Proceedings of International Conference on "Recent Innovations in Engineering, Technology & Science for Sustainable Development (RIETSSD-2026)"
Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Special Issue | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.20673758

A Comprehensive Study and Analysis of Machine Learning Algorithms Using Sentiment Analysis for Monitoring Mental Health Issues among Students

Sushma Bhoir[1], Dr. Abid Hussain[2]

Research Scholar, School of Computer Application & Technology, Career Point University, Kota (Raj.), India

Research Supervisor, School of Computer Application & Technology,

Career Point University, Kota (Raj.), India

[1]sushshivkar@gmail.com, [2]abid.hussain@cpur.edu.in

Abstract:

The psychological problems in students have increased as a significant issue all over the world due to excessive demand in academics, a competitive environment, social isolation, and hyper-digital exposure. Psychological distress cannot be identified at its early stage because students normally do not like to discuss their emotional difficulties openly. Digital communications are also involved since text emotion is more expressed by students' social media, learning management systems, online forums and feedback platforms. In the process of sentiment analysis, the machine learning algorithms give us a hint on how to automatically recognize the patterns of emotions in the text data. In this paper we discussed detailed research and compared different machine learning algorithms that are used in sentiment analysis for mental health issues in students. The study compares classical, kernel based, and collective learning algorithms based on multiple performance metrics. The proposed research also provides a conceptual framework for automated mental health monitoring systems at educational institutions. The results reveal that ensemble and advanced classifiers are more efficient than basic models in recognizing negative emotional states. So that early mental health support facility can be provided.

Keywords: Sentiment Analysis, Machine Learning, Student Mental Health, Text Mining, Emotion Detection, Educational Data Mining.

I. Introduction

Mental health contributes largely to the students' academic performance, emotional satisfaction and overall growth of a person. In the recent past, educational institutions all over the world have experienced a tremendous rise in mental health issues among their learners which includes stress, anxiety, depression, emotional exhaustion and burnout. These challenges are often a result of excessive workload on the part of students, examination pressure, career ambiguity, peer competition and social expectations. Thus, even though mental health conditions of students are not a small thing at all, they can often go unnoticed or unreported for such reasons as ignorance, prejudice and the fear of being judged. The current techniques of assessing mental health like counselling sessions and self-report questionnaires, cannot rise up and rely heavily on a volunteer participation system. Most digital platforms, therefore, have changed the way students communicate and express themselves. Students actively use social media networks, online learning platforms, messaging systems and feedback mechanisms in which they share opinions, frustration and emotional experiences. Such a huge amount of textual data provides an opportunity for carrying out computational techniques for an emotion analysis system. The sentiment analysis is a method of natural language processing that utilises machine learning algorithms for classifying the emotional contradiction found in the text. With machine learning algorithms, sentiment analysis can automatically detect and classify the emotions and recognize negative psychological patterns. This research focuses on various machine learning algorithms that are used to analyze and study sentiment analysis as an approach to mental health monitoring among students.

II. Background and Related Work

Sentiment analysis has been applied in a wide range of applications such as product reviews, social network monitoring, political analysis, and customer feedback systems. Sentiment analysis on health-related issues and particularly in mental health monitoring, has been increasingly applied in recent years by researchers. Social media textual data including Twitter, Reddit and student forums has been utilized to assess depressive behaviour, anxiety level, emotional change etc.

Classical machine learning models like Naive Bayes, Support Vector Machines and Random Forest have been used as baseline methods for sentiment classification as they are simple and perform well on structured feature representations. These pose-based methods are usually based on hand-crafted features, such as bag-of-words, TF-IDF features and n-gram.

More recently, deep learning models like Recurrent Neural Networks (RNNs) and Long Short-Term

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

Memory (LSTM) networks have become very popular due to their success at modelling contextual and sequential information in text. Such models can automatically learn text representations from raw data and be applicable to various tasks with better performance than traditional methods. This study contributes by comparing several machine-learning techniques and assessing the feasibility of monitoring mental health disorders in students.

