in chat applications to improve mental health outcomes for specific populations, such as college students. The research published in the IEEE Journal of Biomedical and Health Informatics, researchers used sentiment analysis to analyze chat conversations between college students and mental health professionals and provided targeted support and resources to those in need [5].

In conclusion, sentiment analysis has shown promise in improving mental well-being outcomes by providing valuable insights into mental health issues and identifying individuals who may need support. It can also be used to monitor for warning signs of mental health issues in chat conversations and provide immediate support and resources to users. The studies reviewed in this paper demonstrate the potential benefits of sentiment analysis in mental health care and highlight the need for further research to explore the effectiveness of sentiment analysis in improving mental health outcomes.

METHODOLOGY

Deep learning or machine learning methods are employed in our suggested methodology. The implementation process involves the following steps:

1. Data collection: Collect data from various sources such as social media platforms, chat applications, or other online sources.
2. Data pre-processing: Cleaning the collected data to remove unwanted characters, stop words, punctuations, and other irrelevant information.
3. Data Labelling: Labelling the data with sentiments such as positive, negative, or neutral.
4. Feature extraction: Features from the pre-processed data are extracted such as bag-of-words, TF-IDF, or word embeddings.

5. Model training: Training a machine learning or deep learning model on the labelled data and extracted features to predict the sentiment of new text data.
6. Model evaluation: The model's performance is assessed using a variety of criteria, including precision, accuracy, and F1-score.
7. Deployment: Deploying the model in the production environment to analyze the sentiment of new text data.

In chat applications, sentiment analysis can be implemented by integrating the model into the chat system to analyze the sentiment of the messages exchanged between users. The system can then provide feedback or suggestions to the users based on the sentiment of their messages. For example, if the system detects negative sentiment in a message, it can suggest ways to improve the user's mood or provide resources for mental health support. This can help users become more self-aware of their emotions and improve their mental well-being.

For this study we have selected to work on RNN classification.

An RNN classifier is a form of neural network that is frequently employed in tasks involving text categorization and natural language processing. RNNs are excellent at handling sequential data, like text, because they are built to remember prior inputs and use them to guide the current prediction. In text classification, RNN classifiers are typically trained on a large dataset of labelled text data to learn patterns and relationships between words and their corresponding sentiment or category labels. The input text is processed as a sequence of words or tokens, and the RNN model considers into account the previous words in the sequence to predict the current sentiment or category label.

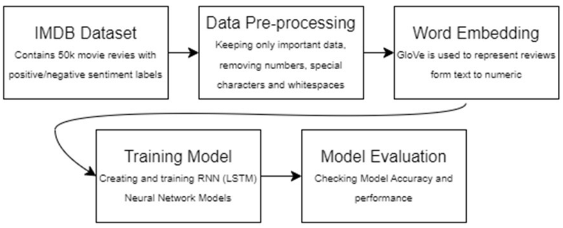


FIGURE 1. Flowchart of Methodology

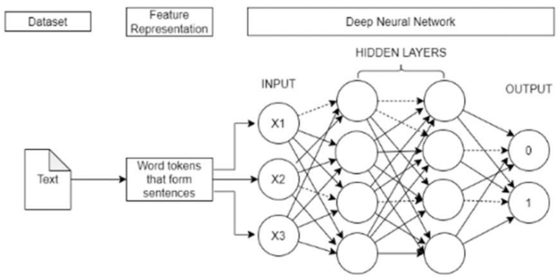


FIGURE 2. Working of RNN

Dataset

The IMDB dataset is utilized for this study and contains 50000 reviews of movies shared by individuals expressing their opinions and sentiments about the movie they watched. The data contains information about the sentiment of the review (positive or negative), user information, movie information, and the text of the review. This dataset is composed of 50% positive reviews and 50% negative reviews, making it balanced.

review	sentiment
50k unique values	2 unique values
One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. The...	positive
A wonderful little production. The filming technique is very unassuming- very old-time-B...	positive
I thought this was a wonderful way to spend time on a too hot summer weekend, sitting in the air con...	positive
Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his par...	negative

FIGURE 3. Sample of dataset

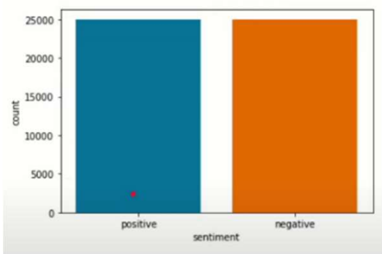
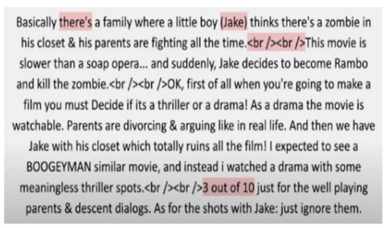


FIGURE 4. Graphical Representation of positive and negative review in dataset



Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his parents are fighting all the time.

