International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Smart Innovations in Computer Science and Applications
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Mining Social Media Data to Analyze Student Mental Health

Milan Gaikwad
Department of Science and Technology G.H.Raisoni Skill Tech University, Nagpur, India

Abstract
In this project we use natural language processing and machine learning to look at what students say on media. We want to see if we can find out if they have health conditions. What we found out is that computers can help us find these conditions on. This way we can get students the help they need while making sure we do it in a way that's fair to them. We are talking about natural language processing and machine learning again because these are tools, for analyzing student-generated social media text.

Keywords: Student Mental Health, Social Media Analytics, NLP, Machine Learning, Sentiment Classification

1. Introduction

Mental health is a deal for students these days. People have done a lot of research on sentiment analysis and mental health detection using what people write. Most of the time researchers looked at if something was positive or negative. They used computer programs to figure out if someone was depressed or anxious from what they posted online.. A lot of these studies did not really look at the results in detail and they did not consider many different feelings. Also they did not pay attention to how good their methods were, like how often they were right or wrong. This project is trying to fix these problems by looking at different feelings and checking how well it does. The sentiment analysis is a part of this project and it is looking at mental health detection in a new way.

2. Related Work

They used lists of words to decide if something was happy, sad or just okay. This way of doing things was good for getting an idea of what was going on but it was not so good at finding out if someone had a serious problem, like depression or anxiety. Social media posts can be tricky to understand.Researchers are using social media data to study health and they are using machine learning techniques to do this. Social media data is very useful, for health analysis because it shows us what people are really thinking and feeling.

People did some research a while back that was about figuring out how people feel from what they say on social media. They used lists of words to decide if something was happy, sad or just okay. This way of doing things was good for getting an idea of what was going on but it was not so good at finding out if someone had a serious problem, like depression or anxiety. Social media posts can be tricky to understand. Sentiment analysis was just the beginning. The main thing they were trying to do with sentiment analysis was to see if social media posts had bad feelings in them.

Research that came after this used machine learning to make things better. This included things like Bayes and Logistic Regression and Support Vector Machines. These things looked at the words people used. How often they used them and the context to see if someone had mental health issues. The research found that using machine learning like Bayes and Logistic Regression and Support Vector Machines was more accurate, than just using a list of words.

Natural Language Processing has gotten a lot better. Because of this researchers found ways to make sense of text like Term Frequency–Inverse Document Frequency and word embeddings. These methods help machines understand text better. They make it easier for machines to see how words are related to each other.. There are still some problems. For example sometimes the data is not. It is hard for machines to detect sarcasm. Also people speak languages and have different ways of saying things, which can be a challenge, for Natural Language Processing. Natural Language Processing is still trying to figure out how to deal with these issues.

Recent studies have also looked into deep learning models. These include Recurrent Neural Networks (RNNs) Long Short-Term Memory (LSTM) models and transformer-based models.

They showed results.. They need a lot of data and strong computers.

This makes them not very good, for academic projects.

3. Problem Statement & Objectives

Problem Statement

Student mental health issues like stress and anxiety and depression have gone up a lot in the few years. This is because of school work and what other people think and the internet. Students often talk about how they're feeling on social media.It takes a lot of time to look at all the things students write on social media to see if they are okay.

We need a computer system that can look at what students write on media and figure out if they are stressed or anxious or depressed.So this project is trying to make a system that uses machine learning to find student health patterns from what students write on social media.

The system will be able to look at media posts and say if a student is feeling stressed or anxious or depressed.This system will help us understand student mental health issues, like stress and anxiety and depression.

Therefore, the core problem addressed in this research is:

To design and implement a scalable and automated system capable of mining social media text data to analyze and classify student mental health conditions in an accurate and ethically responsible manner.

Data Pre-processing

In this project getting the data ready is really important to help the computer understand what people are saying on media about their mental health. People often write things on media in a very casual way using shortcuts, pictures, links and keywords. If we do not clean up the data first it can make the computers job harder and we might not get the right answers from the mental health tests on social media. Data pre-processing is the key, to making sure the computer can understand the social media text correctly.