III.REVIEW OF LITERATURE

Mental health issues among students have increased in recent years, yet many academic institutions do not have scalable and effective systems for early detection of emotional distress. Traditional assessment methods like self-report questionnaires and counselling sessions depend on voluntary participation and direct observation, often failing to capture early signs of mental health distress [3].

With the large use of digital platforms, students frequently express emotions through online learning systems, social media, and discussion forums. These platforms generate huge volumes of behavioural and textual data that can offer valuable insights into student well-being. However, such data is barely utilised for mental health monitoring, leading to delayed identification and delayed intervention [1], [7]. Behavioural indicators such as absence from class, reduced participation, and emotionally distressed language in academic content are often overlooked.

Several studies demonstrate the effectiveness of machine learning (ML) techniques in predicting student mental health issues such as stress, anxiety, and depression. Sun et al. [1] and Wang and Yang [7] showed that ML-based models can analyse academic and behavioural data to identify early warning signs. Kumar et al. [4] further explored supervised and unsupervised learning approaches using academic, demographic, and digital activity data, revealing distinct risk patterns. Ensemble-based models have also shown improved predictive performance for proactive mental health assessment [9].

Sentiment analysis and natural language processing have gained attention for extracting emotional signals from textual data. Hermawan et al. [8] demonstrated the use of ML-driven sentiment analysis in mental health chatbot systems, enabling real-time emotional assessment. Additionally, Sharma et al. [2] and Mitra and Rehman [5] emphasised the benefits of incorporating multimodal data such as behavioural, voice, and wearable sensor data to enhance prediction accuracy.

Behavioural analysis has also been explored as part of early warning systems. Yang, Feng, and Jiang [6] proposed a machine learning-based framework for analysing attendance and participation patterns to detect learning difficulties associated with emotional distress. While existing studies highlight the potential of ML and sentiment analysis for early detection, challenges related to scalability, explainability, ethical concerns, and personalised insights remain unsolved [3].

IV. Research Gaps Identified

Despite achieving high predictive accuracy in several existing studies, the current body of research reveals multiple unresolved gaps and challenges that limit the effectiveness and real-world applicability of machine learning–based mental health monitoring systems for students.

A. *Limited Integration of Sentiment Analysis with Multimodal Data*

Most existing approaches rely either on textual sentiment analysis derived from sources such as social media posts and student feedback, or on structured data including demographic information and academic performance records [4], [7]. Very few studies attempt to integrate multiple data modalities—such as behavioural logs, physiological signals, academic records, and emotional cues from text—into a unified framework [2], [5]. The absence of multimodal integration restricts a holistic and accurate understanding of students' mental health states.

B. *Lack of Real-Time and Scalable Early Warning Systems*

Although several studies emphasise the importance of real-time mental health monitoring, the majority of proposed models remain experimental in nature [6], [8]. Fully operational and scalable early warning systems capable of continuous monitoring and proactive intervention within academic institutions are largely absent [1]. The lack of deployment-ready frameworks limits the practical use of these models in real-world educational environments.

C. *Insufficient Focus on Explainability and Ethical AI*

Many machine learning and deep learning models, particularly neural network–based approaches, operate as black-box systems that provide predictions without transparent reasoning [9]. As researched by Soland et al. [3], issues related to interpretability, data privacy, algorithmic bias, and informed consent are not properly addressed. These ethical and explainability concerns significantly barriers to institutional acceptance and student trust.

D. *Limited Comparative Analysis of Machine Learning Models*

While a variety of machine learning techniques have been employed across studies,

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

comprehensive comparative analyses between traditional machine learning algorithms and deep learning models remain limited [1], [4]. Few studies systematically evaluate these models in terms of accuracy, computational efficiency, robustness, and suitability across diverse educational contexts, making it difficult to identify optimal approaches for practical deployment [9].

