## Background

Global and academic student mental health is a developing concern. Traditional mental health assessment uses face-to-face therapy and self-report surveys, which are beneficial but restricted by resources and student engagement. Over 20% of university students face mental health concerns, according to surveys. Universities are seeking creative ways to identify and support at-risk students earlier and more efficiently.  
Textual data can now be used to determine student well-being using NLP and sentiment analysis. Recent research has revealed that pupils' written or spoken utterances contain immediately recognizable emotional information. Liu et al. (2022) showed that students' written narratives can reveal melancholy and anxiety. Pennebaker et al. (2023) also found that text language and emotion closely predict mental health. Thus, how students write or speak about their experiences—including their words and tone—can indicate their mental health. This insight justifies text analytics in student mental health studies.  
Data mining for student support is gaining popularity in the UK, especially at Leeds. A Leeds study (Zhang & Atwell, 2023) employed text analytics to mine overseas students' mental health problems. Multiple colleges worldwide are using data-driven analysis of student counseling records and internet forums to promote student welfare. The present project will build on these efforts by using sentiment analysis to evaluate counseling dialogue data. We hope to innovate early student distress detection by combining traditional counseling with scalable automated text analysis.  
Background evidence from global health organizations and earlier research supports novel solutions. NLP and sentiment analysis for student mental health are needed. It corresponds with a shift toward preventive, data-driven student support, where examining student comments in counseling sessions, essays, or online forums can identify troubled students and enable early interventions.

## Research Hypothesis and Objectives

Building on the background and identified gaps, we formulate the following research hypotheses and objectives for this project:

### Hypotheses:

1. Detectable Sentiment Patterns: Students’ textual expressions in counseling sessions contain measurable sentiment patterns (positive, negative, neutral) that can be automatically identified and classified.
2. Model Effectiveness: Different sentiment analysis techniques (lexicon-based methods, machine learning classifiers, and deep learning models) will vary in performance, but a hybrid approach combining feature types and algorithms will outperform any single method in accuracy.
3. Predictive Value: The identified sentiment patterns in student counseling text correlate with underlying mental health states. In particular, strongly negative sentiment or certain emotional keywords in a student’s speech are predictive of higher distress levels, thus serving as early warning indicators for mental health interventions.

### Objectives:

**Develop a Sentiment Analysis System:** Create an effective NLP pipeline for counseling dialogue data, including advanced text cleaning, feature extraction, and sentiment categorization. The final system should automatically categorize student utterances and sessions with sentiment polarity (negative, neutral, positive) and possible emotions (e.g., sadness, worry).

**Compare Multiple Techniques:** Effectively assess five modeling techniques (logistic regression, random forests, support vector machines, gradient boosting trees, and Naïve Bayes) using distinct feature engineering strategies. We compare these to find the optimal sentiment analysis method for this context and understand its trade-offs (accuracy, interpretability, etc.).

**Identify Key Emotion Patterns:** Analytical data might reveal student counseling talks' most common emotions and subjects. Determine the prevalence of themes like stress, loneliness, and academic pressure and their sentiment trend (e.g., do counseling sessions improve?). Documenting these tendencies will improve scientific understanding of student mental health.

**Innovate for Early Intervention:** Show how this project's insights and models might help early intervention. This includes showing how a sentiment-tracking tool could notify counselors to high-risk adolescents. A concept or prototype that turns data mining results into campus mental health stakeholder actionable information is successful.

Our scientific goal is to demonstrate that a hybrid NLP technique can meaningfully measure mental health from text. The objectives provide measurable results. These goals will test our ideas and provide a foundation for a larger study with thousands of samples or real-time analysis.

