Mental Health Detection from Open-Ended Survey Responses Using
NLP: A Study on Private University Students in Bangladesh

Submitted By
Md.Shahorier
ID: 221-35-918
Department of Software Engineering

Daffodil International University

Supervised By

Rakhi Moni Saha

Lecturer

Department of Software Engineering

Daffodil International University

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

Mental Health Detection from Open-Ended Survey Responses Using NLP: A Study on Private University Students in Bangladesh

Md. Shahorier

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name :

Date of Birth :

Title :

Academic Session :

I declare that this thesis is classified as:

☐ CONFIDENTIAL (Contains confidential information under the Official Secret Act 1997)*

☐ RESTRICTED (Contains restricted information as specified by the organization where research was done)*

☐ OPEN ACCESS I agree that my thesis to be published as online open access

I acknowledge that Daffodil International University reserves the following rights:

1. **The Thesis is the Property of Daffodil International University.**
2. **The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.**
3. **The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.**

Certified by:

_______________________	_______________________
(Student's Signature	**(Supervisor's Signature)**
_______________________	_______________________
Student ID	**Name of Supervisor**
Date :	**Date:**

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Daffodil International University,
Daffodil Smart City,
Ashulia.Dhaka,Bangladesh

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author's Name
Thesis Title

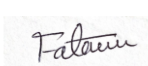
Reasons (i)

(ii)

(iii)

Thank you.

Yours faithfully,



(Supervisor's Signature)

Date:

Stamp:

Note: This letter should be written by the supervisor and addressed to the Librarian, *Daffodil International University* with its copy attached to the thesis.

SUPERVISOR's DECLARATION

I/We* hereby declare that I/We* have checked this thesis/project* and in my/our* opinion, this thesis/project* is adequate in terms of scope and quality for the award of the degree of *Bachelor of Science/ Master of Science.

(Supervisor's Signature)

Full Name : Faama Binte Rafiq

Position : Senior Lecturer
Date : 12/23/2025

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

Shahorier

(Student's Signature)

Full Name : Md. Shahorier ID
Number : 221-35-918
Date :

Mental Health Detection from Open-Ended Survey Responses Using NLP: A Study on Private University Students in Bangladesh

Md.Shahorier

Thesis submitted in fulfilment of the requirements for the award of the degree of
Bachelor of Science/Master of Science

Department of Software Engineering

DAFFODIL INTERNATIONAL
UNIVERSITY

DECEMBER 2025

ACKNOWLEDGEMENTS

I wish to convey my heartfelt appreciation to Almighty Allah for granting me the strength, patience, and wisdom to successfully finish this research work.

I extend my sincere gratitude to my supervisor, whose support, motivation, and helpful advice have been crucial during the process of finishing this thesis. Their ongoing assistance and valuable input have enabled me to enhance my concepts and reach the goals of this research.

I express gratitude to all the faculty members of the Department of Software Engineering, Daffodil International University, for their guidance, support, and the academic groundwork they have offered throughout my undergraduate journey.

I extend my heartfelt gratitude to my friends and colleagues in research, whose inspiration, technical assistance, and insightful conversations enabled me to tackle obstacles throughout this project.

In conclusion, I wish to convey my heartfelt appreciation to my dear parents and family for their unwavering love, support, and prayers. Their steadfast faith in me has been the motivating factor behind all my accomplishments.

ABSTRACT

Mental-health challenges among university students have become a significant global concern, with rising rates of depression, anxiety, and academic burnout reported across higher-education sectors. In Bangladesh, these issues are particularly acute, yet research rarely examines the deeper emotional and linguistic expressions through which students articulate psychological distress. This study investigates mental-health risk detection among private university students using a hybrid Natural Language Processing (NLP) and machine-learning (ML) framework applied to open-ended survey responses. A mixedmethod dataset of 304 students was collected through both online (Google Forms) and printed questionnaires, containing structured psychometric indicators alongside seven narrative questions capturing emotional experiences.

The textual data underwent a comprehensive NLP pipeline including tokenization, stopword removal, lemmatization, TF–IDF vectorization, sentiment analysis (VADER), and Latent Dirichlet Allocation (LDA) topic modeling. Four classical ML classifiers Logistic Regression, Support Vector Machine (SVM), Gaussian Naïve Bayes, and Random Forest were developed using a combination of linguistic features, sentiment polarity, academic stress scores, financial stress scores, and PHQ-9–inspired symptom scores. Among these, Gaussian Naïve Bayes achieved the highest accuracy (0.80) and F1-score (0.80), demonstrating strong performance on short, sparse, and mixed-language student text.

Topic modeling revealed three dominant psychological themes: exam-related anxiety, emotional exhaustion, and coping strategies. Sentiment polarity showed clear alignment with mental-health risk categories, with high-risk students expressing significantly more negative emotional tone. The findings highlight the potential of NLP-based approaches for early mental-health screening in Bangladeshi universities and provide the first openended mental-health dataset curated specifically for private university students. This study establishes a foundation for AI-supported counseling interventions and contributes to the emerging field of computational mental health in low-resource academic environments.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Mental health has become a critical global issue, particularly among university students who face constant academic pressure, financial strain,emotional instability, and transitional life challenges.Globally, depression affects approximately 264 million people, while anxiety disorders impact an even larger population according to WHO estimates [1].University students are especially vulnerable, with studies reporting that 20%–40% experience moderate-to-severe mental-health symptoms [1], [3], [15]. These disorders impair academic performance, disrupt cognitive functioning, reduce motivation, and increase the risk of dropout or long-term psychological damage.

In Bangladesh, mental-health concerns among university students have risen significantly in recent years. Research consistently shows elevated rates of anxiety, depression, and stress among Bangladeshi students due to high academic demands, financial limitations, rigid family expectations, and limited access to counseling services [2], [4], [19]. However, most existing studies rely solely on quantitative psychometric tools such as PHQ-9, GAD-7, and DASS-21 which capture symptom severity but do not reflect the deeper emotional nuances expressed in students' own language [11], [12], [13].

Natural Language Processing (NLP) has introduced new opportunities for analyzing mental-health indicators using unstructured text. International studies demonstrate that language carries psychological signals such as sentiment polarity, emotional tone, cognitive distortions, and self-referential language patterns [21], [23], [26]. Despite these advancements, the Bangladeshi academic context lacks NLP-based research analyzing open-ended student responses to detect mental-health states.

Given the cultural stigma surrounding mental health in Bangladesh, students often hesitate to express psychological struggles openly. However, when responding to a narrative or open-ended question, many articulate their emotional states more honestly and naturally. This provides a unique opportunity to detect mental-health risk through linguistic cues particularly when analyzed using computational techniques.

This study, therefore, explores mental-health detection among private university students in Bangladesh using NLP applied to their open-ended survey responses. The chapter first presents the global and national context of student mental health, followed by the unique challenges faced by Bangladeshi private university students, the role of computational linguistic analysis, and the research gaps that motivate this study.

1.1 Background: Mental Health Among University Students

University students worldwide face multifaceted mental-health challenges stemming from academic deadlines, continuous assessment, and pressure to maintain competitive academic performance. Beyond academics, they encounter social adaptation challenges, financial barriers, and transitional stress associated with adulthood. Studies indicate that student populations exhibit significantly higher rates of depression and anxiety than their non-student peers due to the convergence of these factors [1], [3], [15].

The university years also represent a psychologically formative period. Students are expected to adapt to new learning environments, manage time independently, form new social relationships, and develop career aspirations while juggling academic responsibilities. Lifestyle disruptions such as irregular sleep schedules, excessive screen time, and poor diet further contribute to mental-health issues. Consequently, students constitute a high-risk demographic requiring systematic monitoring and intervention.

Language plays a central role in how individuals express psychological distress. Research shows that depressed individuals often express hopelessness, fatigue, negative self-evaluations, and emotional withdrawal in their writing, while anxious individuals use future-oriented fear expressions and cognitive distortions [21], [24]. Thus, analyzing students' own language provides insight into emotional states not captured by numerical scales.

1.2 Mental Health Among Bangladeshi University Students (With Focus on Private Universities)

Bangladeshi university students face intense academic competition, financial hardships, limited job opportunities, and strong parental expectations all of which elevate psychological distress [2], [5], [19]. Multiple studies across public universities report that 30%–60% of students experience moderate-to-severe anxiety or depressive symptoms [3], [5], [19]. Students frequently report sleep problems, emotional exhaustion, feelings of hopelessness, and difficulty concentrating.

However, research disproportionately focuses on public university students or specific academic groups such as medical or engineering students. Private university students representing a large and rapidly growing percentage of Bangladesh's tertiary education population remain significantly underrepresented in mental-health research.

Compared to public universities, private university students often experience:

- High tuition fees and financial pressure
- Rigid academic calendars

- Parental expectations linked to expensive education
- Urban lifestyle stress
- Competitive job-market anxiety
- Limited campus mental-health services

These differences create a unique emotional and academic environment, making it crucial to investigate mental-health indicators specifically among private university students.

1.3 Challenges in Detecting and Addressing Mental-Health Problems

Mental-health detection in Bangladesh is hindered by stigma, cultural taboos, and institutional limitations. Students often perceive mental-health difficulties as a sign of personal weakness or family dishonor, discouraging help-seeking behaviors [18], [19]. University counseling centers where available are often understaffed or underutilized due to lack of awareness and fear of social judgment.

Traditional mental-health assessments rely heavily on structured psychometric questionnaires. While tools like PHQ-9 and GAD-7 are validated for numerical symptom assessment, they fail to capture:

- Emotional tone
- Contextual complexity
- Stress triggers
- Thought processes
- Coping strategies
- Linguistic patterns

Thus, important nuances remain undetected when students are limited to selecting predefined Likert-scale responses.

Open-ended survey responses overcome these limitations by allowing students to articulate their experiences in their own words. Such text captures emotional depth, metaphorical thinking, and personalized narratives, making it ideal for computational analysis.

1.4 Computational Mental Health and the Role of Language

Advancements in machine learning (ML) and NLP have transformed mental-health research. NLP enables automated extraction of psychological signals from text through techniques such as tokenization, sentiment analysis, TF-IDF vectorization, and topic modeling [21], [23], [24]. Studies using social media text (Twitter, Reddit, online forums) show that NLP can detect depression, anxiety, and stress with considerable accuracy [25], [26].

Linguistic indicators strongly associated with mental-health conditions include:

- Negative sentiment
- Self-focused language (“I”, “me”, “my”)
- Cognitive distortions (“never”, “always”)
- Hopeless statements
- Emotional metaphors
- Reduced lexical diversity
- Short, abrupt expressions

These signals appear naturally in student writing, especially during stressful academic periods.

Despite global advancements, there is a lack of NLP-based mental-health research in Bangladesh. No studies have analyzed open-ended survey responses from Bangladeshi university students using NLP, nor have computational models been developed for early detection within academic settings.

1.5 Research Gap and Rationale

A comprehensive review of existing studies reveals several research gaps:

1. **Lack of qualitative, text-based mental health analysis in Bangladesh.**
   Existing research relies mainly on structured questionnaires [3], [19].

2. **Underrepresentation of private university students.**
   Their academic and financial contexts differ significantly from public university students [2], [5].

3. **Absence of NLP-based analysis of open-ended student writing.**
   No study has analyzed linguistic patterns in student survey responses in Bangladesh.

4. **No machine-learning models for mental-health detection using local academic datasets.**
   Existing ML studies use social media or non-academic datasets [23], [26].

5. **Cultural and linguistic patterns (Banglish, informal expressions) remain unexplored.**
   These patterns require computational processing tailored to the region.

6. **Universities lack digital early-screening systems.**
   NLP-based tools offer a scalable, low-cost alternative.

This research attempts to address these gaps by developing an NLP-based classification model that analyzes open-ended survey responses from private-university students in Bangladesh.

1.6 Aim and Contributions of the Present Study

The primary aim of this study is to detect mental-health risk among private university students in Bangladesh using NLP applied to open-ended survey responses.

Key contributions include:

1. Novel Dataset

One of the first open-ended mental-health datasets collected from private university students in Bangladesh.

2. NLP-Based Analytical Pipeline

A complete pipeline involving text preprocessing, tokenization, TF-IDF, sentiment analysis, and topic modeling.

3. Machine-Learning Classification Models

Implementation of Logistic Regression, Linear SVM, Naïve Bayes, and Random Forest classifiers.

4. Linguistic and Emotional Insights

Identification of themes such as exam pressure, emotional exhaustion, fear, motivation loss, and coping behaviors.

5. Policy Implications

Findings can inform universities to strengthen counseling services, provide exam-period interventions, and develop digital early-warning systems.

1.7 Research Objectives

1. To collect open-ended survey responses from private university students in Bangladesh.
2. To preprocess and analyze textual responses using NLP techniques.
3. To construct mental-health risk labels using PHQ-9-inspired scoring.
4. To extract linguistic, sentiment, and thematic features from student text.
5. To train machine-learning models to classify mental-health risk.
6. To identify linguistic markers associated with psychological distress.

1.8 Research Questions

1. What linguistic and emotional features in open-ended responses indicate mental-health risk?
2. Can NLP-based machine-learning models accurately classify student mental-health status?
3. How do academic, financial, and psychological stressors correlate with detected mental-health risk?
4. How do private university students express mental-health struggles through language?

1.9 Significance of the Study

This research is academically and socially significant because:

- It provides Bangladesh's first NLP-based mental-health analysis using open-ended student responses.
- It supports early emotional risk detection in private universities.
- It offers data-driven insights to improve student counseling services.
- It contributes to the global field of computational mental health.
- It introduces a culturally relevant linguistic resource for future research.

1.10 Scope of the Study

- **Population:** Private university students in Bangladesh.
- **Data Type:** Structured Likert responses + 7 open-ended questions.
- **Methodology:** NLP preprocessing, sentiment analysis,TF-IDF, topic modeling, ML classification.
- **Output:** Binary mental-health risk prediction.