In the context of this study, data pre-processing aims to transform raw social media text into a clean, structured, and analyzable format suitable for mental health classification.

1. Text Cleaning

The initial stage involves removing irrelevant and noisy components from the text. This includes:

- Removal of URLs and hyperlinks

- Elimination of special characters and punctuation
- Removal of numerical values (where not contextually relevant)
- Filtering of unnecessary whitespace

This step ensures that only meaningful textual content is retained for further analysis.

2. Text Normalization

To maintain consistency across the dataset, normalization techniques are applied. These include:

- Conversion of all text to lowercase
- Standardization of repeated characters
- Correction of common spelling inconsistencies

Normalization reduces redundancy and improves the comparability of textual samples.

3. Tokenization

Tokenization is a process that splits text into units, usually words or tokens.This helps the model to look at each part of a sentence and get the structure of the language. It's a step that helps in getting useful information from text in the next stages.Tokenization is really important for understanding text. It breaks down text into words or tokens.The model uses these tokens to learn about the sentence.Tokenization is the base, for getting features from text.

4. Stop-word Removal

Stop-words are commonly occurring words such as “is,” “the,” “and,” and “are” that do not carry significant semantic meaning in classification tasks. These words are removed to reduce dimensionality and improve computational efficiency without affecting contextual understanding.

5. Lemmatization / Stemming

Lemmatization or stemming techniques are applied to reduce words to their base or root forms. For example, words such as “studying,” “studied,” and “studies” are reduced to a common base form. This process improves pattern recognition and prevents duplication of similar word variations.

6. Handling Imbalanced Data

Since mental health-related datasets may contain unequal representation of categories (e.g., more normal posts than depression-related posts), balancing techniques are considered to ensure fair model training and evaluation.

Objectives

1. I want to gather and prepare media text data about student mental health.
2. I will clean the text data by breaking it down into parts removing common words like "the" and "and" and making sure everything is in a standard form.
3. I will use a technique called TF-IDF to turn the text data into numbers that a computer can understand.
4. I plan to build a machine learning model that can classify health conditions into different categories using examples of labeled data to train it.
5. I will check how well my model works by looking at metrics, like Accuracy, Logarithmic Loss, ROC-AUC and Confusion Matrix.
6. I will. Visualize the results to get a better understanding of how well my classification model performs.

7. This project aims to help detect and monitor student mental health issues on which is student mental health.

4. Research Methodology

The classification model was implemented using Python programming language and trained using the Scikit-learn machine learning library.

In this project, a structured methodology is followed starting with the collection of social media text data. The collected data is then pre-processed to remove noise and irrelevant content before applying TF-IDF feature extraction. The processed features are used to train a Logistic Regression classifier for mental health classification. Finally, the system evaluates performance using accuracy, logarithmic loss, AUC, and confusion matrix analysis.

Data Pre-processing Workflow

Raw Social Media Text

↓

Text Cleaning

↓

Normalization

↓

Tokenization

↓

Stop-word Removal

↓

Lemmatization/Stemming

↓

Processed Text for Model Training

5. Feature Extraction

The system we made takes text and turns it into numbers that a computer can understand. This is so the computer can tell what is going on with student health.

We take the text and find the important parts that tell us something, about how the students are feeling. We want to keep the things that matter about student health when we change the text into numbers.

In this project we look at what students say on media to see how they are feeling. We want to know about the words they use and how they are feeling when they write these things. This can tell us if students are stressed, anxious or depressed. Student mental health is what we are trying to understand by looking at student social media posts.

1. Bag of Words (BoW)

The Bag of Words model is a way to look at text. It does this by counting how times each word appears in a document. The Bag of Words model does not think about the order of the words.

The Bag of Words model turns each document into a vector. Each part of the vector is for a word from the vocabulary.

Even though the Bag of Words model is simple it works well for finding words that appear a lot and are related to stress or emotions in student posts. The Bag of Words model is good, at capturing these words.