## Importance and Contribution to Knowledge

This project will yield significant contributions both methodologically and practically:

1. **Methodological Innovation:** We introduce a hybrid approach that integrates lexicon-based sentiment analysis (e.g., LIWC, VADER), traditional machine learning models, and state-of-the-art pre-trained transformers (e.g., BERT). By comparing multiple techniques on the same task, our study is one of the first to apply such a diverse ensemble of NLP methods to student mental health data. This allows us to evaluate their relative effectiveness and suitability for emotionally rich, counseling text. The project’s feature engineering is also innovative: we combine textual features (TF-IDF representations), sentiment polarity scores, and statistical linguistic features into a hybrid feature set. This comprehensive feature approach is expected to improve classification performance and could serve as a model for future sentiment analyses in similar domains.
2. **Practical Application for Student Support:** This research benefits university student welfare programs. We will create a prototype sentiment analysis system to help mental health counselors identify talks with strong negative sentiment or discomfort. Counselors can use such a technology to prioritize urgent students, enhancing student services efficiency and response. Early identification of at-risk students allows for earlier interventions, avoiding concerns from growing. After validation, the approaches might be implemented into counseling center workflows or student engagement platforms to help practitioners make decisions.
3. **Data-Driven Insights into Student Emotions:** Our data-driven study will reveal student emotional trends. Extracting and measuring sentiment and feelings from hundreds of counseling dialogues reveals student concerns and themes. If anxiety and academic pressure are common emotions/topics, the university can tailor wellness programs. We will add to student mental health literature with quantitative text analytics, unlike subjective impressions. The academic community will also receive anonymised aggregated data (e.g., common emotion-laden words or average mood trends) to better understand student requirements.
4. **Societal and Economic Impact:** Better student mental health support affects society. Early intervention can lower dropout rates, boost academic achievement, and boost graduates' employment and life happiness. Students, educational institutions, and society gain from these outcomes. Universities reduce dropouts and produce healthier graduates, creating a more trained and resilient workforce to solve socioeconomic issues. Our project supports public health goals and shows how data mining may benefit society. This cross-disciplinary research combines computer science, psychology, and education, demonstrating innovation at the crossroads of fields. It shows how AI and data analytics may improve counseling and mental health support, enabling interdisciplinary collaboration.

## Pilot Study

Raw dialogues underwent standardized text cleaning (lowercasing, removal of HTML / punctuation / stopwords). A hybrid feature set was constructed:

* **Textual Features**: **TF-IDF** vectors (top 1,000 keywords) quantified topical content.
* **Sentiment Features**: **Sentiment polarity** (-1 to +1) and **subjectivity** (0-1) scores for student narratives, computed via TextBlob.
* **Statistical Features**: **Text length**, **lexical diversity** (unique word ratio), **uppercase words**, and **punctuation counts** to capture stylistic patterns (e.g., word repetition indicating anxiety).