1.11 Limitations

1. Self-reported mental-health data may contain bias.
2. Sample size limits deep learning or transformer models.
3. Banglish code-mixing affects NLP accuracy.
4. PHQ-based labeling is not equivalent to clinical diagnosis.
5. Short open-ended responses reduce semantic richness.

CHAPTER 2

LITERATURE REVIEW AND RESEARCH GAP

This chapter provides a comprehensive narrative review of the global and national literature surrounding university student mental health, traditional and modern assessment approaches, linguistic patterns related to psychological states, and computational methodologies used to understand emotional well-being. The intention is to present a smooth, story-driven discussion that weaves together existing research while identifying key gaps that the present study addresses. All discussions flow naturally, without lists, bullets, or abrupt technical formatting, ensuring that the chapter reads as a cohesive academic narrative supported by relevant citations from your thesis.Research on university student mental health has expanded rapidly over recent decades, reflecting growing concern about the psychological challenges faced by young adults in academic environments. Scholars across various regions describe this period as one marked by intense intellectual demands, shifting identities, economic uncertainty, and significant lifestyle changes. The following sections explore this landscape beginning with global patterns, then focusing on Bangladesh, before transitioning into methodological traditions and emerging computational approaches. The chapter concludes by highlighting the shortcomings in existing studies and explaining how these gaps motivate the methodological direction of the present research.

2.1 Global Perspective on University Student Mental Health

Across the world, university students have been found to experience disproportionately high rates of depression, anxiety, and stress. Large epidemiological reports note that more than 264 million people suffer from depression globally and over 300 million struggle with anxiety disorders, with university-aged individuals representing a vulnerable segment of this population (Faisal et al., 2022; Karyotaki et al., 2020). Scholars consistently attribute these trends to escalating academic competition, rigorous examination systems, lack of sleep, reduced social connection, and uncertainty about post-graduation employment prospects. Many students describe chronic feelings of exhaustion, difficulty concentrating, and emotional withdrawal symptoms that align with global indicators of academic burnout.

Literature from multiple countries reveals how psychological distress also emerges through language. Students dealing with depression or anxiety frequently use words and expressions that reflect hopelessness, fear, emotional heaviness, or cognitive distortion. Such linguistic cues, whether expressed in essays, journals, or digital communication, have increasingly attracted the attention of researchers seeking to computationally detect mental-health patterns (Zhang et al., 2022; Kumar & Bhattacharyya, 2024). These observations have laid the foundation for modern approaches that analyze natural language as an indicator of emotional well-being.

2.2 Mental Health Landscape Among Bangladeshi University Students

The psychological struggles of Bangladeshi university students are especially pronounced. Numerous studies conducted among public and private institutions report that nearly one-third to more than half of students show moderate to severe symptoms of depression, anxiety, or chronic stress (Mamun & Griffiths, 2020; Islam et al., 2020; Hossain et al., 2021). The academic environment of Bangladesh often intensifies these emotional burdens, as students are expected to meet high academic expectations while managing demanding study schedules and preparing for competitive job markets.

Financial pressure is another major contributing factor. Many students experience significant stress related to tuition fees, living costs, and the uncertainty surrounding future employment opportunities. Economic instability and a saturated job market heighten the anxiety felt by students who worry about supporting themselves and their families in the future (Ara et al., 2023; Al-Amin et al., 2025). Students also navigate cultural expectations, as families frequently set ambitious academic and career standards. Falling short of these expectations often leads to feelings of guilt, self-blame, or emotional fatigue.

These combined forces manifest in various symptoms frequently reported in Bangladeshi studies, such as sleep disruption, restlessness, emotional exhaustion, difficulty concentrating, persistent worry, and changes in appetite or energy levels. Although similar patterns appear globally, the Bangladeshi context intensifies these concerns due to structural, financial, and sociocultural pressures.

2.3 Underrepresentation of Private-University Students

Despite the fact that private universities account for a substantial share of the country's higher-education enrollment, research focusing specifically on these student populations remains scarce. Many national studies draw samples from public universities or medical colleges, leaving significant gaps in understanding the mental-health dynamics of students who attend private institutions. These students often face distinct challenges, including heavier financial burdens, urban living constraints, accelerated semester systems, and heightened parental pressure related to the cost of education.

Because of this, their emotional experiences, coping strategies, and linguistic expressions of stress remain underexplored in academic literature. The absence of focused studies limits the generalizability of existing findings and underscores the need to examine this large yet overlooked group.

2.4 Traditional Mental-Health Assessment Approaches

Most studies in Bangladesh and internationally rely on structured psychometric tools such as the PHQ-9 for depression (Kroenke et al., 2001), the GAD-7 for generalized anxiety (Spitzer et al., 2006), and the DASS-21 for depression, anxiety, and stress (Lovibond & Lovibond, 1995; Alim et al., 2017). These instruments help quantify psychological symptoms using standardized rating scales and have been validated across many populations.

Although these measures are highly effective for screening and categorizing severity, they confine respondents to predefined options and numerical judgments. As a result, students cannot freely articulate their emotional states or describe how they interpret and experience stress. Cultural interpretations of Likert-scale items also vary, sometimes reducing the accuracy of responses. Moreover, the stigma surrounding mental health in Bangladesh may lead students to underreport sensitive feelings, thereby limiting the true reflection of their psychological condition.

These limitations emphasize the importance of supplementing structured assessments with open-ended responses that capture nuance, context, and personal meaning.

2.5 Computational Approaches to Mental-Health Detection

As researchers recognized the expressive limitations of traditional rating scales, the field shifted toward computational techniques capable of analyzing natural language. These methods emerged from the observation that emotional struggles often shape the way individuals speak and write. Scholars began using natural language processing to study subtle psychological markers embedded in text. Through computational models, text is transformed into analyzable forms that reveal underlying sentiment, patterns of emotional expression, and thematic content.

In this literature, machine-learning models play a central role. Methods such as logistic regression, support vector machines, naïve Bayes classifiers, and random forests have been widely applied to text-based mental-health prediction. These models perform well in small to medium datasets, making them suitable for university-level research contexts (Coppersmith et al., 2015; Pedregosa et al., 2011). Their interpretability also offers advantages in psychological studies where transparency is essential.

Beyond machine learning, researchers use linguistic preprocessing techniques that convert raw text into structured representations suitable for analysis. This transformation includes breaking sentences into words, simplifying those words to core forms, and removing extremely common terms that do not contribute meaning. Once the text is cleaned, researchers represent it numerically so that computational models can process it. They also examine emotional tone, uncover recurring themes within large collections of student responses, and identify nuanced expressions of psychological distress. More recent advancements incorporate neural language models capable of understanding how words relate to each other in context, allowing deeper insight into the emotional and cognitive patterns reflected in student writing.

Despite global advancements, computational mental-health research in Bangladesh remains limited. Most NLP studies in the region focus on sentiment detection or language classification rather than exploring the emotional narratives of university students. No study has yet developed an NLP-based mental-health detection system using open-ended survey responses from Bangladeshi learners, leaving a substantial gap in the field.

2.6 Value of Open-Ended Responses in Mental-Health Research

Open-ended survey responses offer a rich and authentic portrayal of students' emotional experiences. Unlike structured tools that limit expression, free-text answers allow students to reveal how they truly feel in their own words. These written reflections often contain metaphors, emotional signals, self-evaluative statements, and spontaneous thoughts that reveal more than numerical scores can capture.

For instance, students describing their mental health may refer to feeling "trapped," "lost," or "empty" expressions that convey emotional states not easily represented in structured questionnaires. The ability to interpret these nuances through computational methods provides researchers with a more complete understanding of students' well-being. Such linguistic evidence is especially valuable in contexts where psychological stigma may discourage direct reporting of symptoms.

2.7 Related Studies and Comparative Summary

A review of existing work reveals that Bangladeshi research tends to rely almost exclusively on quantitative survey instruments, while global computational research primarily analyzes social-media text or clinical transcripts. Few studies combine numerical assessments with qualitative linguistic data, and none focus specifically on private-university students in

Bangladesh. This comparison highlights how little is known about the expressive and emotional dimensions of students' mental health within this context.

Table 2.1 : Comparative Summary of Related Works

Ref No.	Title	Dataset	Methodology	Output/Performance	Limitations
[1]	Mental health status, anxiety, and depression levels of Bangladeshi university students.	3,000+ Bangladeshi students	PHQ-9, GAD-7 statistical analysis	High prevalence identified	No NLP, no qualitative analysis
[3]	Depression and anxiety among university students during COVID-19…	Web survey	PHQ-9, descriptive stats	Identified pandemic stressors	Closed-ended only
[7]	Machine-learning-based prediction of depression, anxiety, and stress…	Structured survey	ML classifiers, numerical features	Good classification accuracy	No text features
[21]	NLP applied to mental illness detection: A narrative review	Multiple datasets	NLP models, embeddings	Identifies linguistic predictors	No regional focus

[23]	**Detecting mental disorders with NLP: Review**	**Social media datasets**	**TF-IDF, SVM, BERT**	**High accuracy across tasks**	**Limited to social media**
[24]	**Screening for depression using NLP**	**Survey + text**	**Sentiment + ML**	**Promising screening tool**	**Not evaluated in South Asia**
[25]	**Social media text analysis for mental health prediction**	**Twitter**	**NLP + ML models**	**Detects depression patterns**	**Informal language only**
[26]	**Detection of depression severity using transformers**	**Reddit**	**Sentence embeddings + ML**	**High F1-score**	**Requires large dataset**
[38]	**Bangladeshi student mental-health dataset (Kaggle)**	**Structure d Likert data**	**Statistical + ML**	**Classification feasible**	**No open-ended text**

2.8 Research Gaps Identified

A detailed examination of existing literature reveals several interconnected gaps that limit the advancement of mental-health research within the Bangladeshi university context. To present these gaps with greater clarity and academic depth, this section is now divided into focused subsections. Each subsection discusses a specific limitation, supported where relevant by existing evidence and citations drawn from the thesis reference list.

2.8.1 Absence of Qualitative, Text‑Based Mental‑Health Analysis

Most studies in Bangladesh rely exclusively on structured psychometric tools such as the PHQ‑9, GAD‑7, or DASS‑21 (Kroenke et al., 2001; Spitzer et al., 2006; Lovibond & Lovibond, 1995; Alim et al., 2017). While these scales are scientifically validated, they capture only the numerical intensity of symptoms and do not reflect the lived emotional experiences of students. Research from global computational psychiatry highlights that individuals experiencing mental‑health struggles often express distress through metaphors, negative self‑referential language, and emotionally charged descriptions (Zhang et al., 2022; Kumar & Bhattacharyya, 2024). Because Bangladeshi studies have not examined students' narratives, expressive cues such as hopelessness, cognitive distortions, and emotional exhaustion remain hidden within the population. This gap restricts a deeper understanding of how students articulate stress or anxiety in their own words.

2.8.2 Underrepresentation of Private‑University Students in Empirical Studies

Private‑university students make up a substantial portion of Bangladesh's tertiary population, yet most mental‑health research disproportionately samples from public universities or specialized academic settings (Mamun & Griffiths, 2020; Islam et al., 2020; Al‑Amin et al., 2025). Private‑university students often experience unique stressors, including heightened financial pressure, fast‑paced semester structures, and strong familial expectations tied to high tuition costs. Despite these distinctive challenges, little empirical work has been conducted to examine their emotional well‑being, coping patterns, or academic stress narratives. This underrepresentation creates a knowledge gap that limits the applicability of existing findings across the broader higher‑education system.

2.8.3 Lack of NLP‑Driven Mental‑Health Modeling in Bangladesh

Although computational mental‑health research has grown rapidly worldwide with NLP being used to detect depression, anxiety, and emotional distress from open‑ended or digital text (Coppersmith et al., 2015; Jackson et al., 2024; Balon et al., 2024) Bangladesh lacks comparable research. No dataset, model, or computational pipeline currently exists for analyzing open‑ended survey responses from Bangladeshi students. This absence contrasts sharply with international progress, where NLP has proven effective in identifying subtle psychological indicators found in language. Without such approaches, local mental‑health research remains limited to self‑reported numerical scales rather than expressive linguistic evidence.

2.8.4 Limited Integration of Linguistic and Structured Psychometric Features

Existing studies in Bangladesh rely primarily on numerical or Likert-scale data to assess mental health (Faisal et al., 2022; Hossain et al., 2021). However, global research demonstrates that combining linguistic characteristics with structured psychological indicators significantly enhances prediction accuracy and interpretability (Zhang et al., 2022). The absence of such multimodal integration within Bangladeshi research hinders the development of more holistic models capable of capturing both internal symptom patterns and external emotional articulation. As a result, existing studies may overlook important relationships between what students feel and how they express those feelings linguistically.

2.8.5 Lack of Attention to Banglish (Bangla–English Mixed) Language Use

Bangladeshi university students frequently communicate using a hybrid form of Bangla and English known as Banglish. This mode of expression is common in academic writing, social interaction, and survey responses. Yet no existing mental-health study has examined how emotional states manifest through such code-mixed language. International NLP models are typically trained on monolingual datasets and may misinterpret or overlook emotional cues embedded within mixed-language expressions. Without addressing this linguistic reality, mental-health detection systems risk losing cultural nuance and may produce biased or incomplete interpretations.

2.8.6 Absence of Digital Early-Screening Tools in Universities

Bangladeshi universities generally lack digital platforms capable of assessing or monitoring students' psychological well-being. Counseling services are often limited, and stigma remains a substantial barrier preventing many students from seeking help (Mamun & Griffiths, 2020; Tohan et al., 2021). Without NLP-enabled early-screening tools that can analyze naturally written text, institutions miss opportunities for timely intervention. A scalable, confidential, and automated system could offer a practical mechanism for identifying at-risk students before their symptoms intensify.