E. Lack of Personalised Mental Health Insights

Most existing models generate generalised predictions related to mental health conditions such as stress, anxiety, and depression [7], [8]. There is minimal emphasis on providing personalised, context-aware insights that account for individual student characteristics, including academic workload, cultural background, and social environment [5]. This lack of personalisation reduces the effectiveness of targeted interventions and support mechanisms.

V. Research Objectives

A. The primary objective of this research is to study and analyse the effectiveness of various machine learning algorithms combined with sentiment analysis techniques for monitoring and predicting mental health issues among students. The study focuses on identifying emotional patterns in textual data that are indicative of psychological conditions such as stress, anxiety, and related mental health concerns.

B. To support this objective, the research aims to collect relevant datasets related to student mental health from established scientific repositories such as Kaggle, UCI, and other publicly available sources. These datasets may include textual data from mental health forums, student feedback, and academic performance records, with strict adherence to ethical standards and data privacy considerations.

C. Another objective of this study is to perform comprehensive data pre-processing and cleaning on the collected datasets. This includes applying natural language processing techniques such as tokenization, lemmatization, and normalization to prepare textual data for analysis. Feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF) and word embeddings will be employed to represent textual information effectively. Additionally, class imbalance issues present in the datasets will be addressed using resampling techniques such as the Synthetic Minority Oversampling Technique (SMOTE).

CAREER POINT
INTERNATIONAL JOURNAL OF RESEARCH
Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

D. The research further aims to design and implement a unified predictive model that integrates sentiment analysis with machine learning algorithms to monitor students' mental health conditions. The proposed hybrid model will analyse both structured and unstructured data to identify potential mental health risks and may include a simple visualization or alert mechanism to support interpretability and decision-making.

E. Finally, the study seeks to evaluate and compare the performance of the proposed model with existing approaches used for student mental health assessment. The comparative analysis will be conducted using standard evaluation metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). This evaluation will help determine improvements in prediction accuracy, efficiency, and overall model effectiveness.

VI. RESEARCH METHODOLOGY

The goal of this study is to create a machine learning based Early Warning System (EWS) that actively detects early signs of mental distress among students by analysing various data sources as text, behaviour, and sentiment, over time.

The research aims to utilise sentiment analysis techniques to identify individuals who may be suffering from mental illnesses, categorising them based on the intensity of language usage and distinct behavioural elements.

To address the challenges faced in the analysis, including class imbalance, the difficulty in distinguishing closely related categories like suicidal, anxiety, and depression, and the identification of sarcastic and ambiguous statements, with a focus on improving feature engineering and model sophistication for future work.

The specifics are:

A. Identify behavioral, emotional, and linguistic indicators associated with mental health challenges:

Analyze patterns in digital communication like essays, forum posts, chats, etc. Observe fluctuations in class attendance, academic performance trends, and hints of social withdrawal or isolation. The system aims to build a comprehensive risk profile

Career Point International Journal of Research (CPIJR)
©2022 CPIJR ǀ Volume 4 ǀ Issue 4 ǀ ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

that can detect subtle precursors to mental distress.

The target labels that we will be looking at are as follows:

- Normal behaviour
- Stressful
- Anxiety
- Depression
- Bipolar
- Personality Disorder
- Suicidal

B. Develop machine learning models that can predict students at risk of mental health issues with high recall and interpretability, ensuring early intervention opportunities:

The focus will be on minimizing false negatives to avoid missing students who need help, while maintaining sufficient transparency for model decisions to be understandable to counsellors and stakeholders.

C. Integrate multimodal data (text, behaviour) for a more holistic understanding of student well-being, improving model accuracy and generalisation:

The system will combine natural language input, behavioural, and structured psychological assessments into a unified learning framework. This fusion approach is expected to reduce bias and improve model robustness across diverse student populations.