This movie is slower than a soap opera... and suddenly, Jake decides to become Rambo and kill the zombie.

OK, first of all when you're going to make a film you must Decide if its a thriller or a drama! As a drama the movie is watchable. Parents are divorcing & arguing like in real life. And then we have Jake with his closet which totally ruins all the film! I expected to see a BOOGEYMAN similar movie, and instead i watched a drama with some meaningless thriller spots.

3 out of 10 just for the well playing parents & descent dialogs. As for the shots with Jake: just ignore them.

FIGURE 5. Sample Review

Data Pre-processing

Only the essential columns are kept such as review and sentiment. Figure 5 shows one of the reviews from the dataset. It contains special characters such as commas, brackets, arrows, numbers etc.

The following steps are taken care of:

1. Removal of html tags
2. Removal of punctuations and numbers
3. Single character removal
4. Remove multiple spaces.
5. Remove Stop words.

Then sentiment labels are converted into 0 and 1. Where 0,1 represents positive, negative respectively.

Splitting of Data

For this study, the dataset is split into training and test sets in an 80:20 ratio.

The training set will be used to train the deep learning models, and the test set will be used to evaluate their performance.

Preparing Embedding Layer

Tokenization

The reviews are tokenized in addition. a method of dissecting a text into its component words or tokens. It is a crucial stage in the preparation of data for NLP tasks like sentiment analysis, text classification, machine translation, and others.

In tokenization, a document or a sentence is split into smaller units, which are typically words or sub-words. These units are then used as input to the subsequent stages of an NLP pipeline. There are several methods of tokenization, including whitespace-based tokenization, rule-based tokenization, and statistical tokenization.

It is used as the first layer for the Model as the embedding layer expects the words to be in numeric form. It is with the help of tokenizer function from keras pre-processing text library.

```
#print before tokenization
X_train[0]

'kept asking many fights screaming matches swearing general mayhem permeate minutes
comparisons also stand think one dimensional characters little depth virtually impo
ssible care happens badly written cyphers director hang multicultural beliefs topic
done much better dramas tv cinema must confess really one spotting bad performances
film must said nichola burley heroine slutty best friend wasim zakir nasty bullying
brother absolutely terrible know acting school graduated apply full refund post has
te samina awan lead role manages impress cast called british talent probably never
hear least hope next time hire different scout another intriguing thought hideously
fashionable soundtrack featuring likes snow patrol ian brown keane bit music fan fa
miliar artists output recognise tracks movie apart omnipresent run sides anyone get
many many musical montages telegraph suppose feel accompanied startlingly original
images couples kissing swollen lake canoodling doorways problem none songs convey ...
'
```

FIGURE 6. Before Tokenization

```
#print after padding
X_train[0]

array([32471,   121,   100,   207,   919,  2651,  8472,  8963,  3832,
           1,   843, 16494,   364,  2429,   136,    15,    32,    32,
         553,  9712, 12506,  1197,   119,  4177, 17434,   102,   990,
        4199,  5069, 24901,  2138, 45770, 29970,   320,   486,   577,
        2775,  1114, 16071,  3430,    59,  1343,  1083,  1449,   732,
        1091,   185,  2714,   288,   489,   154,  3373,   845, 13300,
        1103,   111, 17435,   113,     3,    89,  1296,   935,  4726,
       20778, 28006,   117,   238,     3,  3406,    92,  8003,   729,
        3289,  1584,    63,    93,    76,    33,   340,   114,   208,
        3290,  1707,  4526,  1231,     2,  3334,  1852,  2242,   339,
         396,  2386, 17436,  2559,  1608,   449,  3609,    96,   232,
        3671], dtype=int32)
```

FIGURE 7. After Tokenization and Padding

Padding

All the reviews are converted to a fixed length of 100. Adding multiples 0 after the text, making the review size 100.

Word Embedding

After tokenization, the tokens are then converted into numerical vectors, which can be processed by the RNN model. Word embedding techniques such as Word2Vec or GloVe are commonly used to perform this step.

The idea behind word embedding is to map each word in a vocabulary to a high-dimensional vector, where the vector represents the meaning or semantic relationships of the word. Word embeddings are used to convert text data into a numeric representation that can be used as input to machine learning models.

There are many methods to create word embeddings, in this case we shall use GloVe Word Embedding with 100-dimensional vector for text to numeric transformation of our training data. GloVe is a neural network-based method that takes a large corpus of text as input and produces a set of word vectors as output. The basic idea behind GloVe is to train a neural network to predict the context of a given word in a sentence. The neural network is trained to predict the probability of each word in the vocabulary being the context word given a particular input word.