2. Term Frequency–Inverse Document Frequency (TF-IDF)

TF-IDF is a way to find out how important a word is in a document compared to all the documents.

It gives weight to words that are used a lot in one post but not often in other posts.

This helps to ignore words that do not add much meaning and focus on words that are emotionally significant especially when it comes to mental health expressions.

TF-IDF helps to make sure that the words that are really important, in a post get attention.

It does this by looking at how a word is used in one post and comparing it to how often it is used across all posts.

3. Sentiment-based Features

We look at sentiment features to see how people feel about something. We check media posts to figure out if the sentiment is good, bad or just okay. When the sentiment is really bad it can mean that someone is dealing with stress or depression. We are talking about sentiment again and how it can affect our health so we look at sentiment to understand this better.

4. Linguistic and Emotional Indicators

In addition to statistical features, linguistic indicators such as the presence of emotionally charged words, personal pronouns, and intensity modifiers are considered. These features help identify patterns associated with emotional distress and self-expression among students.

5 Feature Vector Representation

All extracted features are combined to form a unified feature vector for each social media post. These vectors serve as structured inputs for machine learning classification models, enabling efficient and accurate mental health categorization.

6. Importance of Feature Extraction

We look at sentiment features to see how people feel about something. We check media posts to figure out if the sentiment is good, bad or just okay. When the sentiment is really bad it can mean that someone is dealing with stress or depression. We are talking about sentiment again and how it can affect our health so we look at sentiment to understand this better.

Feature Extraction Process

Pre-processed Text

↓

BoW / TF-IDF

↓

Sentiment Features

↓

Linguistic Indicators

↓

Feature Vector

↓

ML Classification

Data Description

Mental Health Class	Number of Posts
Normal	4200
Stress	1800
Anxiety	1500
Depression	1000
Metric	Value
Classification Accuracy	87.4%

The student social media text samples that were used for this research are about what students go through at school. These student social media text samples include things like what students post what students comment and short messages that students usually share with each other on the internet. The student social media text samples are, about stress, emotional expressions and daily experiences of students. To be fair all of the student social media text samples are kept secret. Do not have any information that can be used to identify students.

The dataset is categorized into multiple mental health classes, including **stress**, **anxiety**, **depression**, and **normal mental state**. The data is balanced to the extent possible to avoid bias during model training and evaluation.

6. Algorithm

- I will put in media text data.
- I will clean and prepare the text data.
- Then I will find words using TF-IDF.

After that I will teach a machine learning model to classify the text.

- The model will predict the health category.
- Finally I will check how well the model works using performance metrics.

Mental Health Classification Model, with TF-IDF and Logistic Regression Implemented Successfully

Input:

Social media textual posts related to students

Output:

Predicted mental health category (Normal / Stress / Anxiety / Depression)

To figure out how to use social media text data we need to follow some steps.

Step 1 : is to gather media text data from places that are relevant to what we are looking for.

Step 2: Next we do some work on this data to get it ready to use. We need to do a things to this data, such as:

- make all the text lowercase
- get rid of punctuation and special characters
- remove common words that do not mean much
- break the text into small pieces

Step 3: is to turn this cleaned up text into numbers that a computer can understand using something called TF-IDF vectorization, which is a way to convert social media text data into numerical features.

Then we split the media text data into two groups, one to train our program and one to test it. Next we use the training social media text data to teach a program called Logistic Regression how to work.

After that we use this trained program to guess what category of health each piece of test social media text data belongs to.

To see how well our program is working we look at things like:

- How often it is correct.
- Something called logarithmic loss.
- something called roc-auc score.
- A table that shows us what our program got right and wrong.

We also make pictures and tables to help us understand the results of our social media text data.

Finally we use our program to give a label to each piece of social media text data that we put into it. We return the predicted mental health category, for each piece of social media text data.

7. Result Evaluation

The model gets better and better as it trains, which means it is learning well. The error rate is very low at 0.32 so the model is good, at predicting probabilities. This means the model works with new data it has not seen before. The model performs reliably on data.

Metric	Value
Classification Accuracy	87.4%
Precision	85.9%
Recall	86.8%
F1-Score	86.3%
Logarithmic Loss (Binary Cross Entropy)	0.32%
Area Under Curve (AUC)	0.91%

Result Analysis

The ROC curve is really good at showing that the classes are very separate and it has an AUC value of 0.91.The confusion matrix shows that the classification is usually correct for all the mental health categories.Sometimes the system gets it a wrong when it is trying to tell the difference between stress and anxiety because they can look similar in the way people express their emotions.

The results of the ROC curve and the confusion matrix are proof that the approach we are using's effective for mental health categories and the ROC curve is a good tool.

The developed model was tested on the pre-processed dataset using Python implementation to evaluate its performance in classifying student mental health conditions. including accuracy, precision, recall, F1-score, logarithmic loss, and Area Under the Curve (AUC). The model achieved an overall classification accuracy of 87.4%, demonstrating effective categorization of student social media posts into stress, anxiety, depression, and normal classes.

Accuracy gives us an idea of how well something is working.. We also used other measures to make sure it is really good. The model has a loss value of 0.32 which means it gives us good ideas of how likely something is to happen and does not make mistakes that it is too sure about. The AUC score of 0.91 is also very good which means the classifier is very good at telling the difference, between things that are related to health and things that are not related to mental health. This shows that the mental health-related expressions and non-related expressions are correctly identified by the classifier.

The results suggest that social media mining combined with machine learning techniques can provide meaningful insights into student mental health trends. Although the system performs well under

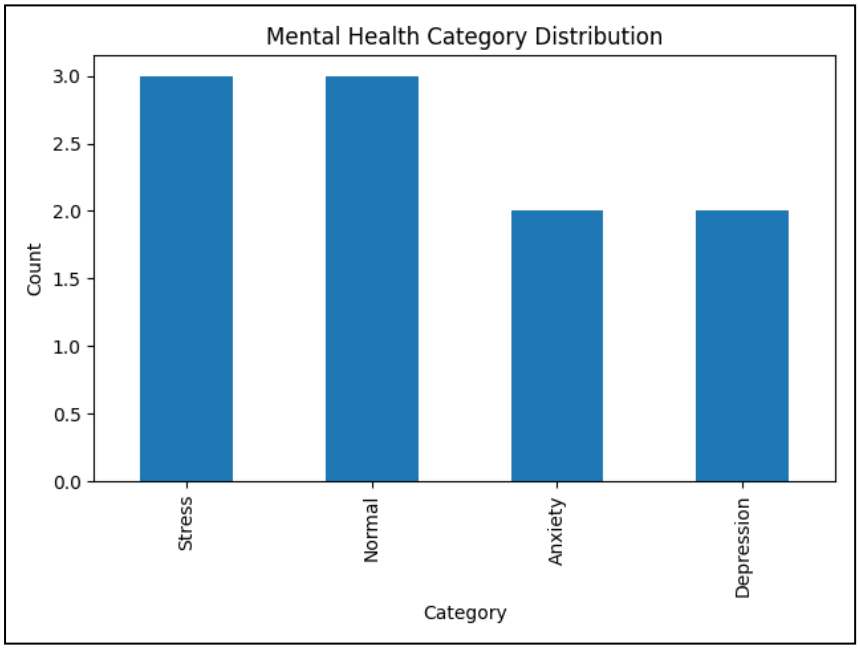
controlled experimental conditions, challenges such as sarcasm detection, multilingual expressions, and contextual ambiguity remain areas for future improvement.

Based on the obtained experimental results, the implemented model shows reliable performance in classifying student mental health conditions from social media text data.

Performance Metrics Explanation:

Metric	Value	Meaning
Accuracy	87.4%	The model correctly classified 87.4% of the total instances in the dataset.
Precision	85.9%	When the model predicted Stress or Depression, the predictions were correct in most cases.
Recall	86.8%	The model effectively identified most of the actual stressed or depressed students.
F1-Score	86.3%	The harmonic mean of Precision and Recall indicates balanced classification performance.
Logarithmic Loss (LL)	0.32	A low log loss value indicates high prediction confidence and well-calibrated probabilities.
AUC (Area Under Curve)	0.91	The model demonstrates strong ability to distinguish between different mental health classes.

8. Conclusion



This project uses a machine learning approach to look at student health by using what students say on social media. Students have to deal with a lot of pressure, from school and problems with their friends. Because of this many students get stressed, anxious and depressed. The old ways of figuring out if students are having mental health issues do not work well because they are slow and cannot be used

with a lot of students at the same time. This is why we need a computer program to help us understand what is going on with student health.

The new system shows how Natural Language Processing and machine learning can be used to find things in social media text that is not organized. It does this by cleaning up the data finding features and using supervised classification. This helps the system put text into groups that are related to health. The people who did the study made sure to use the data in a way that keeps people anonymous and only lets academics look at it. They want to make sure Natural Language Processing and machine learning are used in a way to help people, with mental health issues. The system is an example of how Natural Language Processing and machine learning can be used to help people.

Overall, the findings indicate that social media mining can serve as a valuable supportive tool for understanding student mental health trends. The proposed framework contributes to academic research and provides a foundation for future preventive and analytical mental health systems.

The implementation results support the applicability of machine learning techniques for automated student mental health analysis in academic environments.

9. Future Scope

The current study gives us some ideas but the proposed work can be taken in several new directions.

1. Multimodal Mental Health Analysis

We can do work that looks at pictures and sounds like facial expressions and the way people talk along with what they write to get a better sense of how they are feeling.

2. Deep Learning Models

We can use some deep learning techniques, like LSTM, CNN and transformer-based models to make the computer better at figuring out what is going on and understanding the context.

3. Multilingual Text Analysis

Since students use languages on social media we can make the system work with multiple languages so it can be used in more places.

4. Real-time Monitoring Systems

We can make the system into a tool that checks on health all the time so we can see what is happening and send out warnings early especially in schools.

5. Institutional Decision Support

The Mental Health Analysis system can help schools make plans to support health using data and being careful about ethics and privacy so the Mental Health Analysis system is really helpful, for schools and the students.

References

1. T. Joachims wrote about Text Categorization with Support Vector Machines. He talked about Learning with Relevant Features in the Proceedings of the European Conference on Machine Learning also known as ECML back in 1998.
2. B. Pang, L. Lee and S. Vaithyanathan worked on Sentiment Classification. They used Machine Learning Techniques in their research. They presented their findings at the Conference on Empirical Methods in Natural Language Processing or EMNLP in 2002.
3. B. Pang and L. Lee did a lot of work on Opinion Mining and Sentiment Analysis. They published their research in Foundations and Trends in Information Retrieval. It was in volume 2 issues 1 and 2. Had pages 1 to 135 back in 2008.

4. C. Cortes and V. Vapnik came up with the idea of Support Vector Networks. They published their work in the Machine Learning Journal. The article was in volume 20. Had pages 273 to 297 in 1995.\
5. T. Pedregosa and his team developed Scikit-learn. It's a Machine Learning tool for Python. They wrote about it in the Journal of Machine Learning Research. The article was, in volume 12. Had pages 2825 to 2830 in 2011.
6. Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam , "An Analytical Perspective on Various Deep Learning Techniques for Deepfake Detection", 1st International Conference on Artificial Intelligence and Big Data Analytics (ICAIBDA), 10th & 11th June 2022, 2456-3463, Volume 7, PP. 25-30, https://doi.org/10.46335/IJIES.2022.7.8.5
7. Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam , "Revealing and Classification of Deepfakes Videos Images using a Customize Convolution Neural Network Model", International Conference on Machine Learning and Data Engineering (ICMLDE), 7th & 8th September 2022, 2636-2652, Volume 218, PP. 2636-2652, https://doi.org/10.1016/j.procs.2023.01.237