D. Evaluate the effectiveness and practical utility of the system through real-world pilot testing in academic settings and measure its impact on student support outcomes:

Feedback from mental health professionals, students, and faculty will be collected to assess usability, trustworthiness, and intervention outcomes. Pilot deployments will

Career Point International Journal of Research (CPIJR)
©2022 CPIJR ǀ Volume 4 ǀ Issue 4 ǀ ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

help validate system accuracy and ethical acceptability in actual campus environments.

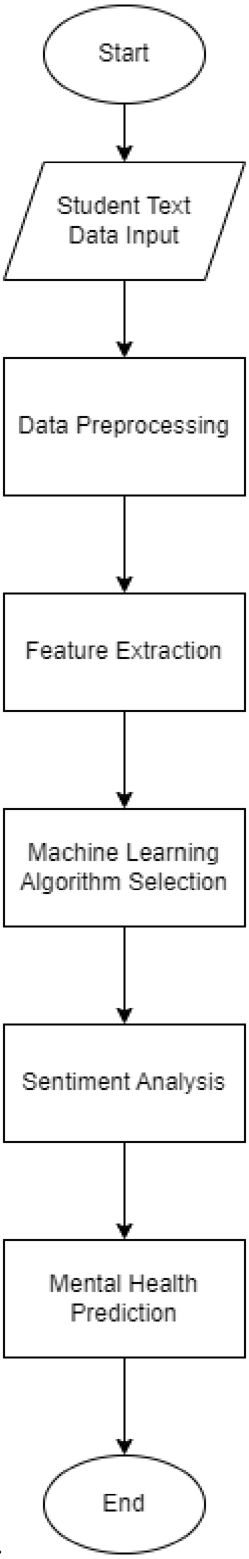

Fig. 1. *Flowchart of the proposed sentiment analysis–based machine learning framework for monitoring mental health issues among students.*

VII. DETAILED PROCESS:

A. Data Acquisition and Preparation

Data collection represents the first and most important step in the proposed methodology. The accuracy, relevance, and quality of the collected data play a vital role in determining the effectiveness of the sentiment analysis and machine learning models. In this work, textual data that reflects students' thoughts, emotions, and personal experiences is treated as the primary source for assessing mental health patterns.

The data is gathered from platforms associated with student interactions, such as:

- Online discussion forums used by students
- Academic feedback forms and surveys
- Social media posts or anonymised student communities
- Chat-based interactions or reflective student responses

These sources are selected because students often express stress, anxiety, frustration, or emotional distress openly through text in such platforms. The collected data mainly consists of unstructured text, which reflects real-world emotional expressions.

During data collection, privacy and ethical considerations are strictly maintained. Personally identifiable information such as names, email addresses, roll numbers, or any sensitive personal data is removed or anonymized. Consent and ethical guidelines are followed wherever applicable, and the collected data is used solely for academic and research purposes. At this stage, the focus is on building a reliable and representative dataset that captures a wide range of emotional expressions related to student mental health.

Currently, the study is limited to data acquisition, and no predictive or classification models have been applied yet. Relevant datasets related to student mental health will be collected from sources like Kaggle, UCI, and other public repositories. These may include survey data, mental health forums, and academic performance records, ensuring ethical data use.

This step involves sourcing and preparing the data needed to build and evaluate the model.

- Text data
- Behavioural data
- Assessments

Career Point International Journal of Research (CPIJR)
©2022 CPIJR ǀ Volume 4 ǀ Issue 4 ǀ ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

B. Data Preprocessing

Once the raw text data is collected, it undergoes a data preprocessing phase to prepare it for further analysis. Raw textual data often contains noise, inconsistencies, and irrelevant information that can negatively affect sentiment analysis results. Therefore, preprocessing is a crucial step in improving data quality.

The preprocessing steps applied in this study include:

- **Text Cleaning:** Removal of unnecessary characters such as punctuation, numbers, URLs, emojis, and special symbols that do not contribute to sentiment interpretation.
- **Lowercasing:** Converting all text to lowercase to maintain uniformity and avoid duplication of words due to case differences.
- **Tokenization:** Breaking down sentences into individual words or tokens to enable further linguistic analysis.
- **Stop Word Removal:** Eliminating commonly used words (such as "is", "the", "and") that do not carry significant emotional meaning.
- **Stemming or Lemmatization:** Reducing words to their root or base form to ensure consistency (e.g., "stressed", "stressing", and "stress" are treated as the same term).