Once the neural network is trained, the weights of the hidden layer are used as the word embeddings. These embeddings are high-dimensional vectors that capture the semantic relationships between words in the vocabulary as shown in the figure. For example, words that are used in similar contexts will have similar vector representations. Many NLP applications, including sentiment analysis, text categorization, and language translation, can make use of word embeddings as features.

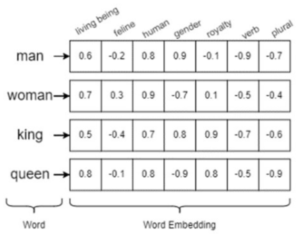


FIGURE 8. Word Embedding

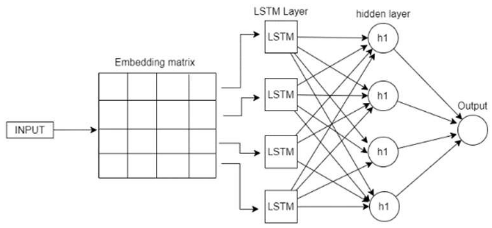


FIGURE 9. Role of LSTM

Preparing Model

The proposed study uses the RNN (LSTM).

LSTM stands for Long Short-Term Memory, and it is a form of RNN architecture. LSTM, like other RNNs, contains a feedback loop that enables it to interpret sequential incoming data. LSTM, on the other hand, has a more complicated structure which enables it to manage long-term relationships while avoiding the issue of vanishing gradients that is common in standard RNNs.The main difference between LSTM and conventional RNNs is that LSTM has a memory cell that can store data for an extended period. A forget gate, an output gate, and three gates make up this memory cell. These gates regulate the flow of information into and out of the memory cell, allowing the LSTM to selectively recall or forget information over time.The input gate chooses whether to keep recent data in the memory cell. The forget gate decides whether to erase the earlier memory. The output gate is responsible for controlling the information that exits the memory cell and passes on to the next stage of the procedure.Because of LSTM's capacity to selectively retain or forget information over time is especially effective for applications such as speech modeling, speech detection, and sentiment analysis, when the input data may have complicated and long-term connections.

Evaluation

The accuracy on training set is **90%** while on the test set is **86.4%.**

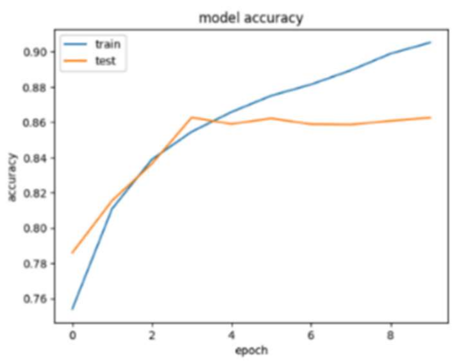


FIGURE 10. Training vs Test Accuracy Graph

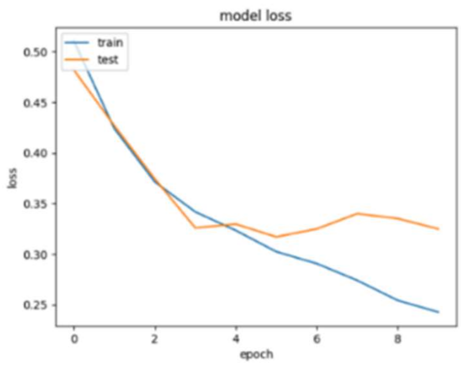


FIGURE 11. Training vs Test Loss Graph

SYSTEM IMPLEMENTATION

Integrating ML Model into Android

Integrating a machine learning model into an Android app can be challenging due to the limited resources available on mobile devices. However, TensorFlow Lite, a framework developed by Google, provides a solution to this problem by allowing developers to deploy optimized machine learning models on mobile and embedded devices. By converting a trained machine learning model into a TensorFlow Lite format, it can be integrated into an Android app to perform tasks such as image recognition, natural language processing, and sentiment analysis, with improved performance and reduced memory consumption. The TensorFlow Lite library also provides various tools and APIs that make it easier for developers to deploy and run machine learning models on Android devices.

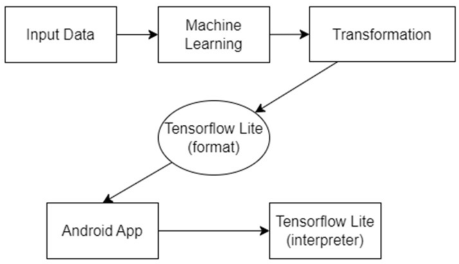


FIGURE 12. TensorFlow lite implementation

About Our Android App

Chat App integrated with sentiment analysis to monitor users' mental wellbeing through recent conservations.