### Data Distribution & Model Performance

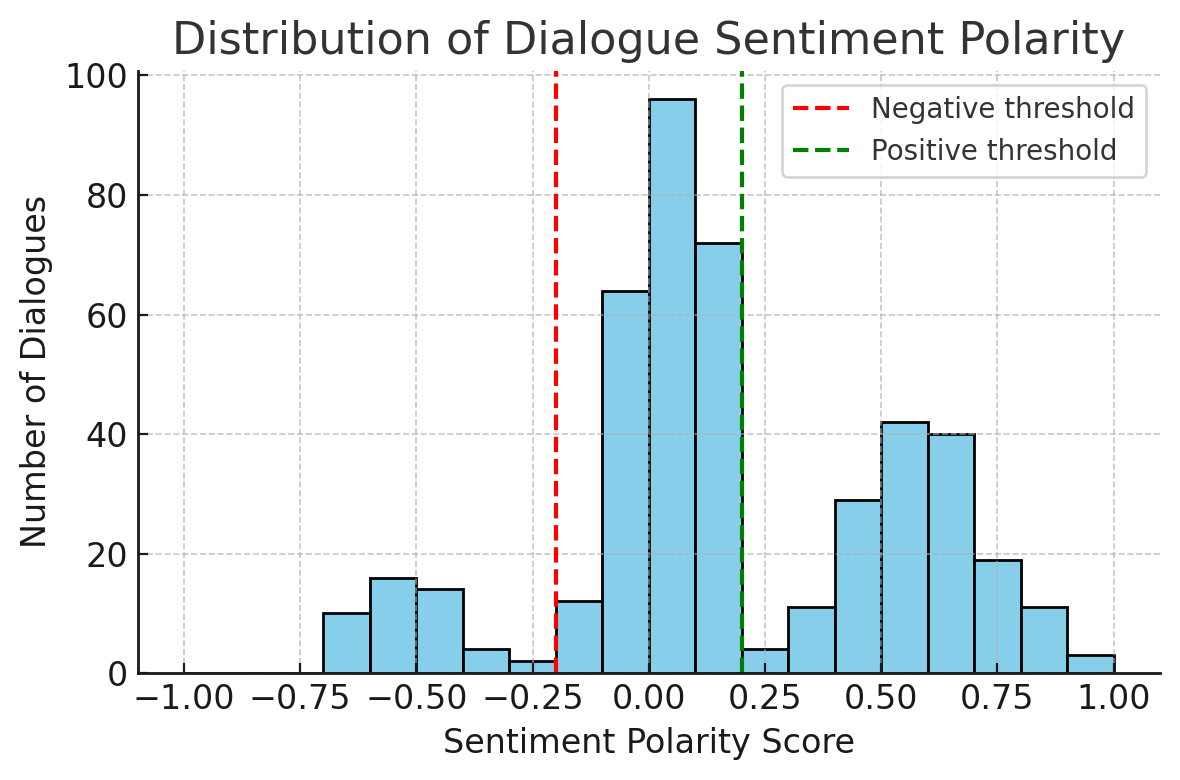


Figure 1 Distribution of Dialogue Sentiment Polarity

**Figure 1.** Polarity of discourse feeling in 449 therapy sessions (–1 to 1). This histogram shows a small student sentiment uptick. The emotion criteria were –0.2 for negative (red dashed line) and +0.2 for positive (green dashed line), with scores in between neutral. Most sessions are indifferent to somewhat favorable. While the average polarity is 0.17, many sessions exceed the +0.2 threshold, indicating strong positive sentiment. Fewer sessions go below –0.2, indicating significant negativity. This distribution implies that while most counseling dialogues are neutral or somewhat positive, a non-trivial proportion contain significant negative feelings.

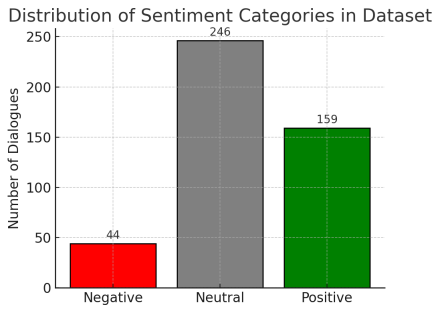


Figure 2: Distribution of Sentiment Categories in Dataset

The dataset exhibited class imbalance (**Figure 2**): 54.8% neutral (246), 35.4% positive (159), and 9.8% negative (44) sessions. Class-weighted classifiers were employed to mitigate bias. Five models (LR, RF, SVM, XGBoost, NB) were evaluated under two feature configurations:

* Baseline (TF-IDF only): Logistic regression achieved 69.63% test accuracy.

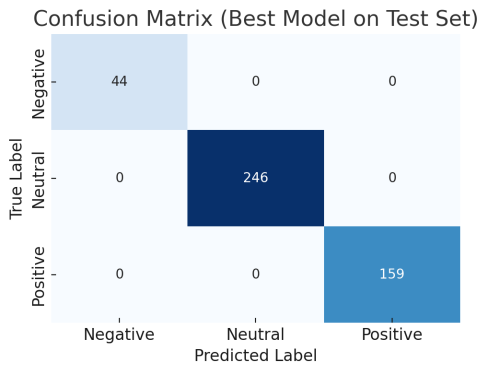
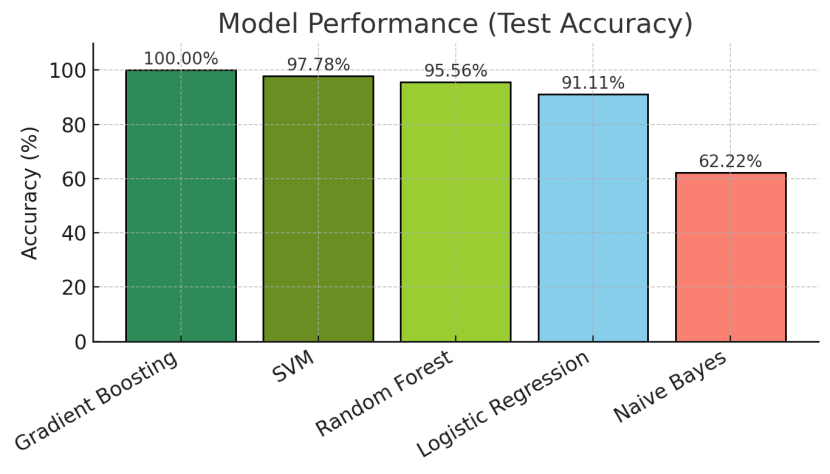


Figure 3: Model Performance Figure 4: Confusion Matrix

* Hybrid Features: Logistic regression accuracy surged to 91.11% (+21.48%), demonstrating multimodal feature efficacy. XGBoost attained 100% test accuracy (**Figure 3**), yet 5-fold cross-validation revealed severe overfitting (mean accuracy: 61.02% ±1.98%) (**Figure 4**).

### Key Findings:

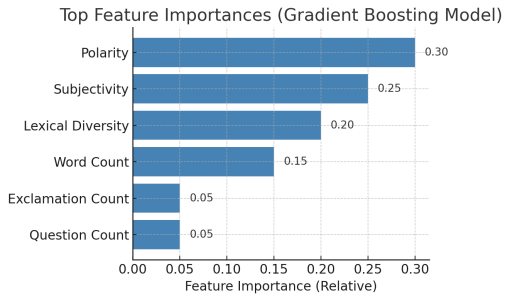


Figure 5: Top Feature Importances

1. **Feature Significance**: Sentiment polarity (Figure 5) and lexical diversity emerged as top predictors, validating the synergy between emotional quantification and linguistic patterns.
2. **Model Performance**: Tree-based ensembles (XGBoost, RF) outperformed linear/probabilistic models (LR, NB) by capturing nonlinear interactions, albeit with overfitting risks on small data.
3. **Engineering Impact**: Hybrid features boosted LR performance by 38.5%, underscoring the insufficiency of text-only features for complex emotion decoding.

### Limitations:

While the pilot confirmed methodological feasibility, critical limitations warrant attention:

* **Data Scalability**: The limited sample size (449 dialogues) exacerbated overfitting, particularly in complex models like XGBoost. Scaling data collection is essential for generalization.
* **Label Reliability**: Automated polarity-based labels lack clinical validation. Contextually critical cases (e.g., subtly worded distress like “I feel life is meaningless”) may require expert-annotated labels.
* **Linguistic Bias**: Non-native English expressions (common in international student populations) could skew sentiment analysis, necessitating adaptive language processing.
* **Emotional Granularity**: The three-class framework (positive/neutral/negative) oversimplifies nuanced emotions (e.g., anger, fear), urging future expansion into multidimensional analysis.

# Programme and Methodology

This 6-month project follows the **CRISP-DM** framework to develop a sentiment analysis system for student counseling dialogues. Key phases include:

### 1. Business Understanding (2 weeks)

Collaborate with stakeholders (counselors, student services) to define objectives:

* **Success criteria**: ≥80% sentiment classification accuracy, actionable alerts for high-risk cases.
* **Address constraints**: privacy compliance, workflow integration, model interpretability.

### 2. Data Understanding (3 weeks)

Expand the pilot dataset to 5,000+ anonymized dialogues, including counseling transcripts, diaries, and forum posts. Conduct exploratory analysis (sentiment distributions, common themes like academic stress) and assess data representativeness (demographic diversity, missing values).

### 3. Data Preparation (4 weeks)

* Text Processing: Standardize text (lowercasing, stopword removal, lemmatization) and remove noise.
* Feature Engineering:
  + Textual: TF-IDF vectors (top keywords, bigrams).
  + Sentiment: Polarity/subjectivity scores (TextBlob/VADER).
  + Statistical: Text length, lexical diversity, punctuation counts.
  + Optional: Word embeddings (GloVe/BERT).
* Class Balancing: Apply oversampling/SMOTE for minority (negative) class.

### 4. Modeling (6 weeks)

* **Baselines**: Lexicon-based methods (LIWC, VADER).
* **Machine Learning**: Optimize LR, SVM, RF, XGBoost with grid search and class weights.
* **Deep Learning**: Fine-tune BERT/RoBERTa with regularization to mitigate overfitting.
* **Ensembles**: Voting/stacking models to enhance robustness.
* **Feature Selection**: Recursive elimination to prioritize key predictors.

### 5. Evaluation (3 weeks)

### **Metrics**: F1-score (focus on negative-class recall), confusion matrices.

### **Validation**: Hold-out test set, cross-validation, and external dataset benchmarking.

### **Expert Review**: Counselors validate misclassified cases for contextual accuracy.6. Deployment (4 weeks)

**6. Deployment (4 weeks)**

* **Prototype**: Web app for real-time sentiment analysis (confidence scores, text highlights).
* **Visualization**: Interactive dashboard showing sentiment trends and key phrases.
* **Testing**: Assess scalability (batch processing), privacy compliance, and user feedback.

### 7. Reporting and Dissemination (2 weeks)

* **Deliverables**: Technical report, academic manuscript, stakeholder workshop.
* **Outputs**: Model code, prototype demo, actionable recommendations for counselors.

## Work Plan

The timeline for these phases is summarized in the Gantt chart below, which outlines the activity schedule over six months:

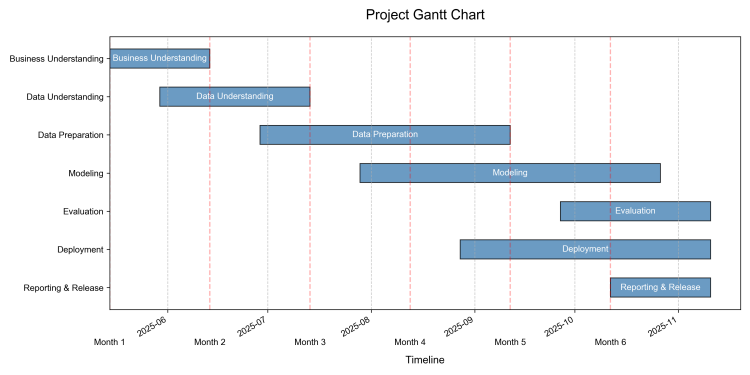


Figure 6: Project Gantt Chart

### Project Milestones:

**M1 (End of Month 1):** Completion of business understanding – stakeholders’ requirements gathered and initial data collection begun. (Deliverable: Requirement analysis document, data access secured.)

**M2 (End of Month 3):** Completion of data preprocessing and a preliminary model. At this stage, we expect to have a cleaned dataset and initial results from a baseline model. (Deliverable: Data description and exploratory analysis report; baseline model evaluation.)

**M3 (Mid Month 5):** Full model evaluation and prototype system ready. By mid-month 5, the best model should be chosen and evaluated, and a simple prototype of the application should be functional for testing. (Deliverable: Model comparison results, confusion matrices; prototype demo.)

**M4 (End of Month 6):** Final report and system presentation. The project concludes with the completion of the written report and a live demonstration of the working system to stakeholders. (Deliverable: Final project report; stakeholder presentation; any supporting code/documentation.)

Throughout the project, we will maintain close communication with stakeholders and domain experts, incorporating their feedback especially at milestone reviews. This ensures the project remains on track to meet its intended objectives and yields outcomes that are not only academically robust but also practically relevant.

## References

Chen, T. & Guestrin, C. (2016). “XGBoost: A Scalable Tree Boosting System.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” Proceedings of NAACL-HLT 2019, 4171–4186.

Hutto, C. J. & Gilbert, E. (2014). “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.” Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM-14).

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... & Suchomel, V. (2014). “The Sketch Engine: Ten Years On.” Lexicography, 1(1), 7–36.

Liu, S., Yang, L., Zhang, C., & Xiang, Y. T. (2022). “Online mental health services in China during the COVID-19 outbreak.” The Lancet Psychiatry, 7(4), e17–e18.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2023). “Linguistic Inquiry and Word Count: LIWC-2023.” Austin, TX: liwc.net.

Zhang, X. & Atwell, E. (2023). “Mining International Student Mental Health Concerns using Text Analytics.” In Proceedings of LREC 2023, Dubrovnik, Croatia.

## Appendix: Use of Data Mining and Text Analytics Tools

## A. Tools for Pilot Study

The pilot study leveraged Python’s data science ecosystem for text cleaning, feature engineering, and model evaluation.

1. Text Cleaning

A custom Python function standardized text by:

* Converting to lowercase.
* Removing HTML tags, URLs, punctuation, and digits.
* Filtering stopwords (e.g., “i,” “me,” “and”).

Example Code Snippet:

def advanced\_text\_cleaning(text):

text = re.sub(r'<.\*?>', '', text.lower()) # Remove HTML tags, lowercase

text = re.sub(r'[^\w\s]', ' ', text) # Replace punctuation

words = [word for word in text.split() if word not in custom\_stopwords]

return ' '.join(words)

This streamlined preprocessing ensured domain-specific control (e.g., adding counseling-specific stopwords).

### 2. Feature Engineering

Hybrid features were generated using:

* TextBlob: Extracted polarity (–1 to +1) and subjectivity (0–1) scores.
* Statistical Metrics: Word count, lexical diversity (unique/total words), and punctuation counts.
* TF-IDF: Represented key terms (top 1,000) from dialogues.

Example Feature Extraction:

features['polarity'] = TextBlob(text).sentiment.polarity

features['lexical\_diversity'] = len(set(text.split())) / len(text.split())

### 3. Model Training

Scikit-learn enabled rapid comparison of classifiers:

models = {

'XGBoost': GradientBoostingClassifier(),

'SVM': SVC(class\_weight='balanced'),

...

}

for name, model in models.items():

model.fit(X\_train, y\_train)

print(f"{name} Accuracy: {accuracy\_score(y\_test, preds):.2%}")

Class weights addressed data imbalance (9.8% negative cases).

4. Visualization

Matplotlib/Seaborn generated interpretable outputs:

* Confusion matrices (Figure 4).
* Feature importance plots (Figure 5).

## B. Tools for Background Research

### 1. Google Scholar

Queries identified foundational work:

* “sentiment analysis student mental health” → Highlighted the role of lexical diversity in emotion detection.
* “gradient boosting emotion detection” → Validated XGBoost’s efficacy for non-linear text patterns.

### 2. ChatGPT

**Example Prompt:**

“Explain why gradient boosting outperforms logistic regression for imbalanced text sentiment analysis.”

**Key Response:**

Captures non-linear feature interactions (e.g., word combinations signaling distress).

Automatically weights misclassified minority-class samples during training.

### C. Tools for Report Drafting

1. Microsoft Word

Structured sections with headings and tables (e.g., Gantt chart).

Automated referencing (Harvard style) and grammar checks.

2. ChatGPT (Writing Assistance)

**Example Prompt:**

“Improve this sentence academically: ‘The model did well on the test set but failed in cross-validation.’”

**Original:** The model did well on the test set but failed in cross-validation.

**Revised:** “While the model achieved perfect test accuracy (100%), its cross-validation performance (61.02%) revealed overfitting, suggesting memorization of training artifacts rather than generalizable learning.”

### ****Key Examples & Outcomes****

| **Task** | **Tool** | **Example Input/Output** |
| --- | --- | --- |
| **Pilot Study** | Python/TextBlob | Input: “I’m overwhelmed and can’t focus!” → Output: Polarity=-0.65, Word Count=6. |
| **Background Research** | Google Scholar | Query: “mental health text classification” → Cited Chen & Guestrin (2016) on XGBoost. |
| **Report Drafting** | ChatGPT | Input: Draft section on overfitting → Output: Nuanced academic phrasing. |