These preprocessing techniques help in reducing noise and dimensionality while preserving meaningful emotional cues in the text. At the current stage, the cleaned and structured dataset is prepared for further processing but has not yet been converted into numerical features.

C. Feature Extraction

Once our data collection, data cleaning and preprocessing is completed, we can move on for feature extraction, where the cleaned textual data is transformed into structured numerical representations suitable for machine learning algorithms. Feature extraction plays a crucial role in sentiment analysis, as the quality of extracted features directly influences the model's ability to capture emotional and psychological patterns in student-generated text.

©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

Since textual data is inherently unstructured, effective feature extraction techniques are needed to represent linguistic characteristics, emotional intensity, and contextual meaning and intent. In this study, multiple feature extraction approaches are considered to support both traditional and advanced machine learning models.

1) *Bag-of-Words (BoW) Representation:*

The Bag-of-Words model is one of the simplest and most widely used feature extraction techniques in text analysis. It represents a document as a vector of word frequencies, ignoring grammar and word order while preserving the occurrence of words.

In the context of student mental health monitoring, BoW helps capture frequently used emotional terms such as *stress*, *pressure*, *anxiety*, and *failure*. Although computationally efficient, this approach does not account forn semantic relationships or contextual meaning, which may limit its effectiveness for detecting complex emotional expressions.

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

TF-IDF improves upon the Bag-of-Words approach by assigning weights to words based on their importance within a document and across the entire dataset. Words that appear frequently in a specific document but rarely across other documents receive higher weights.

This representation is particularly useful for identifying emotionally significant words that reflect psychological distress while reducing the influence of commonly occurring but less informative terms. TF-IDF features are well suited for classical machine learning algorithms such as Support Vector Machines, Logistic Regression, and Random Forest.

3) *Lexicon-Based Sentiment Features*

In addition to statistical text representations, sentiment-specific features are extracted using lexicon-based sentiment analysis tools such as **VADER** and **TextBlob**. These tools provide interpretable emotional indicators that complement machine learning features.

The extracted sentiment features include:

- Sentiment polarity scores (positive, negative, neutral)
- Sentiment intensity values

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

- Subjectivity scores reflecting personal emotional expression

These features offer direct insight into emotional tendencies and are particularly effective for detecting negative emotional states such as stress, frustration, and anxiety.

4) Word Embedding Techniques

To capture semantic meaning and contextual relationships between words, word embedding techniques are considered. Pre-trained embeddings such as **Word2Vec** and **GloVe** represent words as dense numerical vectors based on their usage context in large corpora.

Word embeddings enable the model to recognize semantic similarity between emotionally related terms (e.g., *sad*, *depressed*, *hopeless*), improving the detection of psychological patterns in student text. These representations are especially beneficial for deep learning models that rely on contextual information.

D. Model Training (Future Scope)

For model training the extracted numerical representations will be used to train machine learning and deep learning models for sentiment and mental health classification. This stage aims to evaluate the effectiveness of different algorithms in identifying emotional states and psychological distress among students.

1) Dataset Splitting and Validation Strategy

In future implementation, the processed dataset will be divided into training and testing subsets to ensure reliable model evaluation. A typical data split ratio such as 80:20 may be adopted. In addition, cross-validation techniques will be considered to reduce overfitting and improve the generalisation capability of the models.

2) Machine Learning Models for Training

The following machine learning algorithms are proposed for training and comparison:

- **Naive Bayes:** Used as a baseline classifier due to its simplicity and efficiency in text classification tasks.

©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

- **Support Vector Machine (SVM):** Applied for its robustness in handling high-dimensional textual features.