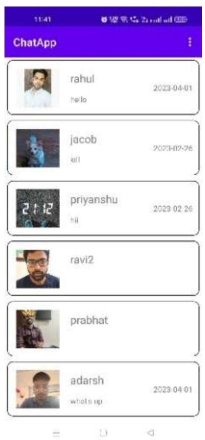


FIGURE 13. Main Page

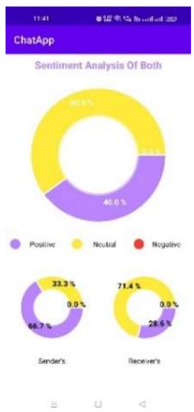


FIGURE 14. Sentiment Analysis of Individual chats

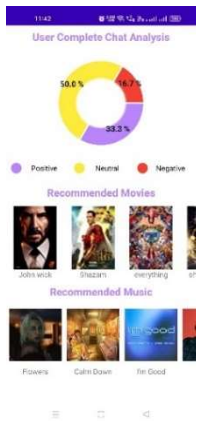


FIGURE 15. Complete Chat Analysis and Recommendations

Selective Movies and Music is recommended by analysing user's recent conservations to uplift user's emotions.

DRAWBACKS

While there are several benefits of using a smartphone Android app for self-monitoring emotional and mental well-being, there are also some potential drawbacks to consider. Here are a few:

1. Privacy concerns: Collecting and storing personal data on a smartphone app can raise privacy concerns. If the app is not designed and maintained properly, users' sensitive information, such as their emotional state, could be at risk of being accessed or used by unauthorized parties.
2. Dependence on technology: Some people may become overly reliant on the app and neglect to develop coping mechanisms to manage their emotional and mental well-being. Relying too heavily on technology to manage emotions could prevent individuals from learning how to effectively deal with their emotions in real-world situations.

3. Inaccurate results: NLP algorithms used in the app may not always accurately detect emotional and mental states. In some cases, users may receive inaccurate feedback or recommendations based on these algorithms, which could lead to inappropriate or ineffective treatments.
4. Limited accessibility: Not everyone has access to a smartphone or the internet, which could limit the reach of the app. This could result in certain individuals being unable to benefit from the app's features.
5. Lack of human interaction: While the app can offer useful comments and ideas, it cannot offer the same degree of psychological assistance or empathy as a real therapist or counsellor.

CONCLUSION

In conclusion, developing an Android chat app with machine learning sentiment analysis using RNN to analyze user's chats and recommend movies and music based on their emotional state can be a valuable tool in promoting mental well-being. With the increasing importance of mental health awareness and the rise in the use of technology, this app can offer a convenient and accessible solution for those looking to improve their mood and emotional state. By leveraging the power of machine learning, the app can provide personalized recommendations based on the user's emotions and help them make informed choices about their entertainment options. Overall, this project has the potential to make a positive impact on the mental health of its users and contribute towards a healthier and happier society.

REFERENCES

1. De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. Journal of Medical Internet Research, 15(2), e28. doi: 10.2196/jmir.2301
2. Nambisan, P., Gustafson, D. H., Hawkins, R., & Pingree, S. (2016). Social support and responsiveness in online patient communities: impact on service quality perceptions. Journal of Medical Systems, 40(6), 136. doi: 10.1007/s10916-016-0489-3
3. Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. Association for Computational Linguistics, 1(1), 1-10. doi: 10.18653/v1/w15-0101
4. Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., & Gonzalez, G. (2016). Utilizing social media data for pharmacovigilance: A review. Journal of Biomedical Informatics, 54, 202-212. doi: 10.1016/j.jbi.2015.12.001
5. Mishra, V., Salehi, N., & Brown, W. (2017). Investigating sentiment analysis techniques for chat-based counseling. IEEE Journal of Biomedical and Health Informatics, 21(1), 230-238. doi: 10.1109/JBHI.2016.2632845
6. Alghadir, A. H., Alyousef, S. M., Al-Bader, B. A., & Gabr, S. A. (2019). Self-monitoring of physical activity using pedometer and smartphone app: A feasibility study. Journal of physical therapy science, 31(11), 879-883
7. Baumel, A., Muench, F., & Edan, S. (2019). Digital self-monitoring in behavioral health: a systematic review. Journal of medical Internet research, 21(5), e13226.
8. Ben-Zeev, D., & Scherer, E. A. (2018). Analyzing Text Message-Based Health Interventions: Systematic Review. JMIR mHealth and uHealth, 6(10), e10268.
9. Bickmore, T. W., Gruber, A., & Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Education and Counseling, 59(2), 21-30.
10. Boschen, M. J., Casey, L. M., & Oei, T. P. (2013). Mobile app-delivered cognitive behavioural interventions for anxiety disorders: A systematic review. Journal of anxiety disorders, 27(3), 267-273.