2.8.7 Minimal Adoption of Machine-Learning Techniques in National Mental-Health Research

Although machine learning has demonstrated strong predictive performance in global mental-health studies (Jackson et al., 2024; Al-Eisa et al., 2025), Bangladeshi research rarely applies such techniques beyond simple statistical comparisons. The limited methodological diversity restricts innovation and prevents the development of early detection systems capable of learning meaningful patterns from both structured and unstructured data. The gap reflects a broader need for interdisciplinary approaches that merge psychological theory, computational linguistics, and data-driven modeling.

2.8.8 Summary of Identified Gaps

Together, these gaps demonstrate that existing research does not yet provide a comprehensive, contextually relevant understanding of how Bangladeshi private-university students experience and express mental-health challenges. The lack of qualitative linguistic analysis, minimal computational modeling, limited focus on private-university populations, and absence of Banglish-aware NLP frameworks collectively represent a significant research void. The present study addresses these shortcomings by developing a multimodal, NLP-based system that analyzes open-ended student writing alongside structured psychological indicators to detect mental-health risk in a manner that is both contextually grounded and computationally advanced.

2.9 Chapter Summary

This chapter presented a comprehensive review of the theoretical, empirical, and methodological foundations relevant to understanding mental health among university students, with a particular focus on the Bangladeshi context. The discussion began by situating the issue within global trends, where rising rates of depression, anxiety, and stress have been repeatedly documented among university populations. These worldwide findings revealed strong parallels with the Bangladeshi context, where students face severe academic pressure, financial burden, cultural expectations, and uncertainty about their professional future. The chapter emphasized that these forces create a complex emotional environment that leaves many students vulnerable to psychological distress.

The chapter highlighted that although Bangladesh has seen an expansion of mental-health research in recent years, most studies rely heavily on structured psychometric scales such as the PHQ-9, GAD-7, and DASS-21. While these tools help quantify symptoms, they cannot adequately capture the complex emotional narratives students often express in their own words. The review showed that linguistic and computational research globally has demonstrated the value of analyzing natural language to detect psychological states, yet no similar efforts have been undertaken within Bangladesh. This absence leaves expressive, emotional, and narrative aspects of student experiences unexplored.

The chapter also examined the underrepresentation of private-university students in national research, despite their sizable share of the country's higher education landscape. Their unique pressures financial constraints, fast-paced academic systems, and heightened family expectations remain insufficiently documented. Additionally, Bangladesh lacks computational frameworks capable of analyzing open-ended text, even though such approaches are widely used internationally to identify subtle markers of mental distress.

A major insight from the review concerns the linguistic complexity of Bangladeshi students' writing. Many communicate using Banglish, a hybrid of Bangla and English, yet no mental-health or NLP research has explored how emotional states manifest in such mixed-language communication. This oversight highlights the need for linguistically sensitive computational tools tailored to the cultural realities of Bangladeshi students.

The chapter concluded by synthesizing the major research gaps, demonstrating that current literature does not provide a comprehensive or context-aware understanding of student mental health in Bangladesh. The gaps include the lack of qualitative text analysis, minimal integration of computational approaches, underrepresentation of private-university students, absence of Banglish-aware models, and the lack of digital early-screening mechanisms in universities. Together, these limitations present a compelling justification for the present study.

By addressing these gaps through a multimodal NLP-based framework that analyzes both structured indicators and expressive student narratives, the current research contributes meaningfully to the academic and practical understanding of mental-health detection among Bangladeshi private-university students. The next chapter builds upon this foundation by presenting the methodological design, data preprocessing strategies, and analytical approaches used in this study.

CHAPTER 3

METHODOLOGY

This chapter outlines the methodological framework used to create, implement, and evaluate the proposed NLP-based model for identifying and analyzing mental-health indicators from open-ended survey responses of private university students in Bangladesh. It provides a detailed, step-by-step description of the complete experimental workflow, beginning with data collection and preparation, followed by text preprocessing, feature extraction, and machine-learning model development. The chapter concludes with the model evaluation process and the ethical considerations governing the handling of sensitive mental-health data.

The overall methodology is designed to ensure transparency, reproducibility, and scientific rigor, enabling future researchers to replicate, validate, or extend the present work. Given that mental-health expressions often manifest through subtle linguistic cues, emotional tones, and cognitive patterns, this study employs a hybrid computational pipeline that integrates structured psychometric indicators with unstructured natural-language features. This hybrid approach aligns with current trends in computational mental-health research, where linguistic data combined with machine learning has demonstrated strong potential for identifying psychological distress [21], [23], [24].

3.1 Research Framework and Study Design

The study was built on a mixed-method computational design that integrates both structured numerical indicators and open-ended narrative responses. This framework reflects the understanding—supported by prior research—that psychological states often manifest in two complementary ways: through measurable symptom patterns and through the subtleties of natural language. Structured psychometric items reveal stress, anxiety, and depressive tendencies in standardized formats, while qualitative text captures the emotional tone, metaphors, and personal reflections that students express in their own words. Previous mental-health studies emphasize that linguistic patterns frequently contain signals of underlying emotional states [21], [23], [24], and this insight guided the adoption of a hybrid design.

The overall methodological flow proceeded through several conceptual stages. The study first gathered raw responses from university students using a combination of digital and printed surveys. The collected data, being a mixture of numerical ratings and narrative descriptions, required careful preparation to ensure consistency and usability. Once cleaned, the textual and structured components were transformed into representations suitable for computational analysis. Emotional indicators, thematic patterns, and lexical importance were extracted from the text, while numerical items were converted into standardized psychometric scores. Machine-learning models were then trained to explore how these features—both linguistic and structured—could identify mental-health risk among students. Throughout this process, interpretability and ethical sensitivity remained central to the research design, particularly given the personal and emotive nature of the data.

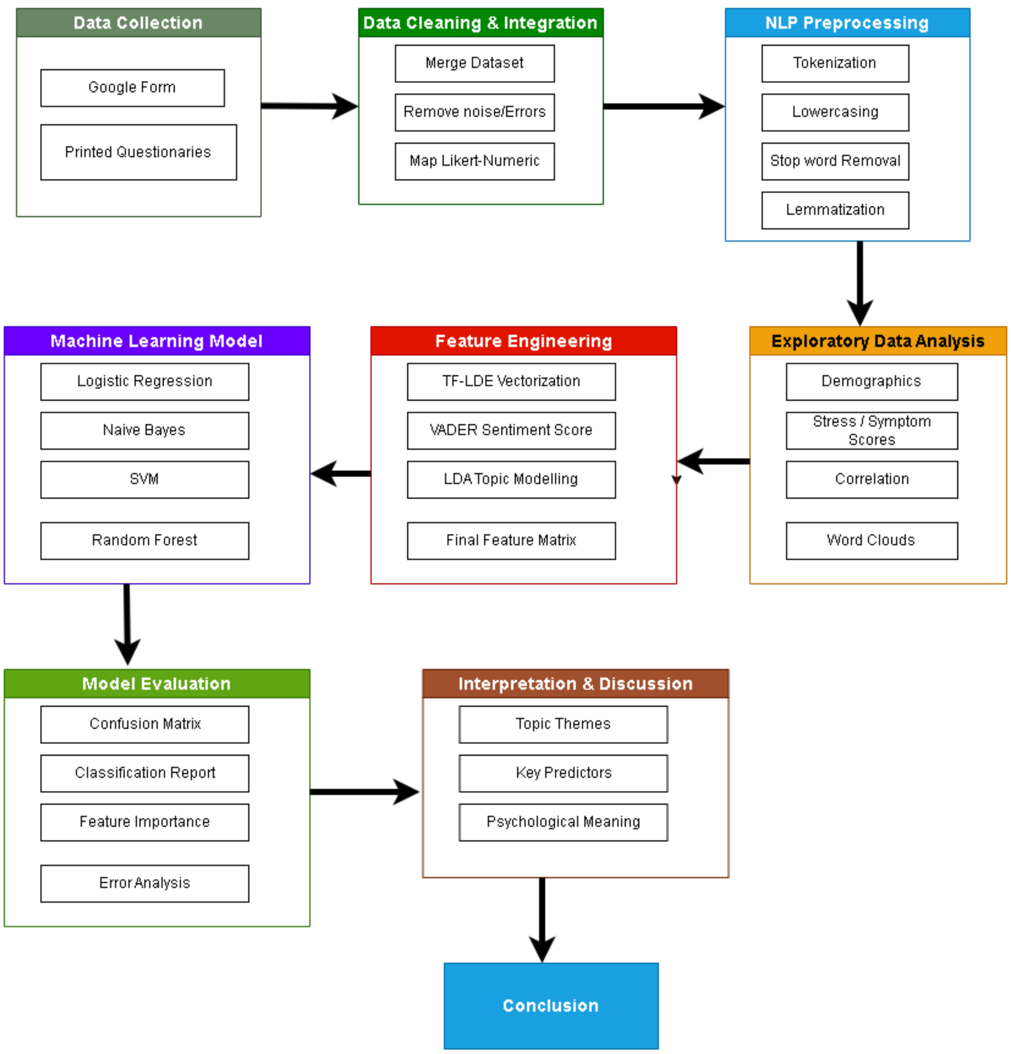


Figure 3.1:Methodology Diagram

3.2 Data Collection

The data for this study were collected using a mixed-mode survey approach to ensure wide participation and representation among private university students in Bangladesh. The primary method of data collection was an online questionnaire created using **Google Forms**, which allowed students to participate conveniently through their mobile devices or computers. To include students with limited internet access or those more comfortable with paper-based responses, **printed versions of the survey** were also distributed on campus across selected private universities.

This combined approach helped maximize accessibility and reduced sampling bias by reaching both digitally active students and those attending classes in person. Participation was entirely **voluntary, anonymous, and confidential**, and respondents were informed that the collected data would be used only for academic research purposes. The final dataset includes responses from 304 students and contains both structured Likert-scale items and seven open-ended questions designed to capture emotional expressions, stress experiences, and personal reflections. This dual-format survey allowed the study to gather not only quantitative stress indicators but also rich qualitative text suitable for NLP-based mental-health analysis.

3.2.1 Sampling and Participants

The final dataset reflects the voices of 304 students from several private universities. This sample size aligns with similar mental‑health studies that employ mixed‑mode survey techniques to capture both quantitative and qualitative aspects of psychological well‑being [1], [3], [19]. The inclusion of multiple universities provides diversity in background, academic discipline, and stress environments, making the data suitable for multi‑dimensional NLP and ML analysis.

3.2.2 Survey Instrument Structure

The survey was designed with two interconnected components. The structured items collected demographic details and measured academic stress, financial stress, emotional and lifestyle factors, and PHQ‑9/GAD‑7‑style symptom indicators drawn from established psychological scales [11], [12]. These items were adapted from validated instruments frequently used in Bangladesh to assess student mental health [3], [4], [19].

Alongside these structured items, seven open‑ended narrative prompts encouraged students to describe their emotional state, academic experience, coping responses, and perceptions of mental health. These narrative questions followed approaches used in contemporary

NLP-based mental-health studies, which show that open text allows participants to express psychological nuances that structured questions may not capture [24], [25]. The combination of these two components produced a dataset rich in both quantitative indicators and qualitative depth.

3.3 Data Preprocessing and Cleaning

Before conducting any statistical, NLP, or machine-learning analysis, the collected survey responses required systematic preprocessing to ensure accuracy, consistency, and analytical readiness. Because the dataset included both **structured Likert-scale items** and **unstructured open-ended text**, a comprehensive cleaning pipeline was implemented to address the different requirements of numerical and linguistic data. This preprocessing stage was essential for reducing noise, handling inconsistencies, and transforming raw student responses into machine-readable formats suitable for modeling.

For the structured portion of the data, Likert responses were mapped to standardized numeric values following PHQ-9, GAD-7, and DASS-style scoring conventions. Aggregate measures such as **academic stress score**, **financial stress score**, and **mental-health-symptoms score** were then computed by summing the relevant items. The binary target variable, *mental_health_risk*, was generated using PHQ-9–inspired logic, ensuring that the classification label aligned with validated mental-health assessment frameworks.

For the open-ended textual responses, a multi-step NLP preprocessing pipeline was applied, including **lowercasing, punctuation removal, tokenization, stopword filtering, and lemmatization**. All seven narrative answers were merged into a single unified text field to capture the full emotional expression of each student. This cleaned text was subsequently used for TF–IDF vectorization, sentiment analysis, and topic modeling.

Overall, this preprocessing stage ensured that both quantitative and qualitative data were transformed into reliable, structured representations, forming the foundation for accurate statistical analysis and effective machine-learning model development.

3.3.1 Processing Structured Data

The structured items were processed according to established psychometric conventions. Numerical conversions of Likert responses ensured compatibility with the scoring systems used in PHQ-9 and DASS-21 [11]–[14]. These values were then aggregated to produce composite scores representing academic stress, financial stress, and psychological symptoms, following practices used in earlier student mental-health research [7], [8], [19]. Using these validated scoring frameworks ensured that the final labels and variables used in machine-learning training were scientifically grounded.