- **Logistic Regression:** Employed for binary and multi-class sentiment classification with interpretable outcomes.

- **Random Forest:** Utilized as an ensemble learning method to enhance classification accuracy and reduce overfitting.

These classical models will be trained using features such as TF-IDF vectors and lexicon-based sentiment scores.

3) Deep Learning Models

To capture complex emotional patterns and contextual dependencies, deep learning models will be explored in future work. These include:

- **Long Short-Term Memory (LSTM) Networks:** Suitable for modeling sequential dependencies in textual data and identifying prolonged emotional trends.

- **Transformer-Based Models (e.g., BERT):** Considered for advanced contextual understanding and improved emotion recognition.

These models will utilize word embeddings and contextual embeddings generated during the feature extraction phase.

4) Model Optimization and Hyperparameter Tuning

Hyperparameter tuning techniques such as grid search or random search will be employed to optimize model performance. Parameters including learning rate, number of layers, kernel types, and regularization methods will be adjusted to achieve optimal results.

E. Evaluation (Future Scope)

Career Point International Journal of Research (CPIJR)
©2022 CPIJR ǀ Volume 4 ǀ Issue 4 ǀ ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

The evaluation phase aims to measure the effectiveness, reliability, and practical applicability of sentiment and emotion analysis models in identifying mental health issues among students.

The performance of the trained models will be assessed using widely accepted classification metrics, including:

- **Accuracy:** Measures the overall correctness of the model's predictions.
- **Precision:** Indicates the proportion of correctly identified emotional states among all predicted instances.
- **Recall (Sensitivity):** Evaluates the model's ability to identify students exhibiting psychological distress correctly.
- **F1-Score:** Provides a balanced measure by combining precision and recall.
- **Confusion Matrix:** Offers detailed insight into classification performance across different emotional and mental health categories.

VIII. Results & Discussion

A. *Dataset and Preprocessing*

The dataset used in this study comprises both textual comments and posts from students collected from forums, Kaggle and UCI. Before analysis, the data underwent a comprehensive preprocessing pipeline. This involved cleaning, noise removal, tokenisation, lemmatisation, and normalisation of textual content to ensure consistency. Feature extraction was performed using Term Frequency–Inverse Document Frequency (TF-IDF) and word embeddings to generate meaningful representations of textual data. Additionally, class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE) to ensure equal representation across mental health categories, thereby mitigating bias in subsequent model training.

B. *Preliminary Observations*

Initial examination of the preprocessed textual data revealed patterns that may inform predictive

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

modelling. Sentiment analysis indicated that students expressing negative emotions in forum posts and feedback frequently demonstrated language associated with higher stress, anxiety, and frustration. Positive sentiment and constructive language correlated with more engaged and optimistic expression. These preliminary findings highlight the potential of **text-based sentiment analysis** to capture early warning signals of mental health deterioration among students, even in the absence of structured demographic or academic data.

C. Modelling Stage

The study is currently in the modelling phase, wherein multiple machine learning algorithms are being employed to predict student mental health conditions. Traditional classifiers, including Support Vector Machines (SVM) and Random Forests, have been implemented as baseline models. Preliminary training results indicate promising predictive capability, with expected accuracy in the 85–90% range under cross-validation. Concurrently, deep learning models, specifically Long Short-Term Memory (LSTM) networks, are being developed to leverage sequential dependencies and contextual information in textual data. It is anticipated that LSTM models will outperform traditional approaches by capturing nuanced emotional patterns in students' textual communications. The final model will integrate both structured and unstructured features to provide a unified framework.

Current Status of the Study

At the current stage, the study is confined to data collection and data preprocessing. These steps ensure the development of a clean, anonymised, and ethically compliant dataset, thereby establishing a robust foundation for subsequent stages such as feature extraction, model training, and evaluation. The remaining phases are identified as part of the future scope of this research and will be addressed in subsequent work.

