

PHENIKAA UNIVERSITY

FACULTY OF ELECTRICAL AND ELECTRONIC ENGINEERING

NATURAL LANGUAGE PROCESSING COURSE

LABELING GUIDELINE

*TOPIC: DETECTING STRESS AND ANXIETY INDICATORS
USING DEEP LEARNING – BASED MENTAL HEALTH TEXT
CLASSIFICATION.*

STUDENTS:

Full Name: Le Thi Thuy – 22010846

Robotics and Artificial Intelligence Engineering Class

Cohort: 16

Email: 22010846@st.phenikaa-uni.edu.vn

Advisor: ThS. Vu Hoang Dieu

Hanoi, February 27, 2026

1. INTRODUCTION

The purpose of this guide is to establish consistent standards for classifying online confessional texts into three states: Stress, Anxiety, and Normal. The labeling is based on the specific linguistic expressions, emotions, and context of the writer.

2. DEFINING LABELS AND CLASSIFICATION CRITERIA

2.1. Stress Label

Concept: Stress is the physiological and psychological response of the body to demands or pressures that exceed an individual's coping capacity. In this study, the label Stress reflects a state of being overwhelmed, experiencing extreme tension due to external factors or pressures from aspects of life such as work, studies, finances, and social relationships.

Identifying characteristics: Texts in this category often clearly reveal both physical and mental fatigue through vocabulary that emphasizes "compressed" and "exhausted" states such as: "exhausted," "overwhelmed," "terrible pressure," "unable to breathe." The writer often directly refers to specific stressors in reality, such as pressure from work deadlines, intense exam schedules, excessive expectations from family, or overwhelming financial difficulties. A key characteristic of the Stress label is its focus on the difficulties occurring at the present moment. The writer expresses feelings of helplessness and exhaustion, yet remains aware of the root cause of this discomfort.

- Example 1: "The workload is overwhelming; I work until 2 AM every night and still can't finish everything. I feel completely exhausted."
- Example 2: "My family keeps expecting me to succeed, while all I want is to rest. Who understands this pressure?"

2.2. Anxiety Label

Concept: Anxiety is a psychological state characterized by pervasive feelings of worry, fear, and unease, often not tied to a specific object or excessive compared to actual danger. Anxiety reflects a state of excessive unease, fear, or worry about events that haven't happened yet or vague future dangers.

Identifying characteristics: Writers often use phrases describing restlessness and emotional instability, as well as physical symptoms, such as: "anxiety," "restlessness," "fear," "inability to sleep due to overthinking," "rapid heartbeat." Unlike stress, anxiety – related writing tends to be

speculative, exaggerates negativity, and frequently poses hypothetical, self – deprecating questions such as: "Can I do it?", "I feel useless," "I'm afraid of being judged." Anxiety states are often accompanied by fear of social judgment or worst – case scenarios that the writer creates in their mind, leading to insomnia or panic attacks even when the actual triggers haven't materialized.

- Example 1: "Every time I think about having to stand in front of a crowd, my heart races, my hands and feet tremble, and I just want to hide somewhere."
- Example 2: "Lately, I've been having trouble sleeping because as soon as I lie down, my mind conjures up all sorts of terrible scenarios for tomorrow."

2.3. Normal Label

Concept: The term "normal" is understood as psychological stability, without exhibiting clinical signs of anxiety disorders or negative stress. Texts labeled "Normal" include those that do not contain pathological psychological manifestations or excessive negative emotions, maintaining balance in linguistic communication.

Identifying characteristics: Texts in this category primarily share objective information, describing daily life states or positive experiences. The most obvious identifying characteristic is the presence of advice, encouraging messages, healing, or sharing about hobbies, weather, and daily life without accompanying depression or mental deadlock. Even when difficulties are mentioned, the language maintains a neutral, calm, or lighthearted tone, not affecting the individual's normal psychological functioning.

- Example 1: "The weather in Hanoi is beautiful today, I just finished a great book."
- Example 2: "Smile every day to find more meaning in life, everyone."

Table 1: Comparison table of labels

Criteria	Stress	Anxiety	Normal
Root Cause / Origin	Resulting from external factors and tangible realities (Work, family, finances).	Stemming from speculation, irrational fears, or perceived future threats.	No significant negative stressors or clinical indicators.

Core Psychological State	Feeling overwhelmed, trapped, or mentally/physically exhausted.	Feeling insecure, fearful, apprehensive, or self – deprecating.	Balanced, stable, optimistic, or emotionally neutral.
Primary Symptoms	Physical fatigue, loss of energy, and burnout.	Palpitations, tremors, and insomnia due to overthinking.	Relaxed mood, clear cognition, and functional daily life.
Timeline Focus	Present – oriented: Focusing on current burdens and ongoing pressures.	Future – oriented: Worrying about "What if" scenarios and upcoming events.	General/Neutral: Focusing on daily activities or general reflections.
Keywords	Exhausted, overwhelmed, deadline, burnout, suffocating, pressure.	Worried, scared, trembling, sleepless, useless, failure, panic.	Happy, peaceful, today, healing, advice, normal, okay.
Assigned Label (Categorization)	0	1	2

3. Labeling Process and Quality Control

To achieve a high Cohen’s Kappa reliability score, the research team followed these four steps:

- o **Training Phase:** Two labelers performed trial labeling on 100 samples to agree on a common understanding of the Labeling Guideline.
- o **Independent Labeling Phase:** Each labeler independently assigned labels to their own data file containing 291 data points without any exchange, ensuring absolute objectivity.
- o **Cross – check Phase:** The system automatically extracted samples with differing labels. The two labelers discussed to identify the cause of the discrepancy.
- o **Finalization Phase:** In case of disagreement, we re – analyzed the data and made a final decision based on the deepest context of the text.

4. Prioritization Principles and Handling of Ambiguous Data

During implementation, to address complex emotional cases, the study established specific prioritization principles to maintain model consistency. For texts containing both Stress and Anxiety, the Anxiety label was prioritized if fear, self – doubt, or anxiety about the future predominated; conversely, the Stress label was chosen if the content strongly focused on real – world pressures and overwhelming circumstances. For texts using negative language but in a playful, satirical, or social media "meme" style, we assigned the Normal label to accurately reflect the writer's psychological nature.

For highly ambiguous data, such as texts that are too short (e.g., "I'm so tired") or lack a defined context, the core principle was to assign the Normal label to avoid interfering with the model's learning process. In cases where the data contains incomplete sentence structures or uses too much slang, making it difficult to identify the dominant emotion, icons, or stickers, we will remove such patterns to ensure data purity for the training set.