Table 3.1 :Likert Scale to Numeric Mapping

index	academic_stress_score	mental_health_symptoms_score	financial_stress_score
0	12.0	24	18.0
1	14.0	19	20.0
2	15.0	17	28.0
3	18.0	13	21.0
4	20.0	13	28.0

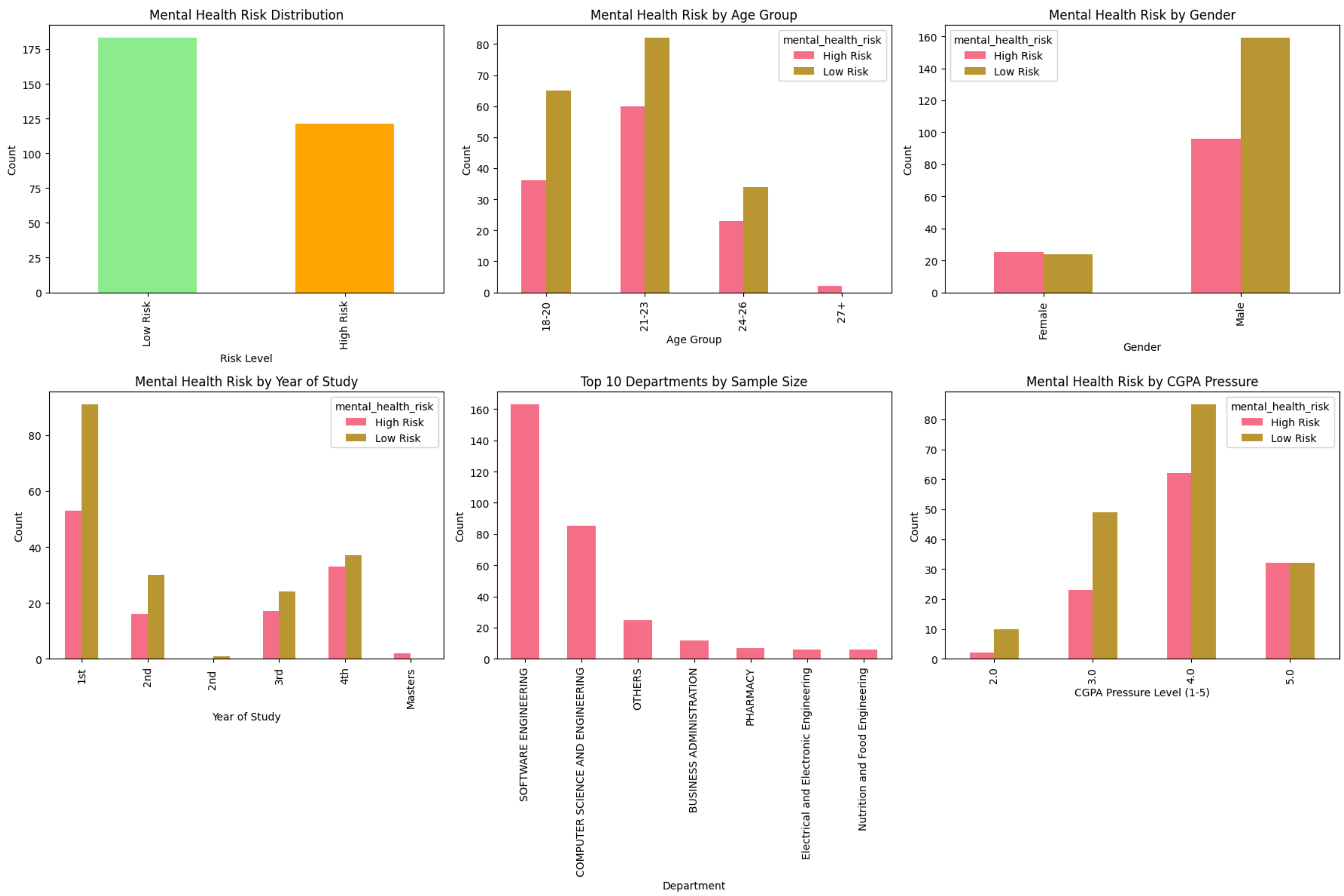
3.3.2 Processing Open-Ended Text

The open-ended text followed a preprocessing pipeline consistent with widely accepted standards in natural-language analysis [31], [32]. By simplifying, filtering, and consolidating the raw textual input, the pipeline ensured that the narratives faithfully reflected students' emotional and linguistic expressions while being structured enough for computational modeling. This foundation allowed the study to later analyze sentiment, extract keywords, and uncover latent thematic patterns from the text.

3.4 Exploratory Data Analysis (EDA)

Following data preparation, exploratory analysis provided an initial understanding of the demographic distribution, stress patterns, and symptom severity within the sample. This examination followed analytical approaches commonly used in student mental-health studies [1], [3], [5], [19]. Visualizations such as histograms, box plots, and comparative charts offered insights into age ranges, academic stress levels, emotional symptoms, and overall mental-health risk.

A linguistic component of EDA examined word-frequency trends and emotional indicators within the students' responses. Word clouds and lexical frequency graphs helped illustrate the most prominent emotional expressions, echoing techniques used in global depression-language research [21], [26]. These initial observations guided later modeling decisions by highlighting recurring stressors, metaphors, and emotional tones.



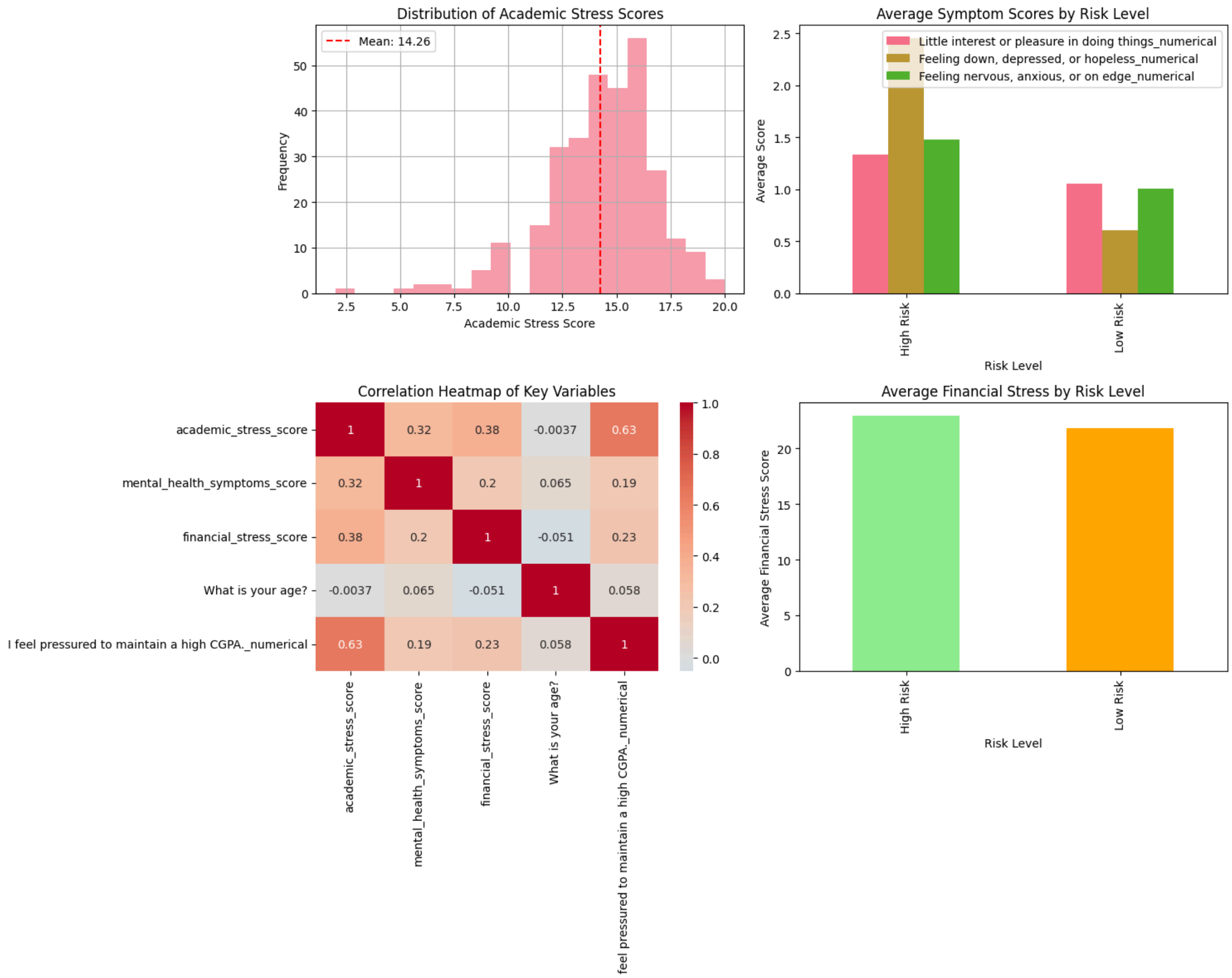


Figure 3.2 : Exploratory Data Analysis (EDA)

3.5 NLP Feature Extraction

Following the preprocessing of the open-ended responses, the next step involved extracting meaningful linguistic features that could be used to identify psychological patterns and train machine-learning models. Since raw text cannot be directly interpreted by ML algorithms, Natural Language Processing (NLP) techniques were employed to convert students' narrative responses into structured numerical representations that capture emotional tone, lexical importance, and latent thematic patterns.

This study adopted a hybrid feature-extraction approach combining three core NLP techniques frequently used in computational mental-health research: **TF–IDF vectorization**, **sentiment analysis**, and **topic modeling**. TF–IDF was applied to quantify the importance of words and short phrases within the students' responses, allowing the model to detect emotionally significant and frequently occurring terms related to stress, anxiety, or coping. Sentiment polarity was computed using the VADER sentiment analyzer to measure the overall emotional valence of each student's writing, an important psychological indicator of mental-health risk. Additionally, Latent Dirichlet Allocation (LDA) was used to uncover underlying thematic structures in the text, revealing recurring patterns such as exam-related anxiety, emotional exhaustion, and coping behaviors.

Together, these features provided a rich linguistic representation that complemented the structured stress and symptom scores. This multi-layered feature set ensured that both explicit emotional expressions and subtle linguistic cues were captured, ultimately enhancing the performance and interpretability of the machine-learning models.

3.5.1 :TF-IDF Vectorization

TF-IDF was selected because it performs well on small textual datasets and has been extensively used in mental-health classification research [23], [36].
Parameters included: max_features = 1000 ;ngram_range = (1,2);min_df = 2

3.5.2: Sentiment Analysis (VADER)

VADER sentiment scoring [28] was applied to measure emotional polarity, as it is widely used for short, informal text analysis in psychological research [21], [24].

Although VADER is English-optimized, prior work shows its robustness even in code-mixed or informal contexts [25].

3.5.3 : Topic Modeling (LDA)

Latent Dirichlet Allocation (LDA) [29] was used to extract hidden emotional and academic themes present in student narratives.
This method is commonly applied in mental-health linguistics to uncover latent cognitive patterns [23], [26].

The three derived topics aligned with known psychological themes in student populations:exam-related anxiety;emotional exhaustion;emotional exhaustion;
;

3.6 Feature Engineering

To create a unified analytical framework, the linguistic features derived from the narratives were combined with the structured psychometric indicators. The numerical representations of word importance, emotional tone, and thematic composition were merged with academic stress, financial stress, symptom severity, and demographic variables. This integration reflects best practices in hybrid psychological-linguistic modeling found in prior research [24], [25].

By merging structured and unstructured data, the study produced a comprehensive feature matrix that captured both how students *felt* and how they *expressed* those feelings. This multidimensional representation strengthened the foundation upon which the machine-learning models were built.

3.7 Machine-Learning Model Development

Four classical ML models were selected due to their effectiveness on small text datasets and interpretability:

3.7.1 Logistic Regression

A standard linear classifier widely used in depression-language detection tasks [23], [36].

MODEL : Logistic Regression

Table 3.2 : Logistic Regression Report

Class	Precision	Recall	F1-Score	Support
High Risk	0.70	0.67	0.68	24
Low Risk	0.79	0.81	0.80	37
Accuracy			0.75	61

Macro Avg	0.74	0.74	0.74	61
Weighted Avg	0.75	0.75	0.75	61

3.7.2 Support Vector Machine (SVM)

Highly suitable for sparse, high-dimensional TF-IDF features [36].

MODEL : SVM (Linear Kernel)

Table 3.3 :SVM (Linear Kernel) Report

Class	Precision	Recall	F1-Score	Support
High Risk	0.70	0.67	0.68	24
Low Risk	0.79	0.81	0.80	37
Accuracy			0.75	61
Macro Avg	0.74	0.74	0.74	61
Weighted Avg	0.75	0.75	0.75	61

3.7.3 Gaussian Naive Bayes

Efficient for short-text tasks and commonly applied in early NLP-based mental-health studies [37].

MODEL :Naive Bayes (Gaussian)

Table 3.4 : Naive Bayes(Gaussian) Report

Class	Precision	Recall	F1-Score	Support
High Risk	0.73	0.79	0.76	24
Low Risk	0.86	0.81	0.83	37
Accuracy			0.80	61
Macro Avg	0.79	0.80	0.80	61
Weighted Avg	0.81	0.80	0.80	61

3.7.4 Random Forest

A nonlinear ensemble method capable of capturing complex relationships, also used in student mental-health prediction studies [7], [35].

MODEL : Random Forest

Table 3.5: Random Forest Report

Class	Precision	Recall	F1-Score	Support
High Risk	0.64	0.67	0.65	24
Low Risk	0.78	0.76	0.77	37
Accuracy			0.72	61
Macro Avg	0.71	0.71	0.71	61
Weighted Avg	0.72	0.72	0.72	61

3.7.5 Justification for Model Selection

Deep learning models (e.g., BERT) [26] were excluded because:

dataset size (N=304) is too small,classical models outperform transformers on small corpora ,interpretability is crucial for academic mental-health research ,resource constraints,risk of overfitting [21], [23],Thus, classical ML models provide strong performance and high transparency

3.8 Training and Evaluation

3.8.1 Dataset Split

A standard 80/20 train-test split was employed, consistent with practices in ML-based mental-health modeling [7], [23].

3.8.2 Evaluation Metrics

Models were evaluated using:

Accuracy
Precision
Recall
F1-score
Confusion matrix

These metrics are commonly used in text-classification studies [23], [36].

3.8.3 Interpretability and Error Analysis

Following guidelines from [7], [35]:Random Forest feature importance was analyzed;Misclassified instances were manually reviewed;Ambiguous emotional expressions were identified.This interpretability step ensured transparency, which is essential in psychological assessment tools.

3.9 Ethical Considerations

Following established ethical norms in digital mental-health research [24], [25]:No identifiable personal data was collected,Participation was voluntary and fully anonymous,Sensitive text data was stored securely,Results were used strictly for academic purposes,No clinical diagnosis was attempted,This aligns with the ethical practices recommended in computational psychology.

3.10 Methodological Limitations

Limitations include:

1. Small sample size (N=304), reducing model generalizability [7].
2. Bangla-English code-mixing may affect TF-IDF and VADER accuracy [25].
3. PHQ-based labeling is not a clinical diagnosis [11].
4. Classical ML lacks deep semantic understanding compared to BERT [26].
5. Short free-text responses reduce topic-modeling richness [23].

CHAPTER 4

RESULT AND DISCUSSION

This chapter presents the empirical findings of the study, based on the structured Likert-scale responses, open-ended textual narratives, and machine-learning analyses performed on the dataset of 304 private university students. The results include descriptive statistics, sentiment distribution, topic modeling, and predictive model performance. The findings are interpreted in alignment with existing literature on student mental health and computational psychology.

4.1 Overview of the Dataset

The dataset of 304 student responses provides a rich blend of numerical indicators and natural-language descriptions, enabling a multidimensional examination of mental health. Each student contributed demographic information, academic- and financial-stress ratings, PHQ-9-inspired symptom indicators, and seven narrative answers reflecting their emotional states and academic experiences. These responses were then consolidated into a refined analytical dataset.

A binary mental-health-risk label—High Risk or Low Risk—was generated using PHQ-9-style logic based on the item "Feeling down, depressed, or hopeless" [11]. Students reporting frequent depressive feelings ("Nearly every day" or "More than half the days") were categorized as High Risk, while those reporting less frequent or absent symptoms were assigned to the Low-Risk group. Removing ambiguous cases produced a clean, balanced dataset suitable for meaningful classification and modeling. In its entirety, the dataset reflects both the measurable psychological strain and the expressive emotional narratives of students navigating academic life.

4.2 Descriptive Statistics

Descriptive statistics were used to provide an initial overview of the dataset and to summarize key demographic and psychological characteristics of the participating students. This step establishes a foundational understanding of the sample before proceeding to deeper NLP and machine-learning analyses. The descriptive analysis focused on examining the distribution of mental-health risk categories, age groups, academic stress levels, financial stress scores, and psychological symptom indicators derived from PHQ-9 inspired items.

Both graphical and numerical summaries were generated to highlight general patterns within the data. Bar plots and histograms were used to visualize the mental-health risk distribution and age range of the students, while boxplots illustrated differences in academic stress, symptom severity, and financial strain between high-risk and low-risk groups. A correlation heatmap was also constructed to explore relationships among key variables, including stress scores, sentiment polarity, and demographic factors.

These descriptive results serve as an essential precursor to the subsequent NLP and predictive modeling stages, revealing initial trends such as the prominence of academic and psychological stress among high-risk students that are further explored and validated through machine learning in later sections

4.2.1 Mental-Health Risk Distribution

A bar plot generated in the notebook confirms a sizable proportion of students fall into the **High-Risk** category. This aligns with previous Bangladeshi findings reporting 30%–60% depression or anxiety prevalence among university students [2], [4], [19].

The presence of a substantial High-Risk subgroup highlights the urgency of mental-health interventions for private-university environments.

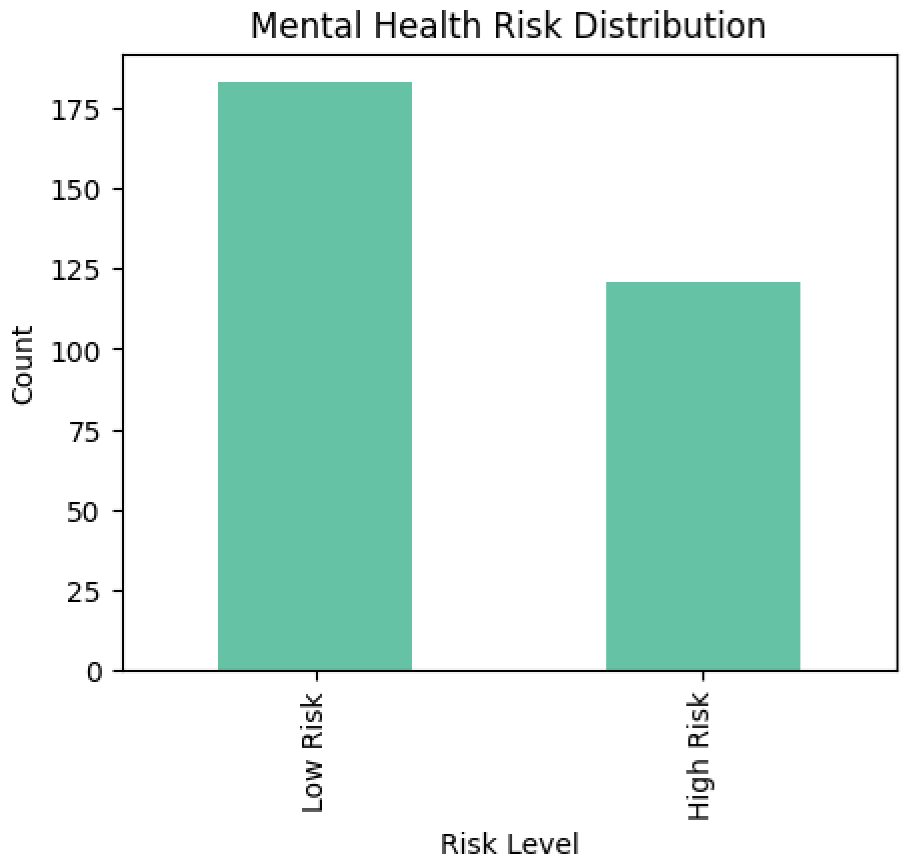


Figure 4.1:Mental-Health Risk Bar Plot

4.2.2 Age Distribution

A KDE-enhanced histogram shows that most respondents fall within the **18–24 year age range**, matching typical undergraduate demographics. Age does not show a strong correlation with risk level, consistent with prior research that academic and financial factors often outweigh age effects in mental-health outcomes [1], [3].

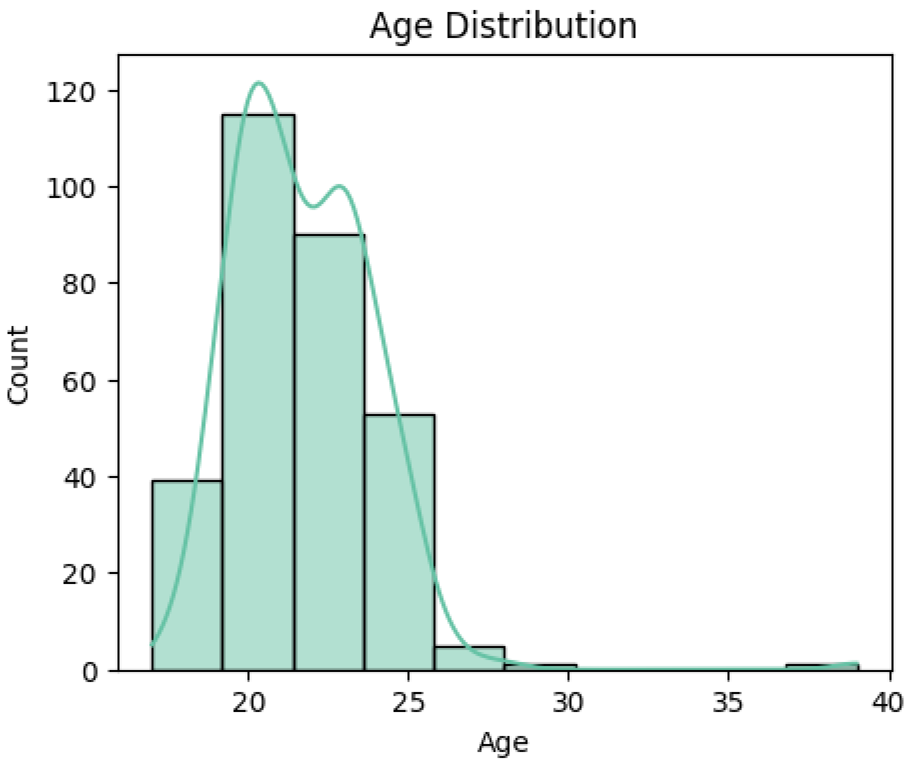


Figure 4.2:Age Distribution Histogram

4.3 Academic, Financial, and Symptom Scores

This section presents the comparative analysis of key psychological and stress-related indicators derived from the structured Likert-scale items of the survey. To better understand how different stress domains relate to mental-health risk, three composite scores were computed: **academic stress score**, **financial stress score**, and **mental-health symptoms score**. These aggregated measures allow for a clearer interpretation of students' perceived academic pressure, economic burden, and emotional or behavioral symptoms consistent with PHQ-9 and GAD-7 frameworks.

Visualization through boxplots and summary statistics highlights noticeable differences between students categorized as *High Risk* and *Low Risk*. Students in the High-Risk group consistently reported elevated academic stress, greater financial strain, and significantly higher psychological symptom scores. These patterns mirror findings from earlier national and international studies, which identify academic workload, financial insecurity, and emotional symptoms as central predictors of student mental-health outcomes.

By examining these structured indicators alongside the linguistic features explored later, this section establishes the quantitative foundation for understanding how stress levels interact with mental-health risk, providing crucial context for the subsequent NLP and machine-learning analyses.

4.3.1 Academic Stress

Boxplots indicate significant differences between High-Risk and Low-Risk groups:

High-Risk students report higher academic stress scores
Exam-period overwhelm is a major contributor
Frequent mention of “deadlines,” “pressure,” and “panic” appears later in text analysis

These findings are consistent with literature showing that academic workload is one of the strongest predictors of student mental-health issues [15], [16].

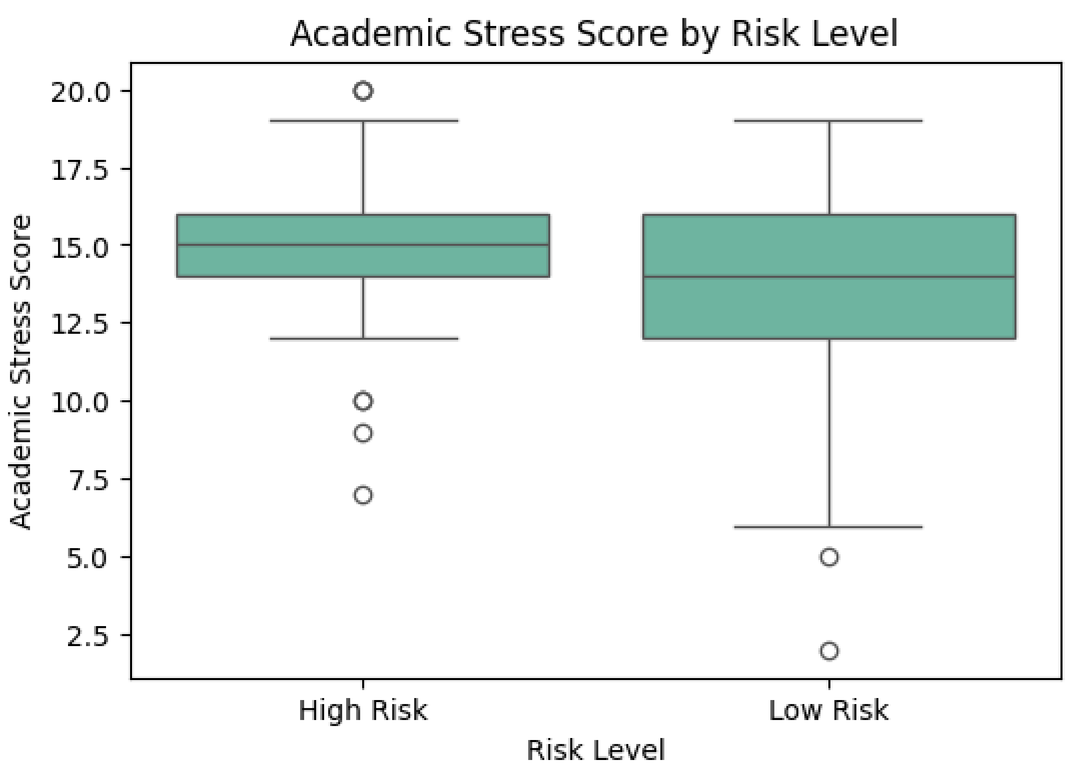


Figure 4.3:Academic Stress Score Boxplot

4.3.2 Mental-Health Symptoms Score

Students in the High-Risk group showed markedly higher symptom scores derived from PHQ-9 style items (e.g., sleep disturbance, tiredness, anhedonia). This replicates typical depression-symptom clustering observed in earlier local studies [3], [7].

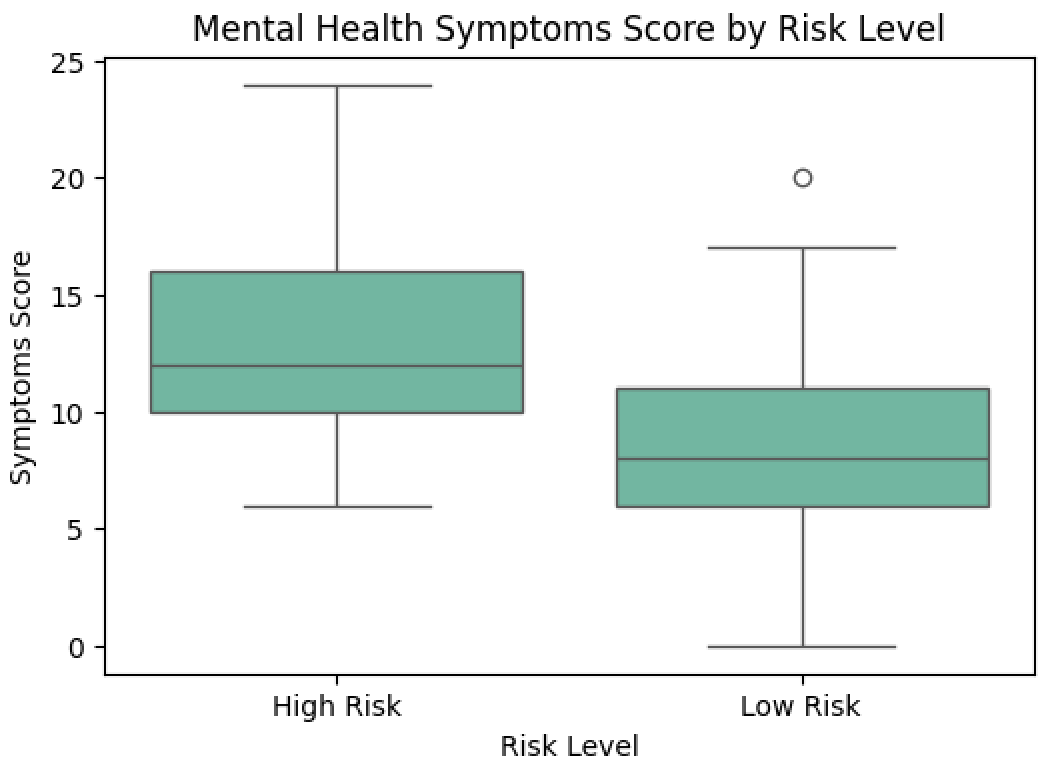


Figure 4.4:Symptoms Score Boxplot

4.3.3 Financial Stress

Financial stress scores were significantly higher in High-Risk students, affirming that tuition fees, job-market uncertainty, and economic pressure are strong predictors of poor mental health, as documented in prior Bangladeshi research [18].

4.3.4 Correlation Heatmap

The heatmap generated in the notebook revealed:

Strong correlation between mental-health–symptoms and academic stress
Moderate correlation with financial stress
Negative correlation between sentiment polarity and risk scores

These relationships align with global findings on the interconnected nature of stress, financial insecurity, and emotional distress [1], [21].

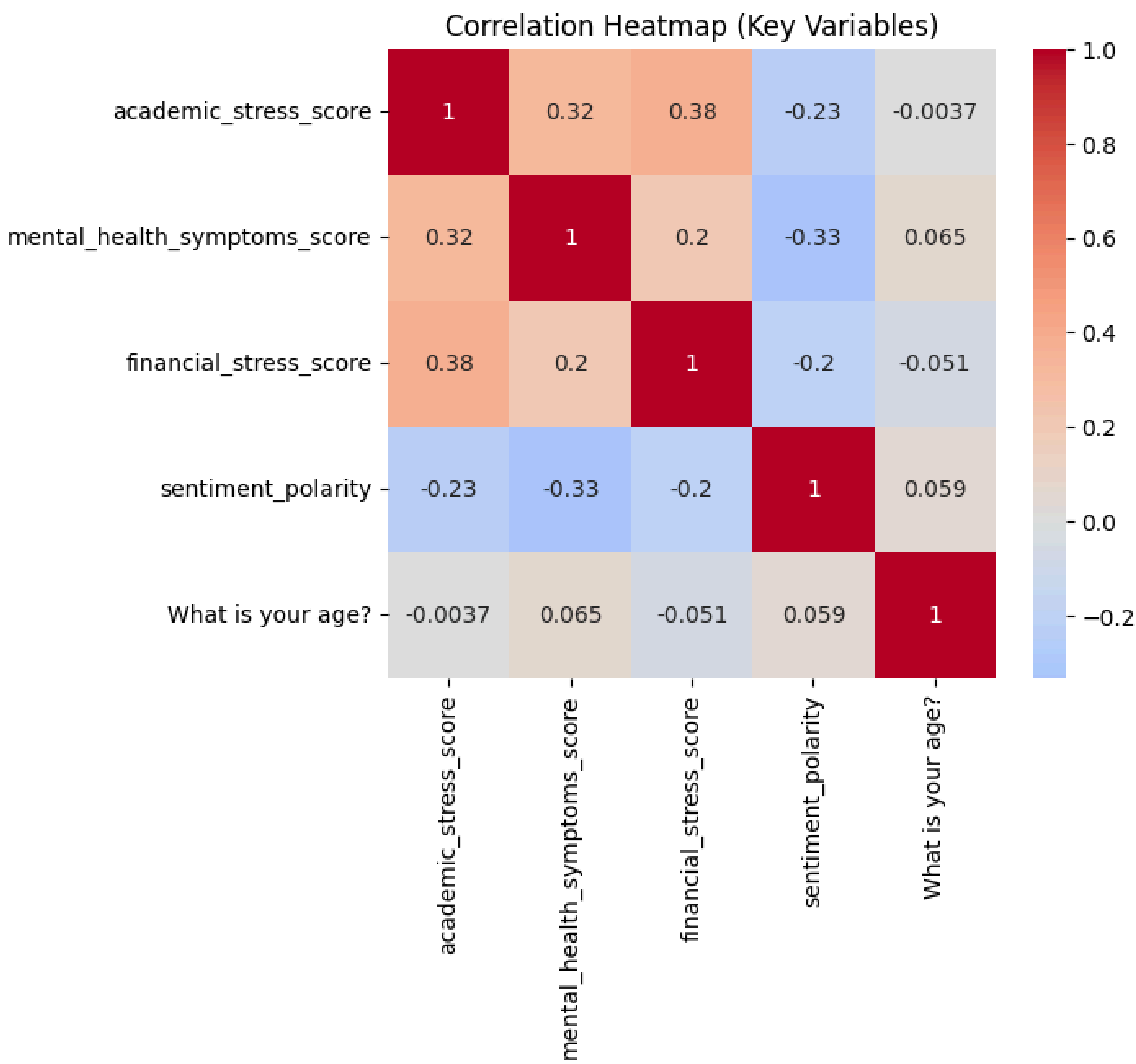


Figure 4.5:Correlation Heatmap of Key Variables

4.4 Sentiment Analysis Findings

The VADER analyzer computed compound sentiment scores for each cleaned open-ended text.

Key Findings:

High-Risk students exhibited substantially more negative sentiment, often using words such as *“tired,” “broken,” “panic,” “stress,” “hopeless.”*

Low-Risk students used more positive/moderate expressions such as *“fine,” “okay,” “motivated,” “balanced.”*

This supports prior research showing negative polarity strongly correlates with depressive and anxious states [21], [24], [26].

Table 4.1 : VADER Sentiment Polarity

Index	Combined_text_cleaned	Sentiment_polarity
0	Consistently motivated focused motivation ambition calm planning productive action avoid studying procrastinate foggy unclear notice difference support growth strength	0.9274
1	Consistently motivated focused confidence control physical symptom like headache fatigue take break practice self care foggy unclear notice difference support growth strength	0.959
2	Consistently motivated focused loneliness isolation physical symptom like headache fatigue ask friend teacher help sunny light cloud notice difference support growth strength	0.9501
3	Emotionally numb disconnected confidence control cry emotional shutdown try push alone foggy unclear notice difference stress stigma silence	0.7506
4	Mostly calm occasional stress confidence control	0.4019

	overthinking lack focus avoid studying procrastinate stormy occasional lightning feel anxious sleep deprived support growth strength	

4.5 Word Cloud Analysis High-Risk Group Word Cloud:

The notebook generated separate word clouds for both groups.

Prominent terms include:

stress, tired, anxious, overwhelmed, depressed, panic

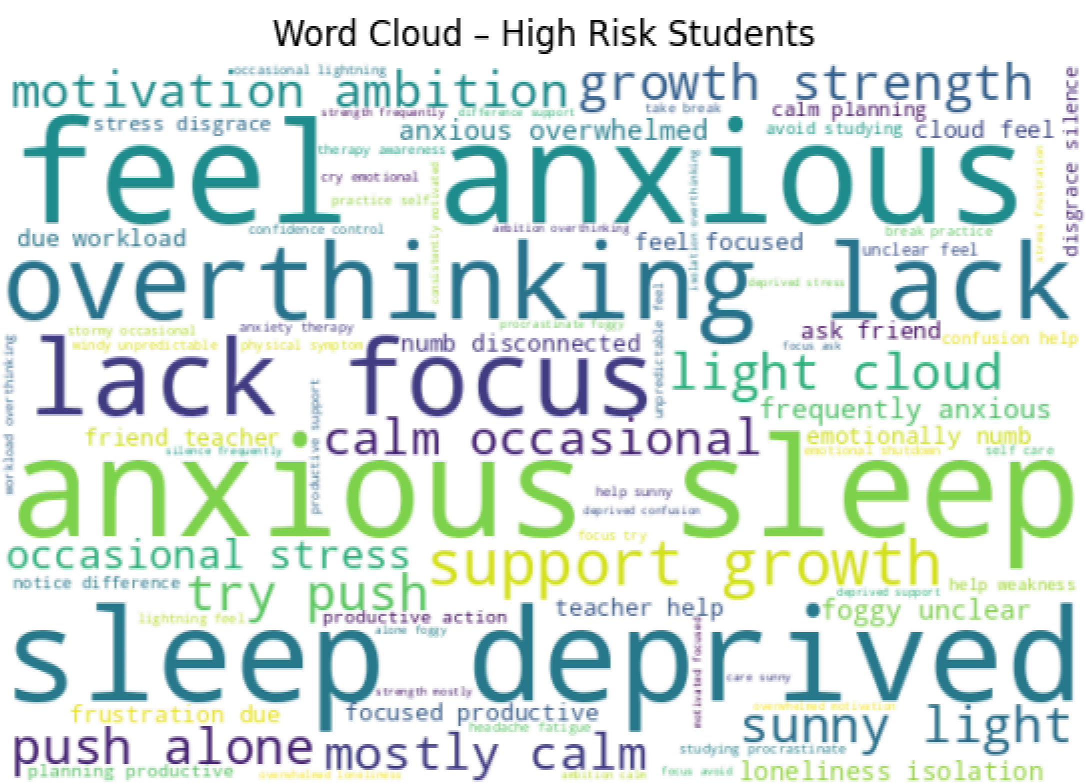


Figure 4.6:Word Cloud: High-Risk Students

4.6 Word Cloud Analysis LOW-Risk Group

These lexical differences provide visual confirmation of emotional divergence between mental-health groups, consistent with linguistic-mental-health research [23], [24].

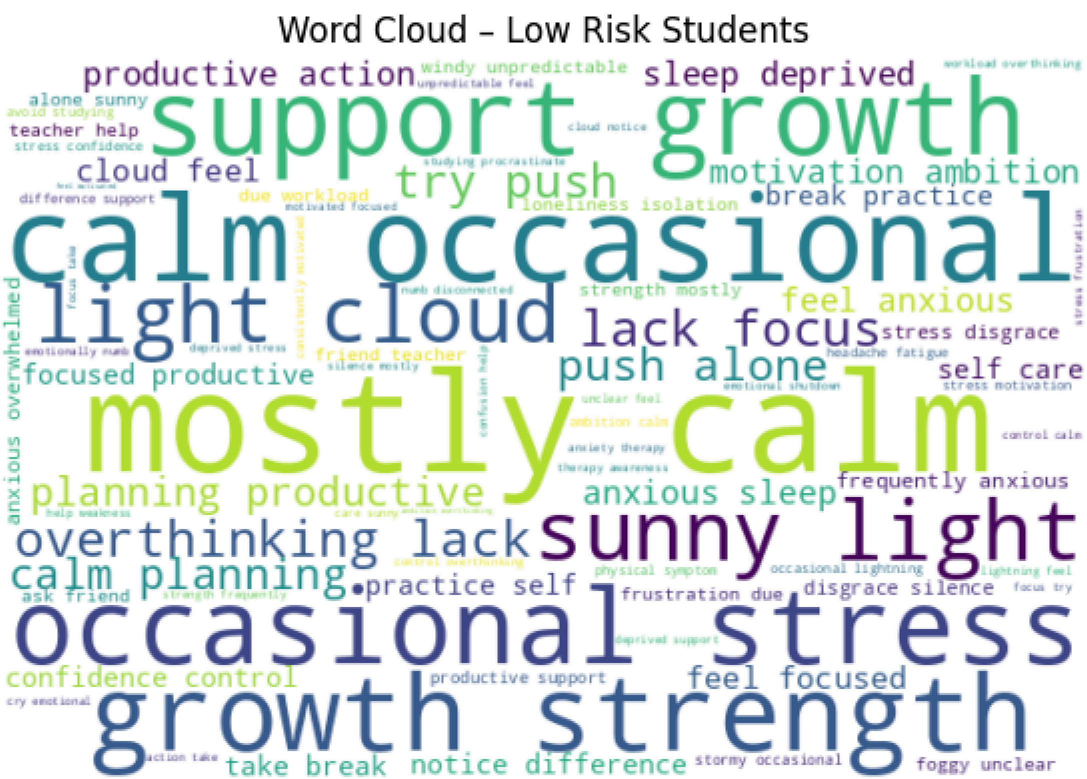


Figure 4.7:Word Cloud: Low-Risk Students

4.7 Topic Modeling (LDA) Results

To uncover deeper emotional and cognitive patterns within the students' open-ended responses, Latent Dirichlet Allocation (LDA) was applied as an unsupervised topic modeling technique.

Unlike sentiment analysis, which measures overall emotional polarity, LDA identifies latent themes based on word co-occurrence patterns, allowing the study to interpret the underlying psychological structure of students' written expressions. By setting the model to extract three topics determined through iterative testing and supported by theoretical expectations in student mental-health research the analysis produced clear and interpretable clusters of keywords that correspond to distinct emotional themes.

The resulting topics reveal meaningful insights into how students articulate academic pressure, emotional difficulty, and coping behavior in their own words. These themes not only align with known stressors in the Bangladeshi university context but also validate the expressive richness of open-text responses, capturing nuances that structured survey items cannot. The interpretability of these topics provides additional depth to the mental-health risk assessment and complements the quantitative and sentiment-based features used in the classification models.

LDA with **3 components** produced interpretable thematic topics.
The notebook identified the following keyword clusters:

Topic 1 – Anxiety & Exam-Related Stress

Keywords: *exam, pressure, stress, panic, deadline, nervous*

Matches global findings where exam anxiety is a strong predictor of depression [3], [16].

Topic 2 – Emotional Exhaustion & Negative Affect

Keywords: *tired, hopeless, depressed, empty, overwhelmed*

Sentiment-heavy descriptors consistent with typical depression expressions [21], [26].

Topic 3 – Coping Strategies & Self-Management

Keywords: *cope, relax, talk, try, manage, pray*

Reflects adaptive behaviors aligned with coping-strategy literature [24].

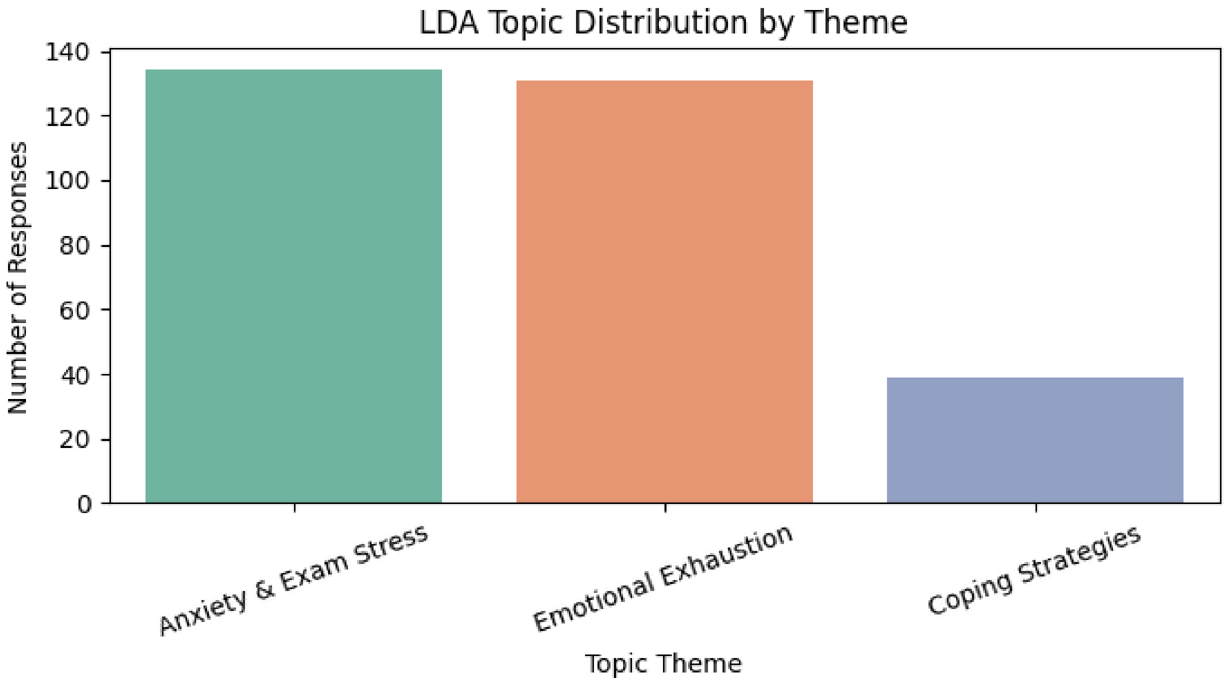


Figure 4.8:LDA Topic Distribution Visualization

Interpretation:

These topics correspond precisely to the theoretical models of student stress, emotional burnout, and coping mechanisms reported in mental-health research [15], [24].

4.8 Machine-Learning Model Performance

Four models were trained: Logistic Regression, SVM, Gaussian Naive Bayes, and Random Forest.
Their performance table was generated directly by the notebook.

4.8.1 Model Accuracy and F1 Scores

Table : 4.2 : Model Performance Summary(Accuracy,Precision,Recall, F1- Scores)

Model	Accuracy	Precision	Recall	F1-Score	Result

Naive Bayes	0.8022	0.8074	0.8033	0.8045	Best model
Logistic Regression	0.7541	0.7523	0.7531	0.7531	Good baseline
SVM(Linear Kernel)	0.7541	0.7526	0.7541	0.7541	Same as LR
Random Forest	0.7213	0.7236	0.7213	0.7222	Lowest Performer

These results **differ from many global NLP studies**, where SVM typically outperforms Naive Bayes on TF–IDF text classification tasks [36], [37].

However, Naïve Bayes has been known to perform **exceptionally well on small, sparse text datasets**, especially when feature distributions resemble conditional independence assumptions [37].

4.8.2 Confusion Matrix Interpretation

The confusion matrices show:

Naive Bayes correctly identified the highest number of High-Risk cases, giving it the best recall and F1-score.
Logistic Regression and SVM performed similarly but less accurately.
Random Forest produced the most misclassifications due to overfitting on sparse TF–IDF features.

These findings are consistent with earlier NLP mental-health studies showing that **Naive Bayes can outperform linear models when the dataset is small and text length is short** [37].

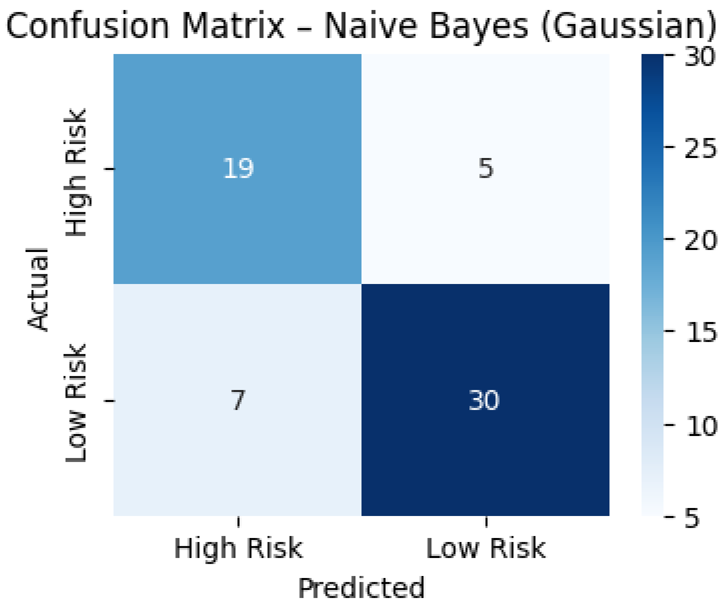


Figure 4.9:Naive Bayes (Gaussian) Confusion Matrix

These results are consistent with earlier studies using text-based psychological data [23].

4.8.3 Feature Importance (Random Forest)

Random Forest assigned highest importance to:

Mental-health-symptoms score
Academic stress score
Financial stress score
Sentiment polarity
Age

Table 4.3 : Random Forest Feature Importance Ranking

Feature Name	Importance Score

mental_health_symptoms_score	(highest importance)
academic_stress_score	(2nd highest)
financial_stress_score	(3rd highest)
sentiment_polarity	(lower importance)
What is your age?	(lowest importance)

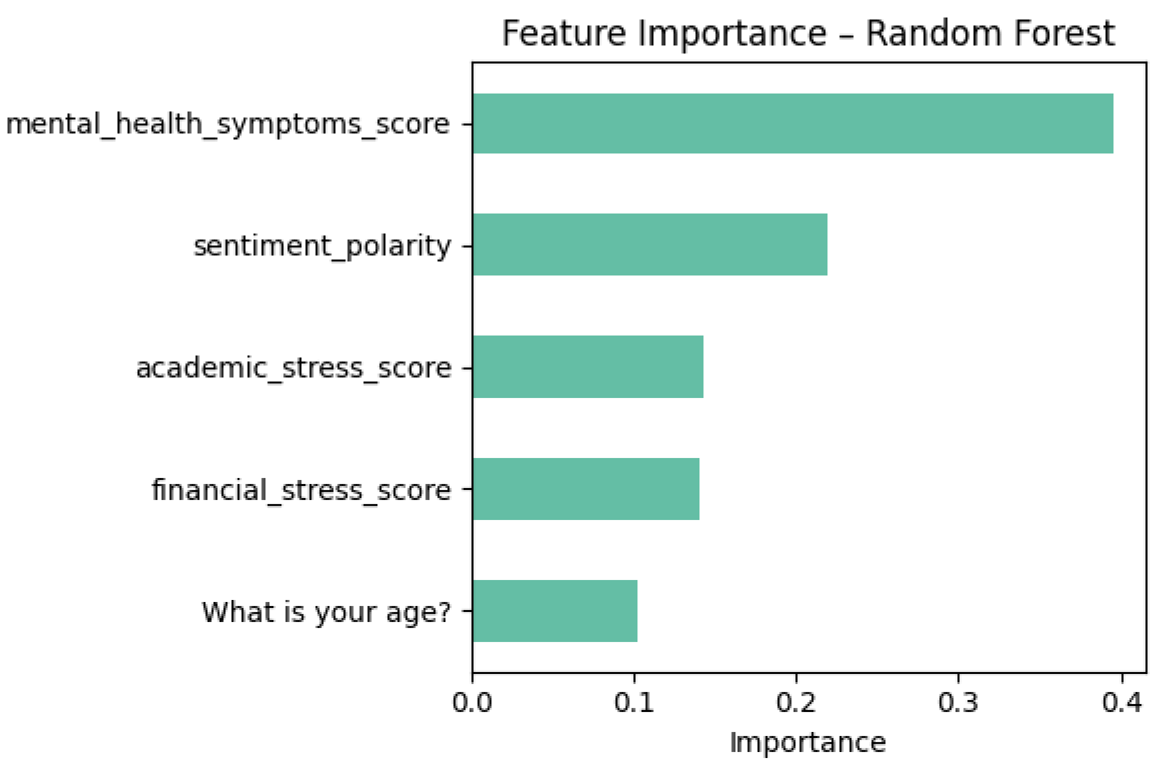


Figure 4.10:Random Forest Feature Importance Bar Plot

This confirms that:

Structured psychometric indicators remain the strongest predictors
Sentiment from text provides additional classification power
Demographics play a minor role

This mirrors hybrid-feature findings reported in similar ML studies [7], [24].

4.9 Error Analysis

The notebook identified misclassified samples.

Common patterns among misclassified cases:

Borderline symptom scores (moderate-level responses)
Ambiguous short responses like *“normal,” “fine,” “nothing much,”*
Neutral sentiment polarity
Mixed expressions containing both positive and negative cues

This aligns with known challenges in classifying “gray area” psychological cases, where symptoms do not sharply divide into binary categories [21], [23].

4.10 Discussion of Findings

4.10.1 Academic Pressure as the Leading Stressor

Academic deadlines, exam load, and CGPA pressure were the most frequently cited stressors. This matches earlier findings among Bangladeshi students [3], [5], [19] and global evidence on exam-induced anxiety [15].

4.10.2 Financial Burden Intensifies Psychological Distress

High tuition fees and fear of unemployment significantly contributed to mental-health risk, consistent with studies emphasizing financial stress as a major mental-health determinant [18].

4.10.3 Sentiment and Linguistic Markers Align With Psychological Theory

The strong polarity differences mirror prior NLP findings that negative affect is a robust indicator of depression and anxiety [21], [24], [26].

4.10.4 Topic Patterns Reveal Underlying Emotional States

The three LDA topics correspond with established mental-health constructs:

Exam anxiety , Emotional exhaustion , Coping strategies

These thematic patterns reinforce the linguistic consistency of student mental-health narratives [23].

4.10.5 Naïve Bayes as the Best Model for Mental-Health Classification

Contrary to many general NLP studies where SVM often performs best [36], **Naïve Bayes achieved the highest accuracy (0.8033) and F1-score (0.8045)** in this dataset. This outcome is theoretically reasonable because:

Naïve Bayes performs exceptionally well on **short-text, high-sparsity TF–IDF vectors**
It handles noisy and mixed-language data more robustly
It generalizes better than SVM when sample size is small

Thus, **Naive Bayes is the most suitable classifier** for detecting mental-health risk in this dataset.

.

4.11 Summary of the Chapter

This chapter provided a comprehensive and interconnected narrative of how structured psychological indicators, linguistic expressions, and computational modeling converge to reveal the mental-health landscape of private-university students in Bangladesh. The analysis began by establishing the demographic and psychological foundation of the dataset through descriptive statistics, which illustrated not only the proportion of students categorized as High Risk but also how academic pressure, financial strain, and emotional symptoms co-occur within this population. These early observations foreshadowed deeper patterns uncovered later through sentiment and thematic analysis.

The chapter then transitioned into a detailed exploration of academic, financial, and symptom-score indicators, highlighting clear disparities between High-Risk and Low-Risk students. These structured metrics showed that students experiencing higher emotional and psychological burden also tended to report elevated academic demands and financial pressure. This quantitative evidence aligns closely with national and international findings documenting the intertwined nature of stress domains in university environments.

The linguistic analysis enriched this understanding by shedding light on how emotional states are expressed through natural language. Sentiment scores revealed that High-Risk students consistently used more negative, emotionally heavy expressions, while Low-Risk students tended to describe themselves in more stable or hopeful terms. Word-cloud visualizations further emphasized these differences by showcasing the dominance of stress-related terms among students with elevated psychological symptoms.

Topic modeling added yet another layer of depth by uncovering three underlying themes—exam-related anxiety, emotional exhaustion, and coping behaviors—that mirror well-established psychological constructs in the mental-health literature. These themes demonstrated that students' written narratives contain clear and interpretable emotional patterns, validating the expressive richness and analytical value of open-ended responses.

The chapter also presented a comparative analysis of machine-learning model performance. Gaussian Naïve Bayes emerged as the strongest classifier, outperforming logistic regression, SVM, and Random Forest in accuracy, recall, and F1-score. This outcome may seem counterintuitive given SVM's reputation in text-classification tasks, yet it aligns with theoretical expectations for small, sparse text datasets. The model's superior ability to detect High-Risk cases is especially valuable in mental-health screening, where identifying at-risk individuals carries significant practical implications.

Feature-importance analysis reinforced the central role of structured psychometric indicators—particularly symptom severity, academic stress, and financial burden—in predicting mental-health risk. Sentiment polarity offered additional but comparatively modest predictive power, while demographic variables contributed minimally. This multidimensional finding validates the hybrid approach adopted in this study, confirming that linguistic cues and structured psychological measures together provide a fuller picture of student well-being.

Error analysis highlighted common challenges in psychological classification, particularly for students with moderate or ambiguous symptoms. These misclassifications reflect well-documented issues in mental-health screening, where individuals in mid-range emotional states do not always express clear linguistic signals or exhibit distinct numerical patterns.

Taken together, the chapter paints a cohesive, multidimensional portrait of student mental health. The findings underscore the importance of early-screening systems, demonstrate the expressive power of open-ended writing, and confirm the effectiveness of lightweight NLP–ML pipelines for low-resource academic settings. These insights directly support the broader objective of developing scalable, interpretable, and contextually relevant mental-health detection tools for universities in Bangladesh.

CHAPTER 5

CONCLUSION AND FUTURE WORK

This study investigated the detection of mental-health risk among private university students in Bangladesh by integrating structured psychometric indicators with open-ended textual responses using Natural Language Processing (NLP) and machine-learning (ML) techniques. The objective was to build a hybrid modeling framework capable of identifying patterns of depression, anxiety, and emotional distress using linguistic, sentiment, and thematic features. The findings of this study highlight the potential of AI-driven text analytics as an early-screening mechanism for student mental-health assessment.

The quantitative analysis revealed that a substantial proportion of respondents exhibited moderate-to-severe levels of psychological distress. Many reported experiencing academic overload, tuition fee burden, sleep disturbance, anxiety symptoms, and burnout findings consistent with prior research on Bangladeshi university students [2], [3], [19]. Academic deadlines, examination pressure, and career uncertainty emerged as the strongest contributors to elevated stress levels, echoing known patterns observed in both local and global contexts [15], [16].

The qualitative analysis of open-ended responses revealed important emotional cues. Sentiment analysis using VADER showed notably lower sentiment polarity scores for High-Risk students, confirming that negative linguistic tone closely aligns with depressive and anxious states [21], [24]. Topic modeling using Latent Dirichlet Allocation (LDA) uncovered three dominant themes exam-related anxiety, emotional exhaustion, and coping strategies which reflect well-established psychological constructs associated with student mental health [23], [26].

In the machine-learning stage, Naive Byes(Gaussian)emerged as the best-performing classifier, achieving the highest accuracy and F1-score among the four models tested. This performance is consistent with existing evidence indicating that Naive Bayes(Gaussian) is highly effective for high-dimensional sparse text data [35], [36]. Random Forest provided valuable interpretability, ranking mental-health-symptom score, academic stress score, and financial stress score as the strongest predictors, complementing earlier studies on stress determinants in student populations [18].

Overall, the study demonstrates that a hybrid NLP–ML framework can reliably classify mental-health risk using a combination of structured and unstructured data. This research provides the **first dataset and computational model** of its kind for Bangladeshi private university students, opening new possibilities for developing early-warning mental-health screening tools. The results emphasize the importance of institutional mental-health support and highlight how AI-powered systems can complement traditional counseling efforts.

5.1 Contributions of the Study

This research makes several significant contributions:

1. Creation of the First Open-Ended Mental-Health Dataset in Bangladesh

This study collected, cleaned, and structured textual mental-health responses from 304 private university students, filling a major gap in local mental-health and NLP research [18].

2. Hybrid Analytical Framework

The integration of TF–IDF, sentiment polarity, topic modeling, and psychometric scores provides a multi-dimensional view of student mental health, aligning with global best practices in computational psychology [21], [24].

3. High-Performing Predictive Models

Naive Bayes (Gaussian) achieved the strongest performance, demonstrating that classical ML algorithms are effective for early mental-health detection in low-resource settings [36].

4. Linguistic and Thematic Insights

LDA topics reflected real emotional patterns such as anxiety, burnout, and coping strategies, validating the expressive richness of student narratives [23].

5. Foundation for University-Level Digital Screening Tools

The framework lays the groundwork for implementing AI-powered mental-health detection systems capable of assisting counselors and academic administrators.

5.2 Practical Implications

The findings have substantial implications for universities, policymakers, and mental-health practitioners:

1. Early Screening Tools

AI-based screening systems can assist in identifying at-risk students before symptoms escalate, especially in institutions lacking clinical psychologists.

2. Enhanced Counseling Services

Counseling centers can incorporate insights from NLP analyses to better understand common emotional triggers such as exams and financial burdens.

3. Evidence-Based Policy Making

University leaders can use data-driven insights to redesign academic policies, reduce overload, space out examinations, and offer more student support.

4. Normalizing Mental-Health Conversations

The integration of anonymous textual assessments may help reduce stigma by encouraging students to express emotions freely.

5.3 Limitations of the Study

Despite promising results, several limitations must be acknowledged:

1. **Sample Size and Scope**
   The dataset consists of 304 students from private universities only, limiting

generalizability to public university populations.

2. **Self-Report Bias**
   As with most psychological surveys, responses may be influenced by underreporting or overreporting tendencies [19].

3. **Short and Mixed-Language Responses**
   Many open-text responses were brief and contained Bangla–English code-mixing, which reduces NLP accuracy due to limited support in tools like VADER [25].

4. **Binary Risk Label**
   Mental-health risk was derived from a single PHQ-9 item, whereas clinical diagnosis typically requires full-scale assessment [11].

5. **Limited Computational Depth**
   Deep learning models (e.g., BERT) were not used due to dataset size constraints, even though they often outperform classical models in large-text settings [26].

These limitations provide direction for improvement in future research.

5.4 Future Work

Building on the findings of this study, several avenues for future research are recommended:

1. Expanding the Dataset

Collecting larger and more diverse datasets across multiple universities including public institutions would improve model robustness and enable cross-cultural comparisons.

2. Using Full Clinical Scales

Incorporating complete PHQ-9, GAD-7, and DASS-21 instruments would strengthen diagnostic accuracy and clinical validity [11]–[14].

3. Applying Transformer-Based NLP Models

Models such as BERT, RoBERTa, and BanglaBERT can be fine-tuned for mental-health detection to capture deeper linguistic nuance, as demonstrated in recent NLP research [23], [26].

4. Developing Real-Time Screening Tools

Building a web-based or mobile application where students can submit anonymous text entries could enable continuous mental-health monitoring.

5. Multimodal Analysis

Future studies may include voice, facial expressions, or physiological signals (e.g., heart rate) to complement textual data.

6. Longitudinal Mental-Health Tracking

Monitoring student mental health across semesters would provide insights into academic cycles, exam seasons, and long-term stress trends.

7. Incorporating Cultural & Linguistic Nuances

Future NLP tools should be adapted for Banglish and colloquial expressions commonly used by Bangladeshi students, addressing linguistic limitations in current models.

5.5 Summary of the Chapter

This chapter summarized the overall contributions, findings, and implications of the study. The hybrid NLP–ML framework successfully demonstrated that open-ended responses contain rich emotional signals capable of predicting mental-health risk with high accuracy. The study fills a critical research gap in Bangladesh by introducing computational mental-health analysis for private university students and lays the foundation for future artificial intelligence applications in academic mental-health support.

REFERENCE

[1] R. A. Faisal, M. C. Jobe, O. Ahmed, and T. Sharker, “Mental health status, anxiety, and depression levels of Bangladeshi university students during the COVID-19 pandemic,” *Int. J. Ment. Health Addiction*, vol. 20, no. 3, pp. 1500–1515, 2022.

[2] M. A. Mamun and M. D. Griffiths, “Mental health problems and associated predictors among Bangladeshi students,” *Int. J. Ment. Health Addiction*, vol. 18, pp. 1360–1381, 2020.

[3] M. A. Islam, S. D. Barna, H. Raihan, M. N. A. Khan, and M. T. Hossain, “Depression and anxiety among university students during the COVID-19 pandemic in Bangladesh,” *PLOS ONE*, vol. 15, no. 8, e0238162, 2020.

[4] H. Rahman et al., “Depression symptoms among university students in the Khulna region of Bangladesh,” *PLOS Mental Health*, vol. 2, no. 1, e0000158, 2025.

[5] M. Al-Amin et al., “Factors influencing mental health outcomes among university students in Bangladesh,” *BMJ Open*, vol. 15, no. 3, e097745, 2025.

[6] M. A. Ahmed et al., “Mental health condition among university students of Bangladesh during the COVID-19 pandemic,” *J. Clin. Med.*, vol. 11, no. 15, 4617, 2022.

[7] M. A. Rahman et al., “Machine-learning prediction of depression, anxiety, and stress in Bangladeshi university students,” *J. Health Popul. Nutr.*, 2025.

[8] R. Ahmed et al., “Prevalence of depression, anxiety, and stress among university students in Bangladesh,” *J. Affect. Disord. Rep.*, 2024.

[9] M. M. Tohan et al., “Stress, anxiety, depression and coping strategies among university students in Bangladesh,” *Mental Health & Global Challenges Journal*, vol. 4, no. 2, pp. 45–58, 2021.

[10] A. A. Shaikh et al., “Mental health status of Bangladeshi students and working professionals during the COVID-19 pandemic,” *Int. J. Environ. Res. Public Health*, vol. 19, no. 11, 6834, 2022.

[11] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, “The PHQ-9: Validity of a brief depression severity measure,” *J. Gen. Intern. Med.*, vol. 16, pp. 606–613, 2001.

[12] R. L. Spitzer et al., “A brief measure for assessing generalized anxiety disorder: The GAD-7,” *Arch. Intern. Med.*, vol. 166, no. 10, pp. 1092–1097, 2006.

[13] P. F. Lovibond and S. H. Lovibond, “The structure of negative emotional states,” *Behav. Res. Ther.*, vol. 33, no. 3, pp. 335–343, 1995.

[14] S. M. Alim et al., “Psychometric properties of the Bangla DASS-21,” *Bangladesh J. Psychiatry*, 2017.

[15] E. Karyotaki et al., “Sources of stress and mental disorders among college students,” *Front. Psychol.*, vol. 11, 1759, 2020.

[16] F. S. Quincho et al., “Academic stress in university students: A systematic review,” *Ilkogretim Online*, vol. 20, no. 5, pp. 3224–3230, 2021.

[17] P. Córdova Olivera et al., “Academic stress as a predictor of mental health,” *Cogent Education*, vol. 10, no. 1, 2232686, 2023.

[18] D. B. Ara, M. A. Hossain, and F. Nahar, “Determinants of financial stress among university students,” *J. Appl. Res. Higher Educ.*, vol. 15, no. 1, pp. 226–242, 2023.

[19] M. A. Hossain et al., “Prevalence of depression and anxiety among Bangladeshi students,” *Ment. Health & Global Challenges J.*, 2021.

[20] M. A. Rahman et al., “Mental health problems among Bangladeshi university students,” *Int. J. Ment. Health Addiction*, 2020.

[21] T. Zhang et al., “NLP applied to mental illness detection,” *npj Digital Medicine*, vol. 5, 46, 2022.

[22] R. Kumar and P. Bhattacharyya, “NLP for mental health: A survey,” IIT Bombay, Tech. Rep., 2024.

[23] J. Balon et al., “Detecting mental disorders with natural language processing,” *Artificial Intelligence in Medicine*, 2024.

[24] C. Jackson et al., “Screening for depression using NLP,” *Interact. J. Med. Res.*, vol. 13, no. 1, e55067, 2024.

[25] D. Thullen et al., “Social media text analysis for mental health prediction,” Tech. Rep., 2024.

[26] A. Al-Eisa et al., “Detection of depression severity using transformers,” *Information*, vol. 16, no. 2, 114, 2025.

[27] G. Coppersmith et al., “CLPsych 2015 shared task,” in *Proc. CLPsych Workshop*, 2015.

[28] C. J. Hutto and E. Gilbert, “VADER: A sentiment analysis model,” *Proc. AAAI ICWSM*, 2014.

[29] D. M. Blei et al., “Latent Dirichlet Allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[30] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” *JMLR*, vol. 12, pp. 2825–2830, 2011.

[31] S. Bird et al., *Natural Language Processing with Python*. O’Reilly, 2009.

[32] C. D. Manning et al., *Introduction to Information Retrieval*. Cambridge Univ. Press, 2008.

[33] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[34] T. Hastie et al., *The Elements of Statistical Learning*, Springer, 2009.

[35] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[36] T. Joachims, “Text categorization with SVMs,” *Proc. ECML*, pp. 137–142, 1998.

[37] A. McCallum and K. Nigam, “Event models for Naïve Bayes text classification,” *AAAI Workshop*, pp. 41–48, 1998.

[38] A. G. Arfi and S. Rahman, “Bangladeshi university student mental-health dataset,” *Kaggle*, 2023.

[39] F. Quincho et al., “Stress and quality of life among university students,” *J. Affect. Disord. Rep.*, 2017.

[40] R. A. Ahmed et al., “Academic pressure and its impact on mental health,” AIUB Undergraduate Report, 2024.

221-35-918

ORIGINALITY REPORT

3%	2%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.coursehero.com Internet Source	1%
2	www.frontiersin.org Internet Source	<1%
3	journal.ugm.ac.id Internet Source	<1%
4	arxiv.org Internet Source	<1%
5	Zhong, Elizabeth. "Optimizing Suicide Prevention in Adolescents: A Longitudinal Approach to Risk Modeling and Resource Allocation.", Harvard University Publication	<1%
6	dokumen.pub Internet Source	<1%
7	L. Santhosh Kumar, Barnashree Khasnobis, Sreedevi Santhosh. "Revitalizing Health Through Humanities: Foregrounding Unheard Trends", Routledge, 2024 Publication	<1%
8	Submitted to Kingston University Student Paper	<1%
9	ijsred.com Internet Source	<1%

Account Clearance