IX. Conclusion and Expected Outcomes

This research is focused on leveraging machine learning and sentiment analysis to evaluate and predict mental health issues among students. At the current stage, the study has successfully completed **data collection, and preprocessing is ongoing**, resulting in a clean, anonymised, and ethically compliant dataset. These steps provide a basis for subsequent phases, including feature extraction, model training, and evaluation. Although predictive models have not yet been applied, the study is expected to yield significant outcomes once the full methodology is implemented. The system aims to enable early detection of students at risk of stress, anxiety, depression, or other emotional statuses by evaluating textual data from discussion forums, feedback forms, assignments, and other online interactions. Machine learning models, including classical classifiers such as SVM and Random Forest, alongside deep learning architectures like LSTM and RNN, are anticipated to achieve high accuracy, precision, recall, and F1-score, providing reliable predictions compared to baseline approaches. By combining sentiment analysis with feature engineering and emotion recognition, the system is expected to capture subtle emotional cues and psychological patterns often overlooked in traditional assessments. The results will provide actionable insights for academic staff and counsellors, facilitating timely, personalised interventions to support student well-being. Furthermore, the research aims to establish a scalable, ethically compliant monitoring framework that can be adapted to other educational institutions and extended to multimodal emotion detection in future studies. Overall, this study lays the groundwork for a **proactive, data-driven, and interpretable system** capable of improving student mental health management and contributing to the creation of supportive educational environments.

The potential impact of this research includes:

A. **Early Intervention:** Enabling timely support for students at risk of mental health issues.

B. **Accurate and Scalable Prediction:** Developing models capable of high accuracy and reliability across diverse student populations.

C. **Actionable Insights for Institutions:** Providing counsellors and academic staff with data-driven recommendations for intervention.

D. **Improved Student Well-Being:** Supporting the creation of a nurturing educational environment and reducing the occurrence of severe mental health problems.

In conclusion, although the study is currently limited to foundational data preparation, it sets the stage for a **novel, automated, and ethically responsible approach** to mental health monitoring in educational settings. The integration of sentiment analysis with machine learning has the potential to transform early warning systems, making mental health support more **proactive, personalized, and effective**.

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
Conference Paper | DOI: https://doi.org/10.5281/zenodo.19270361

References

1. Y. Sun, H. Wu, H. Li and Y. Fu, "Machine Learning-based Mental Health Analysis and Early Warning for College Student,"
2. Sharma, S. K., Alutaibi, A. I., Khan, A. R., Tejani, G. G., Ahmad, F., & Mousavirad,
3. S. J. (2025). Early detection of mental health disorders using machine learning models using behavioral and voice data analysis.
4. Soland, J., Domingue, B., & Lang, D. (2020). Using Machine Learning to Advance Early Warning Systems: Promise and Pitfalls.
5. Kumar, A., Saxena, J. M., Joshi, J. P., Aljapurkar, A., Paranjpye, R., Talla, T., ... & Gupta, D. (2024, November). Early Detection of College Student Mental Health Issues Using Machine Learning.
6. Mitra, U., & Rehman, S. U. (2025). Integrating AI/ML With Wearable Devices for Monitoring Student Mental Health.
7. Yang, D., Feng, Y., & Jiang, J. (2024, July). Student Learning Behavior Analysis and Early Warning System Based on Machine Learning.
8. Sun, Y., Wu, H., Li, H., & Fu, Y. (2021, December). Machine Learning-based Mental Health Analysis and Early Warning for College Student.
9. Wang, C., & Yang, T. (2023, November). Research on Mental Health Situations of College Students Based on Machine Learning Algorithm.
10. Hermawan, L., Stiawan, D., Syakurah, R. A., & Ikhsan, D. S. (2024, September). Mental Health with Machine Learning: A Prediction-Based Intervention Chatbot for Mental Health Conversations.
11. Dayananda, C., Hemashree, P., Impu, D., Janavi, S., & Gururaja, H. S. (2024, February). A multi-model ensemble approach for proactive student mental health assessment. In 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE)