

RESEARCH ARTICLE

Detection and Analysis of Stress-Related Posts in Reddit's Academic Communities

NAZZERE ORYNGOZHA¹, PAKIZAR SHAMOI¹, AND AYAN IGALI¹

School of Information Technology and Engineering, Kazakh-British Technical University, 050000 Almaty, Kazakhstan

Corresponding author: Pakizar Shamoï (p.shamoï@kbtu.kz)

This work was supported by the Kazakh-British Technical University.

ABSTRACT Nowadays, the significance of monitoring stress levels and recognizing early signs of mental illness cannot be overstated. Automatic stress detection in text can proactively help manage stress and protect mental well-being. In today's digital era, social media platforms reflect various communities' psychological well-being and stress levels. This study focuses on detecting and analyzing stress-related posts in Reddit's academic communities. Due to online education and remote work, these communities have become central for academic discussions and support. We classify text as stressed or not using natural language processing and machine learning classifiers, with Dreddit as our training dataset containing labeled Reddit data. Next, we collect and analyze posts from various academic subreddits. We identified that the most effective individual feature for stress detection is the Bag of Words, paired with the Logistic Regression classifier, achieving a 77.78% accuracy rate and an F1 score of 0.79 on the pre-labeled DReaddit dataset. To validate our model's applicability to detect stress in the specific context of academia, we conducted a supplementary experiment by manually annotating 100 posts from academic subreddits, achieving a 72% accuracy rate. Our key findings reveal that the overall stress level in academic texts is 29%. Posts and comments in professors' Reddit communities are the most stressful compared to other academic levels, including bachelor's, graduate's, and Ph.D. students. This research contributes to our understanding of the stress levels within academic communities. It can help academic institutions and online communities effectively develop measures and interventions to address this issue.

INDEX TERMS Stress detection, Reddit, academia, social media analysis, natural language processing, machine learning, academic stress factors, logistic regression, mental health, emotion detection.

I. INTRODUCTION

In our fast-paced world, stress affects countless individuals, impacting their mental well-being. Among 48% of Gen Z adults report feeling anxious, sad, depressed, experiencing FOMO (fear of missing out), and having lowered self-esteem or insecurity [1], [2], [3]. According to these findings, Gen Z adults' stress levels may rise due to social media-related worries, highlighting the importance of efficient stress detection and support mechanisms in digital platforms.

Stress statistics worldwide indicate the need for effective stress management [4]. Stress permeates various aspects of people's lives, from the workplace to educational institutions and domestic environments. Automatic stress detection has

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan¹.

the potential to address this global health concern, providing support and resources to those in need. Kazakhstan, like many regions around the world, grapples with the multifaceted challenges of mental health and stress. In the context of this Central Asian nation, several factors have been identified as significant contributors to mental health problems, such as education, traditional belief systems, stigma, and shame [5].

One key aspect to consider is that individuals may be unaware of their stress levels or underestimate their symptoms. Stress can often go unnoticed until it reaches a critical stage. Automatic textual stress detection can help bridge this awareness gap by identifying stress markers that might go unnoticed.

Moreover, societal factors like the stigma surrounding mental health and the hesitance to seek professional help create barriers to receiving timely support [6], [7]. While

psychological help is valuable, it can be expensive and inaccessible for many individuals. Automatic stress detection can be an accessible and complementary tool to traditional services, assessing stress levels without direct interaction with specialists. This tool can help individuals overcome hesitations and encourage early intervention by specialists by identifying individuals who require further assistance.

As for the existing psychological tests, they rely heavily on self-reporting [8], which can be influenced by biases or individuals not acknowledging their stress levels. Automatic stress detection of texts in social media offers a data-driven approach, reducing reliance on self-reported data.

In the dynamic landscape of higher education, marked by rigorous academic demands and the pursuit of future goals, stress has become a pervasive issue for both students and pedagogical staff [9], [10], [11]. The other study reveals that students aged 18-19 exhibited more symptoms of common mental disorders than non-students, though the effect size was small [12]. Another study identified that a substantial portion of Ph.D. participants were at high risk for depression [13]. It was found that unhealthy working conditions and increased mental health risks among Ph.D. students. Identified life satisfaction, perceived stress, and negative support as primary predictors of anxiety and depression, and the impact of supervision and the work environment on the mental health and overall well-being of Ph.D. students. Detecting and addressing stress early is essential to prevent it from developing into more severe issues like anxiety or burnout.

Nowadays, people share their stressful experiences and challenges with mental health through online forums, microblogs, and tweets as Internet usage has grown [14]. A 2019 survey conducted by Hill Holiday revealed that 94% of Gen Z adults (individuals born between the mid-1990s and early 2010s) reported having at least one social media account [3]. So, the alarming prevalence of stress worldwide highlights the urgent need for innovative approaches to address this issue.

Reddit is one of the most popular social networks. It consists of thousands of communities called "subreddits" (r/subredditname) that focus on specific interests, hobbies, or themes. Every post strongly follows the topic and rules of the subreddit where it was posted, which makes it easy to analyze the behavior and opinions of the community. Posts on Reddit have a score that is responsible for popularity and it is calculated with the sum of upvotes and downvotes. Reddit was chosen as the data source due to its community's open and supportive nature, allowing longer posts where people readily share their problems, including those related to anxiety [15]. This choice aligns with the importance of studying online platforms like Reddit in the context of well-being and the recognition of the presence of toxicity and anxiety within Reddit communities [16], [17].

The present paper aims to use Natural Language Processing (NLP) and Machine Learning (ML) classification algorithms to automatically detect stress in social media such as Reddit. We further apply the proposed method to analyze

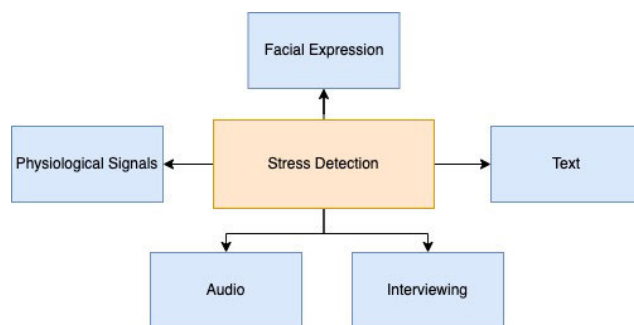


FIGURE 1. Main techniques to detect stress.

the stress levels in texts of professors, bachelor's, graduate, and Ph.D. students. For that, we collected and analyzed a dataset containing 1,584 posts and 122,684 comments on academia-related topics from the Reddit social network.

The contributions of this paper are presented as follows:

- Comparative analysis of ML classifiers for stress detection
- Evaluation of the applicability of the method in the context of academia. We initiated the annotation process to evaluate our method further due to the absence of available academic stress-related datasets labeled by humans.
- Identification of the prevailing emotions associated with stress messages at different levels of academia.
- Understanding stress patterns across various academic levels, including professors, Ph.D. students, graduate students, and bachelor students.
- Use of long-length text that can capture stress indicators.

The rest of the paper is organized as follows. Section I is this Introduction. Section II provides a literature review of research in stress detection. Section III describes Methods, including data preprocessing, feature extraction, and classification. Data collection, description, and annotation procedures are also presented there. Section IV compares the results and presents the most powerful machine-learning method for stress detection. We also present the results of stress evaluation within Reddit's academic communities. Section V compares our results with those of similar studies. Finally, Section VI offers concluding remarks. Additionally, it offers suggestions for how the technique might be enhanced in the future.

The following section provides an overview of stress detection techniques.

II. RELATED WORK

This section provides an overview of stress detection studies in Social Media.

There are different ways of detecting stress, including Physiological Signals [18], [19], [20] (e.g., Heart Rate, Humidity response, Temperature response), Audio [21], [22], Facial Expression [23], Interviewing or Psychological tests [24], [25], [26], and Text-based approaches [27], [28], [29], [30] (see Figure 1).

Many text-based approaches for stress detection were proposed, and some of them identified that the Bidirectional Long Short-Term Memory (BLSTM) with attention mechanism is the most effective approach [28], [31]. The other study focuses on continuous lexical embedding and the BLSTM model [32]. Another study introduces the KC-Net, a mental state knowledge-aware and contrastive network effective for early stress detection [33]. A light and robust method to detect depressive texts in social media was proposed in [34]. The authors used an attention-based bidirectional LSTM and CNN-based model. Some works focus on applying the SVM-based approach [35], achieving the 0.75 accuracy. To increase classification accuracy, some approaches combine lexicon-based features with distributional representations [30]. The other work proposes a TensiStrength method to identify the stress or relaxation level in the text, it applies a lexical approach and a set of rules [36].

Some works concentrate on the analysis of textual data in social media. In the study of Nijhawan et al. [27], the authors employ large Twitter datasets, machine learning techniques, and BERT to classify sentiment. They also use Latent Dirichlet Allocation, an unsupervised ML approach, to find document patterns and forecast pertinent subjects. These algorithms simplify the process of identifying people's feelings on internet platforms.

In the other study, authors look at the use of emotion detection for explainable psychological stress detection [37]. They research multi-task learning and the empathetic modification of language models. Their models, trained by the Dreddit dataset [29], produce results comparable to BERT models. Another work that used the dataset Dreddit is [38]. They used various NLP tools, BERT tokenizers, and a Bag of Words method to prepare textual data for machine learning models. Each method's outcomes are reviewed, with an accuracy of 76%. The study [39] classified the Dreddit data using the VM, LSTM, and BLSTM algorithms, and reported that BLSTM had the best accuracy. However, they did not specify how much accuracy was achieved.

Authors of the related paper [30] discuss the suggested technique for classifying stress, which combines Word2Vec, GloVe, and FastText word embeddings with lexicon-based characteristics. It gets strong classification results with 84.87% F-Score utilizing word embeddings on the Dreddit dataset and 82.57% F-Score using lexicon-based features. While balancing performance and computing resources, their approach shows potential in reliably diagnosing stress from the text. Notably, the technique beats more intricate systems like BERT.

Next, the other study presented Stress Annotated Dataset (SAD) [40], and by developing a model and exploring how it may be used in mental health contexts, they demonstrate how the dataset can be used in real-world situations. The study provides a classification method for everyday stress, a collection of SMS-like phrases that represent stressors, and an assessment of the method's efficiency in categorizing various themes.

While Chauhan et al. [18], [19] concentrate on physiological markers like body temperature, GSR, and blood pressure to detect stress, Raichur et al. [23] use image-processing techniques. Both strategies, however, have limitations since they depend on actual presence. While physiological measures demand specialized equipment and can be intrusive, image processing requires controlled conditions and may not be effective in all circumstances. Researchers are investigating several non-invasive, scalable approaches, such as analyzing textual data, voice, and behavioral clues to solve these constraints. Researchers want to create more adaptable and user-friendly methods for precisely detecting stress levels by extending the spectrum of stress detection techniques beyond a physical presence.

As we can see, stress detection is challenging and worthy of further exploration for several reasons, including subjectivity, context dependency, and multimodal nature. Research on textual stress detection is still limited in using short-length texts for analysis. Short-length text can be enough for sentiment analysis, but when it comes to detecting mental health problems detections, including stress detection it would be proper to use long texts as they are more informative [41].

III. METHODS

In this research paper, we build a system that classifies whether the text is stressful. The schematic representation of the approach is shown in the Figure 2. To implement it, we need to collect data from various sources or datasets and use the NLP techniques such as word embedding, and Bag of Words(BoW). They are intended to create word/sentence data that can be fed to ML models.

A. DATA COLLECTION

We used the following datasets in our study:

1) DReaddit

To train the model, we used Dreddit, a text data collection specifically designed for detecting stress [29]. It is a publicly available dataset that contains a collection of Reddit posts and associated metadata. A total of 190 thousand posts from Reddit are included in the dataset, of which 3553 were labeled manually. The authors offer initially supervised learning algorithms for stress detection that combine neural and traditional methodologies. The data corpus contains stress-indicative posts in 5 domains: abuse (703 posts), anxiety (728 posts), financial (717 posts), PTSD (711 posts), and social (694 posts). Table 1 demonstrates the examples of stressed posts for each domain.

2) TEST DATASETS - REDDIT ACADEMIC COMMUNITIES

For this study, data was collected from the most popular posts based on scores of various subreddits connected with academic topics (“r/Professors”, “r/PhD”, “r/GradSchool”, “r/csMajors”, and “r/EngineeringStudents”) from Reddit

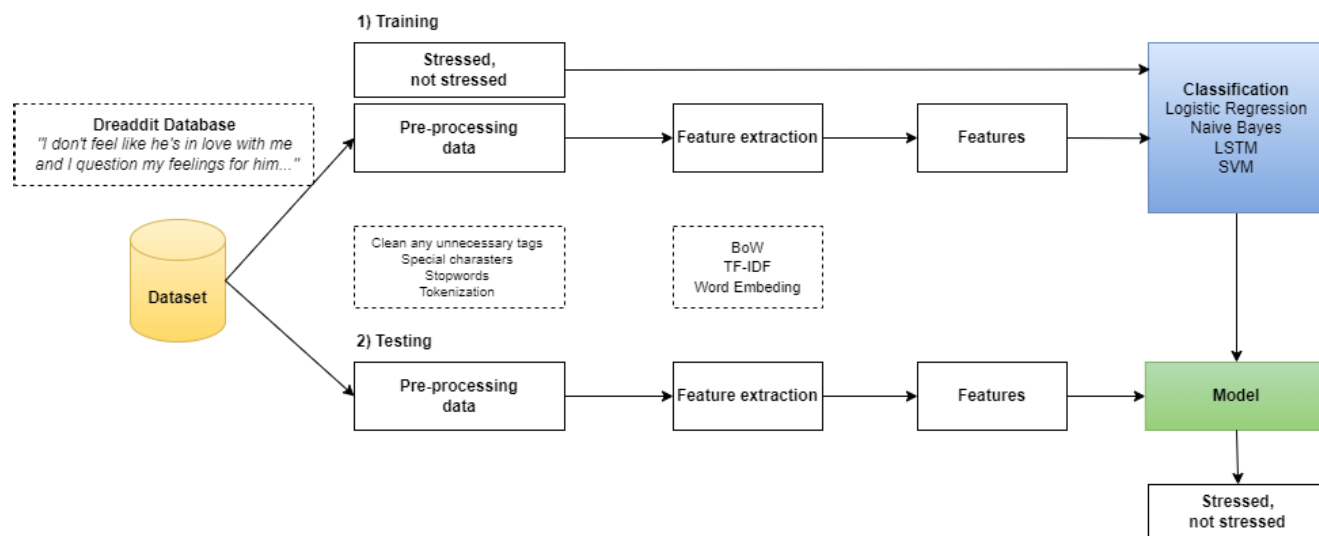


FIGURE 2. Schematic representation of the proposed approach. As a training dataset, we use the Dreddit pre-labeled dataset [29].

TABLE 1. Stressful Dreddit posts with domains.

Post	Stress label	Domain
I've mostly come to terms with it but every time I see his name or face or his girlfriend's it comes back. My question is, when the justice system fails, how do you deal with living near a child molester when you can't just pick up and leave?..	1	abuse
I cried for hours and at one point, something came over me and just slammed my head into my bathroom door. Sadly, since I'm in a dorm, it's a shitty hollow core door and it broke bad and now there is a hole that I have to figure out how to fix. It's a \$100 fine if I can't figure out what to do with it so that's just compounding on the already existing stress. ...	1	anxiety
I asked for nothing but a declaration (a document detailing finances) from the divorce. He stole from me. I asked for nothing but restitution. He fought, forced me to hire a lawyer (more money), dragged out the case, and led to having alimony imposed upon if he failed to make restitution...	1	financial
I cried for hours and at one point, something came over me and just slammed my head into my bathroom door. Sadly, since I'm in a dorm, it's a shitty hollow core door and it broke bad and now there is a hole that I have to figure out how to fix. It's a \$100 fine if I can't figure out what to do with it so that's just compounding on the already existing stress. ...	1	ptsd
That's it. My mom made me delete it and said "that's what is wrong with your generation, you act too grown". Nothing about my picture was "grown" I was actually wearing a long white t-shirt and boyfriend jeans. Even when we go to the beach with my siblings, me and my sister were wearing the exact same bikini and she said something to me about "needing to sit my fast-tailed butt down"...	1	social

using a Python module “PRAW”¹(Python Reddit API Wrapper).

PRAW is a Python library that provides easy access to Reddit’s API. PRAW aspires to be user-friendly and adheres to all of Reddit’s API restrictions. Reddit API has limits, saying it allows 60 requests for one minute.

B. DATA DESCRIPTION

Using Reddit Scrapper we collected 1584 posts and 122684 comments for the posts. The data was collected on June 2, 2023. Each piece of data has information about *Date, Title, Text, Score, Tag*. Reddit is frequently used as a platform for online discussions within various communities or “subreddits”. Table 2 illustrates the distribution of

TABLE 2. Number of posts and comments from each subreddit.

Subreddit	# of posts	# of comments
r/Professors	313	37235
r/PhD	386	22406
r/GradSchool	785	44762
r/csMajors	70	9436
r/EngineeringStudents	30	8845
	1584	122684

posts and comments among subreddits that were scrapped. We only took those posts with “self-text,” the main text beside the title. On Reddit users can make posts with title and image, video and gif, or title and text, it is impossible to combine them. This explains why r/csMajors and r/EngineeringStudents have less data: they tend to post

¹<https://github.com/praw-dev/praw>

TABLE 3. Professors tags.

Tag	# of posts	# of comments
Rants / Vents	40	5968
Humor	18	1892
Advice / Support	6	1320
Academic Integrity	3	296
Teaching / Pedagogy	3	358
Mindblown	1	336
COVID-19	1	69
Other (Editable)	1	28
Where Humor and Rants Collide	1	159
FML	1	83
Technology	1	118
Untagged	237	26608
	313	37235

TABLE 4. Graduates tags.

Tag	# of posts	# of comments
Health & Work/Life Balance	84	5908
Fun & Humour	36	2185
Academics	38	1804
Research	18	998
Professional	8	639
Finance	8	796
Admissions & Applications	9	589
News	4	295
Untagged	580	31548
	785	44762

memes more than graduates or professors and usually do not categorize their posts with tags.

Posts on Reddit can have a flair or tag; the number and topic of tags vary from subreddit to subreddit. They specify the post's topic, and by looking at tags, we can understand what kind of posts are usually popular and get attention. Table 3, Table 4, Table 5, and Table 6 present the distribution of

TABLE 5. PhD tags.

Tag	# of posts	# of comments
Vent	128	9684
Post-PhD	36	1610
Dissertation	75	3197
Other	92	4975
Humor	26	1490
Admissions	3	134
Preliminary Exam	7	260
Need Advice	10	773
Untagged	9	283
	386	22406

TABLE 6. Bachelors tags (csMajors and EngineeringStudents).

Tag	# of posts	# of comments
Shitpost	12	818
Flex	4	296
Others	4	770
Rant	5	972
Company Question	2	91
Advice	3	541
Course Help	1	303
Untagged	69	14490
	100	18281

scraped posts and comments across various tags. Even if not every post was tagged, we can see some tendencies. Posting memes are quite common for all communities ("Humor", "Fun and & Humor", "Shitposting", "Memes"). Another theme common to all communities is asking for advice, which is a general topic on Reddit, as each subreddit can be considered a separate online forum. Professors and Ph.D. students/graduates have the same habit of venting.

Next, Figure 3 presents word clouds of text from posts we scrapped from Reddit communities. Common words for all communities are "think" (13330 overall occurrences from all communities), "time" (19460), "work" (18730) and "people" (16842). In total, the Reddit data encompassed 43,209 distinct words. Figure 3 also shows us that bachelors write a lot about getting a job, companies, interviews, and internships. This can be because our data is from Computer Science and Engineering communities. Bachelors used 16085 unique words.

Graduates and Ph.D. students are different with words like "dissertation", "advisor", "research", "paper", "lab", "academic" or "academia", etc. Graduate and Ph.D. students used 32449 unique words. In addition, Figure 3 illustrates that professors mainly discuss students, emails, classes, teaching, grading/evaluation, parents, and online education. Professors used 25266 unique words. Bachelors and professors have a connection regarding words like "lecture", "class", "professor", and "grade".

C. STRESS DETECTION

1) PRE-PROCESSING

Before applying ML algorithms, posts were preprocessed. During the preprocessing stage, various necessary adjustments were made to prepare the text data for ML classification. The preprocessing steps applied to the dataset are as follows (see Figure 4):

- 1) **Lowercase:** All text data has been converted to lowercase to ensure consistency and uniformity of references.
- 2) **Removal of non-characters:** Removed non-characters, such as letters and special characters, from the text data: removing HTML tags, "@", "_", and other special characters. This step helps eliminate unnecessary noise and information.
- 3) **Tokenization:** The text was tokenized using a tokenizer, which splits the text into individual tokens. This step is necessary to divide the text into basic units.
- 4) **Stopword Removal:** Common English stopwords (e.g., "and", "the", "in") are removed from the tokenized text. Grammatical terms are often discarded because they have no special meaning for classification tasks.
- 5) **Stemming:** The stem process was used to reduce the remaining words to their original form. This helps form similar words and simplifies vocabulary.



FIGURE 3. Word clouds.

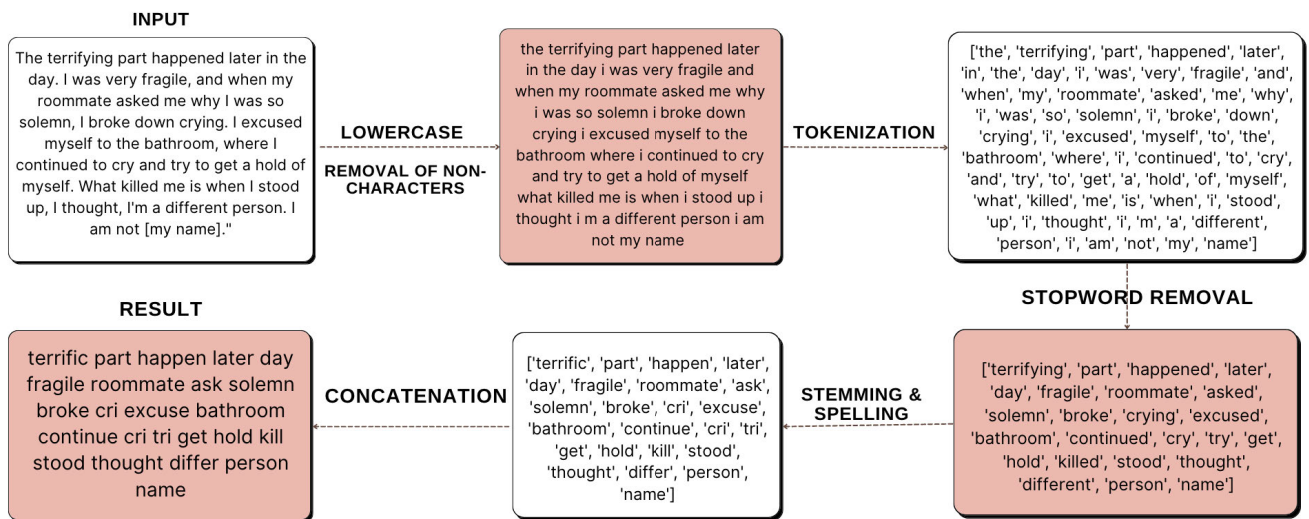


FIGURE 4. Illustration of preprocessing steps.

- 6) **Combining Words:** Finally, the preprocessed words are recombined as coherent text strings, ready to be used in machine learning classification.

Together, these preprocessing steps improve the quality and usability of the texture data, making it more suitable for subsequent ML classification tasks. The difference between raw and processed data can be seen in Fig. 5.

If the data is labeled, it goes to the classification step; otherwise, it goes to the Feature extraction step.

2) FEATURE EXTRACTION

Reducing the dimensionality of the original raw data is achieved by feature extraction, resulting in more manageable groupings that can be processed quickly [42]. Different feature extraction techniques, like BERT embeddings, Word2Vec with TF-IDF weights, bigram TF-IDF, unigram TF-IDF, and Bag of Words, are available. In this work, we use Bag of Words.

▼	Text	▼	Pre-processed	▼
0	He said he had not felt that way before, suggested I go rest		said felt way suggest go rest trigger ahead you hypocond	
1	Hey there r/assistance, Not sure if this is the right place to		hey r assist sure right place post go current student interi	
2	My mom then hit me with the newspaper and it shocked me		mom hit newspaper shock would know like play hit smack	
3	until I met my new boyfriend, he is amazing, he is kind, he i		met new boyfriend ama kind sweet good student like thin	
4	October is Domestic Violence Awareness Month and I am a		actor domes violence war month domes violence survivor	
5	I think he doesn't want to put in the effort for the relationsh		think want put effort relationship work difficult work relat	
6	It was a big company so luckily I didn't have to see him all t		big company lucklll see time act though exist tri talk upda	
7	It cleared up and I was okay but. On Monday I was thinking		clear okay monday think human brain work trip got worry	
8	I actually give an assistant half my emergency fee to come		actual give assist half emerge fee come help sometime fe	
9	I just feel like the street life has fucked my head up. There's		feel like street life fuck head much even know talk anymo	
10	Next week I'll be flying for our family vacation. The flights v		next week fl family vacant flight long msi la lax msi way h	
11	Everything sets me off and I'm almost having a panic attack		every set almost panic attack go dad soon also make anx	
12	I've been taking 12.5 mgs zoloftr for about 6 months (yeah,		take mg nolft month yeah super small dose take ed think	
13	No place in my city has shelter space for us, and I won't pu		place city shelter space us put baby liter street fine nev	
14	These past couple of months have been the worst. My anxii		past couple month worst anxiety gotten bad effect sleep	
15	I asked him three time what happened. And after the third t		ask three time happen third time cri went home month lat	
16	The sensation lasted hours. It seemed like a circulation pro		sest last hour seem like circus problem panic cours end e	
17	It did hurt but didn't leave any marks. When I lied down I ke		hurt last mark lie kept cough today feel kinda construct l	
18	This is helping to get kids in the area tan ability to practice		help get kid area tan bail practice strong group girl look w	
19	We had plans to move in together, a fancy \$1,800/month h		plan move together fancy month home would contribute s	
20	he still holds me back in life. he still finds ways to get to me		still hold back life still find way get yet still talk away fe	

FIGURE 5. After pre-processing.

a: BAG OF WORDS

The principle of BoW is computing the frequency of occurrence for each word while considering the complete text

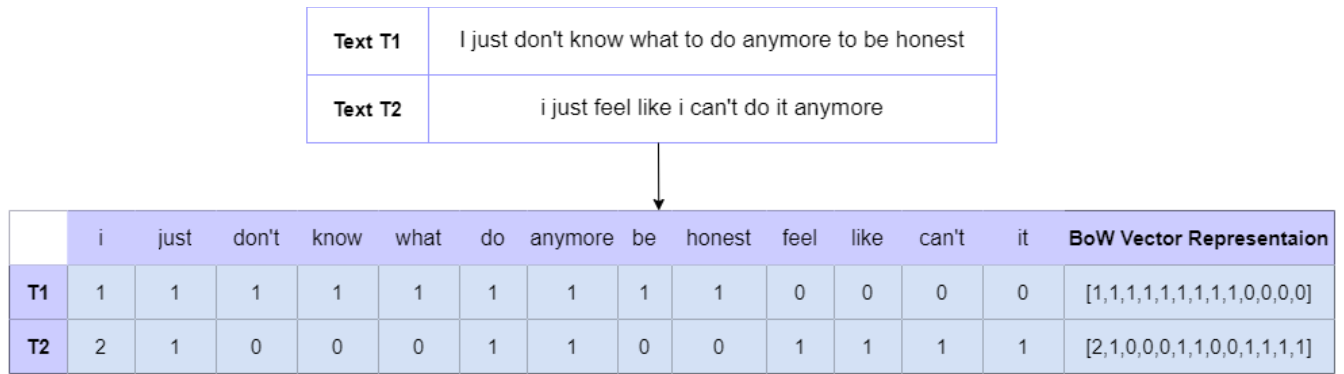


FIGURE 6. Bag-of-words work principle.

data [42]. Figure 6 illustrates the principle of how Bag-of-Words works. Following that, the terms are mapped to their corresponding frequencies. This method ignores word order while accounting for the frequency of each word.

b: TF-IDF

A statistical tool called TF-IDF evaluates a word’s significance inside a document or set of documents. It combines inverse document frequency (IDF), representing a word’s rarity throughout the corpus, with term frequency (TF), which determines how frequently a word appears in a text. To capture their relative significance, TF-IDF gives terms common in a text but uncommon in the corpus larger weights.

c: WORD EMBEDDING

A feature of the learning method called word embedding uses the contextual hierarchy of words to map them to vectors. The feature vectors of comparable words will be the same. For example, the places of the words “dog” and “puppy” are closer, because of their meaning.

3) CLASSIFICATION

To estimate the presence of stress in text, we employ approaches to classification to estimate the likelihood of stress within the users. The following ML algorithms have been used:

a: SVM

An ML approach called the Support Vector Machine (SVM) is generally employed for classification tasks. It establishes a decision boundary that divides the data into several classes to optimize the margin or separation between the decision boundary and the support vectors. Using kernel functions, SVM can handle data that may be separated into linear and non-linear categories. It has been widely used for successful categorization across many fields [20], [41], [43], [44].

b: LOGISTIC REGRESSION

Consider a contextual setting within regression analysis wherein the dependent variable exclusively manifests two

potential outcomes, specifically 0 or 1. This configuration aligns with the framework of binary logistic regression. In contrast, applying multinomial logistic regression becomes pertinent when the dependent variable encompasses a multitude of potential outcomes that surpass the binary domain. These values might be assigned based on qualitative observations. Consider a model structured as follows: $y_i = x_i' \beta + \varepsilon_i$ $x_i' = [1, x_{i1}, x_{i2}, \dots, x_{ik}]$ [45]. Because of its interpretability and simplicity, it is a well-known algorithm that can quickly train and evaluate [20], [41].

The Binary Logistic Regression model estimates the probability that the dependent variable Y equals 1 given input X . It uses the logistic (sigmoid) function to transform a linear combination of input features into a probability value between 0 and 1 as in Equation 1.

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (1)$$

where: $P(Y = 1|X)$ is the likelihood that Y equals 1 given input X . The model coefficients $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are values learned during training, and X_1, X_2, \dots, X_n represent the input features.

The logistic function converts a linear combination of input data into a probability value between 0 and 1. The data point is given to the positive class (1) if the estimated probability is greater than or equal to 0.5; otherwise, it is assigned to the negative class (0).

c: NAIVE BAYES

An efficient probabilistic approach for classification tasks is Naive Bayes. Given the class label, it assumes that the characteristics are conditionally independent, making the probability calculation easier. The effectiveness and scalability of Naive Bayes in handling huge datasets are well established [20], [41], [43].

d: LSTM

Long Short-Term Memory (LSTM) can recognize long-term relationships in sequential data. It is commonly utilized in jobs involving sequential data and natural language processing. To effectively simulate sequential patterns, LSTM

features an internal memory system that enables it to preserve and selectively forget information across extended durations [46], [47].

D. DATASET ANNOTATION ON AN 11-POINT SCALE

We initiated the annotation process to evaluate our method due to the absence of available academic stress-related datasets labeled by humans. The trial dataset comprised 100 posts or comments (25 for each academic level) randomly sampled from our unlabeled stress-related Reddit dataset. 5 annotators, including four regular humans and one psychologist, assessed the stress levels of the posts using an 11-point scale. Each annotator was instructed to score the sentiment from -5 (strong stress) to +5 (unstressed).

It is desired that data annotation is done by someone who has a full understanding of the topic, hence a psychologist (Ph.D. in Psychology) was requested to contribute. She was given double weight because she is more familiar with the context’s terminology. The human raters labeled the stress levels of the posts using separate Google sheets obtained online. Figure 7 shows a screenshot of the form (Google Sheets).

We applied the strategies described in [48]: if 60% or more of annotator labels are considered outliers, the annotator judgments are removed from the job. However, we decreased the threshold to 40% for more accurate resultant annotations.

We use the following formula - Equation 2 to identify if a judgement $A_{i,j}$ is an outlier [48]:

$$|A_{i,j} - avg(A_{i,j})| > std_t(t_j), \tag{2}$$

where i is the index of a particular annotator, and j is the index of a specific tweet/post. The notation $std_t(t_j)$ is the standard deviation of all scores given for a tweet/post t_j .

One annotator, with a 41% outlier rate, had their annotations excluded from the analysis, and we recalculated the weighted average scores for each post.

IV. EXPERIMENTAL RESULTS

A. ACCURACY EVALUATION

Figure 8 illustrates a schematic representation of how the accuracy of the proposed model was evaluated in our study using different datasets. In the initial evaluation phase, the Dreaddit dataset, comprising 3553 human-labeled posts, served as the foundational training set for our model. The accuracy of the model reached 77.78% on this dataset when employing Logistic Regression and the Bag of Words methodology.

Following this, in the academic evaluative phase, we employed a collection of unlabeled posts from an academic context. We performed data annotation on a small dataset involving five annotators (including a psychologist) and applied the model, resulting in an accuracy of 72%. This comprehensive approach allowed us to evaluate the model’s performance thoroughly. The detailed description is provided in the following subsections.

TABLE 7. The output of prediction of the model BoW + logistic regression.

Text	Preprocessed	Expected	Actual
My mom then hit me with the newspaper and it shocked me that she would do this, she knows I don’t like play hitting, smacking, striking, hitting or violence of any sort on my person...	mom hit newspaper shock would know like play hit smack strike hit violence sort person ...send vibe ask universe yesterday decide take friend go help not friend move new place drive friend move strike shoulder	1	0
No place in my city has shelter space for us, and I won’t put my baby on the literal street...	place city shelter space us put baby liter street ...	1	1
but I’m really, really afraid of public embarrassment and awkward situations. So I was in the train station and saw this homeless lady asking for food and money. I always help homeless people if I have a change in my wallet. I walked up to her, took out my wallet, and pulled out this £5 cash to give her.	really really afraid public embarrass awkward stat train station saw homeless laid ask food money away help homeless people chang wallet walk took wallet pull cash give kinda old perfect physics health	0	1

1) DREADDIT (TRAINING DATASET)

The feature extraction method BoW was paired with the SVM, Naive Bayes, Logistic regression, and Word Embeddings with the LSTM algorithm to classify stressed or non-stressed texts.

Table 7 shows the result of a preprocessing tool, in cases when the model made incorrect predictions because of preprocessing. The text must be classified as stressed in the first row, but let us look at the preprocessed data and consider the words separately. If we look at the words “mom”, “hit”, “newspaper”, “shock”, “like”, “play”, it is difficult to understand that in the original text, the mother hits the author with the newspaper, and that’s why the author was shocked. A recap of the third text is that the author helps the homeless lady. However, the words of preprocessed data made the model think that the author may be an anxious person or a homeless who needs food and money.

The models’ accuracies have been calculated and are presented in Table 8. Notably, the Logistic Regression model achieves the highest accuracy at 77.78%.

We use metrics such as Accuracy, Precision, Recall, and F_1 score to evaluate the classification algorithms. Precision estimates how many positively recognized samples are correct, while Recall estimates the fraction of positive samples correctly identified. The higher the F_1 score, the closer both values are.

#	Reddit Post/Comment	Annotation on 11 points scale	Annotator X
1	No accommodations should force overworked graduate students, or even ...	-2	Assign a score to each post from -5 to 5, 11 possible values. -5 - for posts conveying an extreme stress 5 - for posts in which stress is totally absent 0 - for posts in which stressed and unstressed is balanced
2	Yeah i refuse to teach regular undergrads now, just honors and grad students for me	-1	
3	Read a lot of papers, and use something like Zotero to track everything you've read...	5	
4	I graduated a few years ago and I remember how pissed I got when one of...	-3	
5	Congratulations on your victory. You've earned it.	5	
...	
100	My ADHD is very difficult to deal with as it is. Between financial issues and having ...	-4	

FIGURE 7. Sample form (Google Sheets) with Reddit posts or comments for annotation provided to each annotator. Score descriptions and data annotation methods were taken from the following study [48]. The page design was taken from [49]. The form contains 100 posts (25 posts each in 4 categories) from our unlabeled stress dataset.

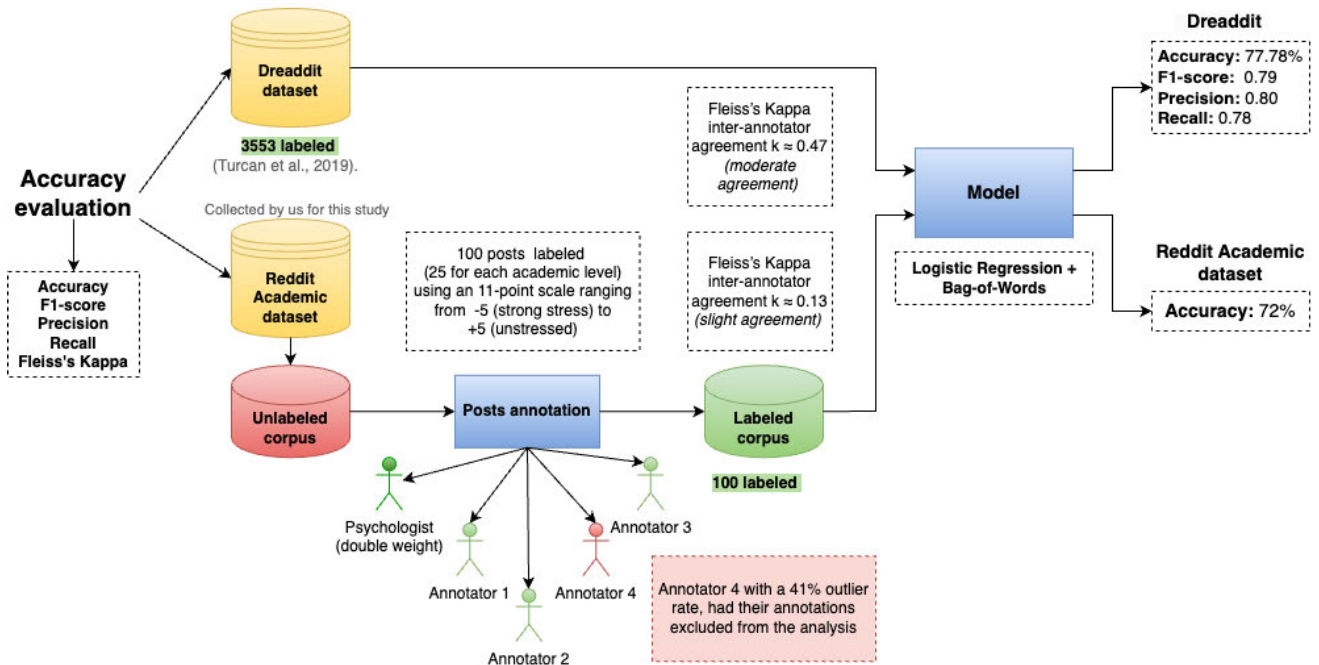


FIGURE 8. Representation of the accuracy evaluation on the Dreaddit dataset [29] and on the Reddit academic dataset.

TABLE 8. Classification results of ML algorithms paired with BoW.

Features	ML	Accuracy,%	Precision	Recall	F score
BoW	SVM	69.90	0.69	0.50	0.58
BoW	Naive Bayes	71.31	0.69	0.61	0.65
BoW	Logistic Regression	77.78	0.80	0.78	0.79
Word Embeddings	LSTM	70.2	71.23	52.10	60.19

a: PRECISION

Correct positive predictions relative to total positive predictions [50]:

$$Precision = \frac{TP}{TP + FP}$$

b: RECALL

Correct positive predictions relative to total actual positives [50]:

$$Recall = \frac{TP}{TP + FN}$$

F₁ score is [50]:

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

c: ACCURACY

Correct predictions to the total prediction [50]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Our results are closely aligned with the findings of the creators of the Dreddit dataset [29]. Their best-performing model achieved an F1-score of 0.798 using Logistic Regression with domain-specific Word2Vec and additional features. Using the Bag of Words approach paired with Logistic Regression, our model achieved a comparable F1 score of 0.79 on the same dataset.

2) REDDIT ACADEMIC DATASET

Let us use the human-annotated dataset discussed earlier to evaluate the performance of the proposed approach.

In the annotation experiment, we classified 100 post texts using the proposed model (BoW + Logistic Regression). The experiment involved five participants, including one psychologist and four outliers. The psychologist’s labels were given higher importance due to their expertise, indicated by a weight of 2. The calculated accuracy of 72% reflects the model’s performance in correctly classifying the texts based on the input from all participants, with a higher emphasis on the psychologist’s judgments. Figure 9 shows the heat map with correlations of ratings provided by annotators. We can see that annotator 3 correlates highly with annotators 1 and 2.

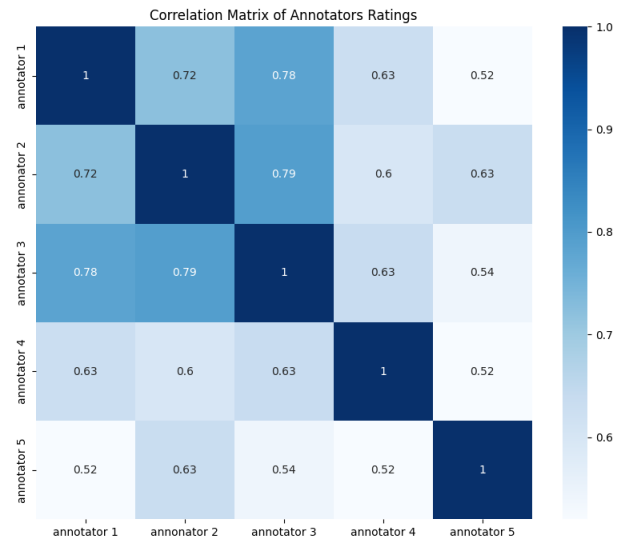


FIGURE 9. The heat map showing correlation of annotators’ rating.

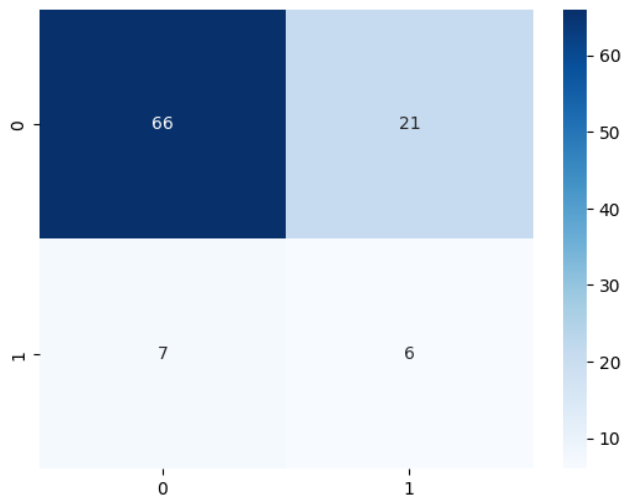


FIGURE 10. Confusion matrix.

The calculated agreement on labeled data is $k \approx 0.13$, using Fleiss’s Kappa inter-annotator agreement [51]. According to the interpretation of Fleiss’s Kappa, the k value shows slight agreement between annotators.

The model’s performance is assessed using a confusion matrix, which provides insights into the model’s ability to classify texts accurately. The confusion matrix for our experiment is as follows in Figure 10.

Our results do not corroborate the findings of [52] on the influence of the sentiment on information diffusion. According to their research, positive messages reach a larger audience, meaning that individuals are more willing to repost positive information, a phenomenon known as positive bias. Figure 11 shows upvotes of posts and comments as a function of the average stress score assigned by the annotators. The graph shows that most viral posts are neutral or slightly stressed, rather than excessively not stressed.

TABLE 9. Mean, median, and standard deviation of upvotes of posts and comments that were classified as “stressed”/“not stressed”.

Dataset	Stressed upvote mean	Stressed upvote median	Stressed upvote std	Not stressed upvote mean	Not stressed upvote median	Not stressed upvote std
Bachelor students	39.5	2.0	255.1	27.0	2.0	163.6
Graduate students	26.7	4.0	97.6	19.9	3.0	71.4
Ph.D. students	20.4	3.0	67.6	13.8	2.0	50.7
Professors	35.6	7.0	120.8	25.7	6.0	84.2
Human-annotated dataset	41.1	9.0	70.5	12.5	2.5	26.0

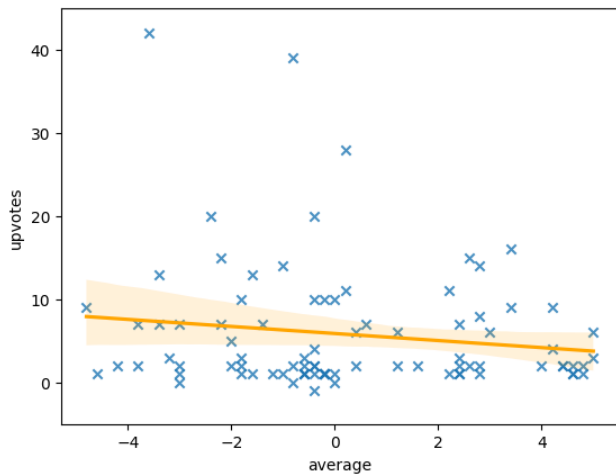


FIGURE 11. Regression plot of upvotes and average of annotators rating.

Table 9 shows us the mean and median upvotes of posts and comments of different academic levels in the human-annotated dataset classified as *stressed/not stressed*. By the phenomenon of positivity bias, it is said that people tend to like positive views on life (not stressful) rather than negative (stressful) [52]. It also implies that people like and share online posts that do not evoke negative emotions. However, after analyzing our datasets, we see that this does not apply to our case. This can possibly be explained by the nature of the Reddit social network.

B. STRESS DETECTION RESULTS – ACADEMIC COMMUNITIES IN REDDIT

For our experiment, we analyzed posts and comments in popular Reddit communities of professors, bachelors, graduates, and Ph.D. students. The most successful model trained on the Dreddit dataset utilizes the BoW representation in combination with Logistic Regression. This model achieved an accuracy rate of 77.78% and F_1 score of 0.79. So, this model is used to make an experiment using the dataset that we parsed from Reddit.

Table 10 presents the stress detection results in considered academic levels: professors, bachelor, graduate, and Ph.D. students.

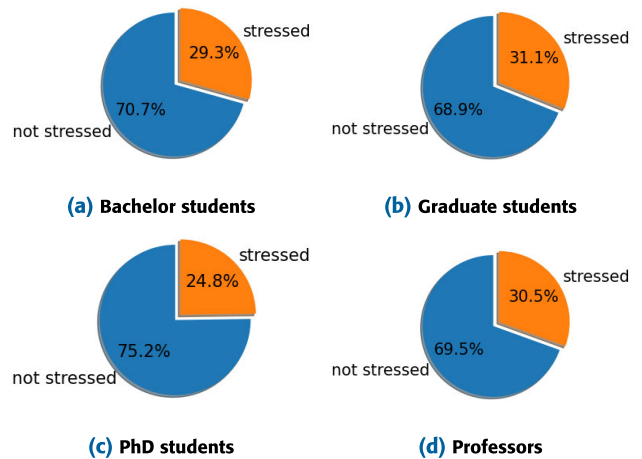


FIGURE 12. Pie chart representing the percentage ratio of stressed/not stressed in posts and comments of different academic levels.

We used NRCLex² Python library [53], [54] to retrieve negative emotional affect(anger, fear, sadness, disgust, surprise) from Reddit data that were classified as *stressed*. NRCLex is based on the National Research Council Canada (NRC) affect lexicon and the NLTK library’s WordNet synonym sets. NRCLex analyzes the text based on the words used and returns the frequency of emotional affect within the text. We filtered emotions with a value of 0.3 or higher to capture the most prevalent ones.

Table 11 represents some comments on Reddit from different levels of academia that were classified as *stressed*. Along with academic level and text, emotion frequency was detected with NRCLex.

Table 12 shows a lexicon of more common terms among academic stressed data. The most common words at all academic levels are “work”, “time”, and “get”. It is noticeable that graduate students and professors use the word “student” more frequently than “people” than bachelor and Ph.D. students. It is also interesting that undergraduate students are more likely to use the word “say” when others are more often using the word “think”.

Words alone don’t tell us much about stressed posts and comments. To get an idea of the key factors contributing to academic stress, we performed topic modeling, and the

²<https://github.com/metalcorebear/NRCLex>

TABLE 10. Stress detection results in various academic levels. The mean stress level is 29%.

Level in Academia	Subreddit name	# of posts	# of comments	Total	Stressed	Stressed, %	Not stressed	Not stressed, %
Bachelor students	r/csMajors, r/EngineeringStudents	100	18281	18381	5389	29.3%	12992	70.7%
Graduate students	r/GradSchool	785	44762	45547	14156	31.1%	31391	68.9%
PhD students	r/PhD	386	22406	22792	5647	24.8%	17145	75.2%
Professors	r/Professors	313	37235	37548	11455	30.5%	26093	69.5%

TABLE 11. Example of post/comments that were classified as “stressed”.

Academic Level	Text	Prevailing emotion
Professors	My grandfather died the day before an exam. I attended the exam in mourning clothes.	sadness: 0.33
PhD	You don’t want to hear this but it’s worse in industry. They tell you this then just fire you :/	fear: 1.0
Graduates	I understand how you feel. It is a very frustrating thing to be trying to tackle grad school and mental health at the same time! :(anger: 0.33
Bachelors	Is it too late to change if I wasted my freshman year (going into sophomore in the fall)? Always felt and still feel too intimidated to try these things.’	sadness: 0.5

TABLE 12. 10 most frequently used words in posts/comments that were classified as “stressed” of different academia levels.

Bachelors	Graduates	PhD	Professors
'work', 1290	'work', 6762	'phd', 2700	'student', 10841
'get', 1174	'school', 5220	'work', 2665	'class', 5076
'people', 1123	'get', 4871	'get', 2280	'work', 3189
'like', 1100	'student', 4835	'like', 2095	'one', 3017
'class', 1036	'time', 4833	'time', 2030	'get', 2979
'engine', 997	'like', 4266	'people', 1520	'would', 2539
'time', 879	'go', 4247	'go', 1472	'time', 2487
'go', 834	'grad', 4126	'think', 1438	'make', 2353
'one', 806	'people', 3704	'know', 1337	'like', 2312
'say', 794	'think', 3677	'one', 1248	'think', 2113

results can be seen in Figure 13. For identifying topics, we used the BERT language representation model [55]. BERT, which stands for Bidirectional Encoder Representations from Transformers, is a comprehensive tool for analyzing short texts with little context and is less time-consuming. We can see what kinds of words are frequently used together, and it gives us a little understanding of topics mostly discussed on Reddit posts and comments that can be considered as *stressed*. The posts and comments about *studying* and *classes*, *professors* and their *English skills*, *major IT companies* and *internships*, and *sleep* resonate the most with stress among bachelor students. *Balancing jobs and education*, *research*, and *defense of a master’s dissertation* probably are stressful for graduate students. *Research and supervisors*, *working*, *mental health* and *therapy*, and *talking to big audiences at conferences* are topics mostly associated with stress among Ph.D. students. Everything connected with *teaching* and *students*, *answering their questions*, *grading them*, *classes*, and *responding to emails* is stressful for professors.

Some sample students’ posts (or comments) classified as stressed and unstressed are presented in Figure 14.

Figure 15 shows the stress change among professors, bachelor, graduate, and Ph.D. students during the year (stress tracking). Posts/comments classified as stressed were categorized monthly, starting in September and ending in August, to match the academic year. The illustrated bar charts show that professors and graduate students tend to be stressed throughout the year. Specifically, there are noticeable spikes in stress for all levels of academia, which coincides with months (December and May) when exams are usually conducted. Bachelor students’ stress levels remain relatively stable throughout the academic year except in October, and stress levels are lower during summer. It may appear that bachelor students experience less stress compared to other academic levels, but this is mostly due to the smaller size of the parsed Reddit dataset used for analysis.

Figures 16 and 17 present line and whisker plots depicting the distribution of identified emotion ratings in stressed Reddit posts/comments from various academic levels. Line plots represent how emotions change throughout the academic year. A clear pattern for all academic levels is a low level of negative emotions in the summer months, which then rises from September until the end of the academic year. As we can see prevailing emotions of stressful posts across all academic levels are sadness and fear. The whisker plot illustrates how emotions are overall distributed and the outliers that appear there.

V. DISCUSSION

Let us consider how the current research findings compare to previous studies. Mental stress has become a significant issue among young people, affecting various aspects of their lives.

With nearly 55 million daily active users and millions of posts publicly broadcasted to a large audience, Reddit is one of the most popular social networking sites [15]. Several works employ machine-learning algorithms to detect stress [56], [57]. Some works focused on the identification of the presence of depression in Reddit texts using ML and NLP techniques [15], [58], Reddit Comment Toxicity Score Prediction using BERT [17].

Our findings support previous research that found stress factors in students: academic, teaching, and learning-related stressors [11]. The same topics were detected in our research and are shown in Figure 13. In contrast, our results presented in Table 9 do not support the phenomenon of positive bias discussed in [52], according to which people tend to share and

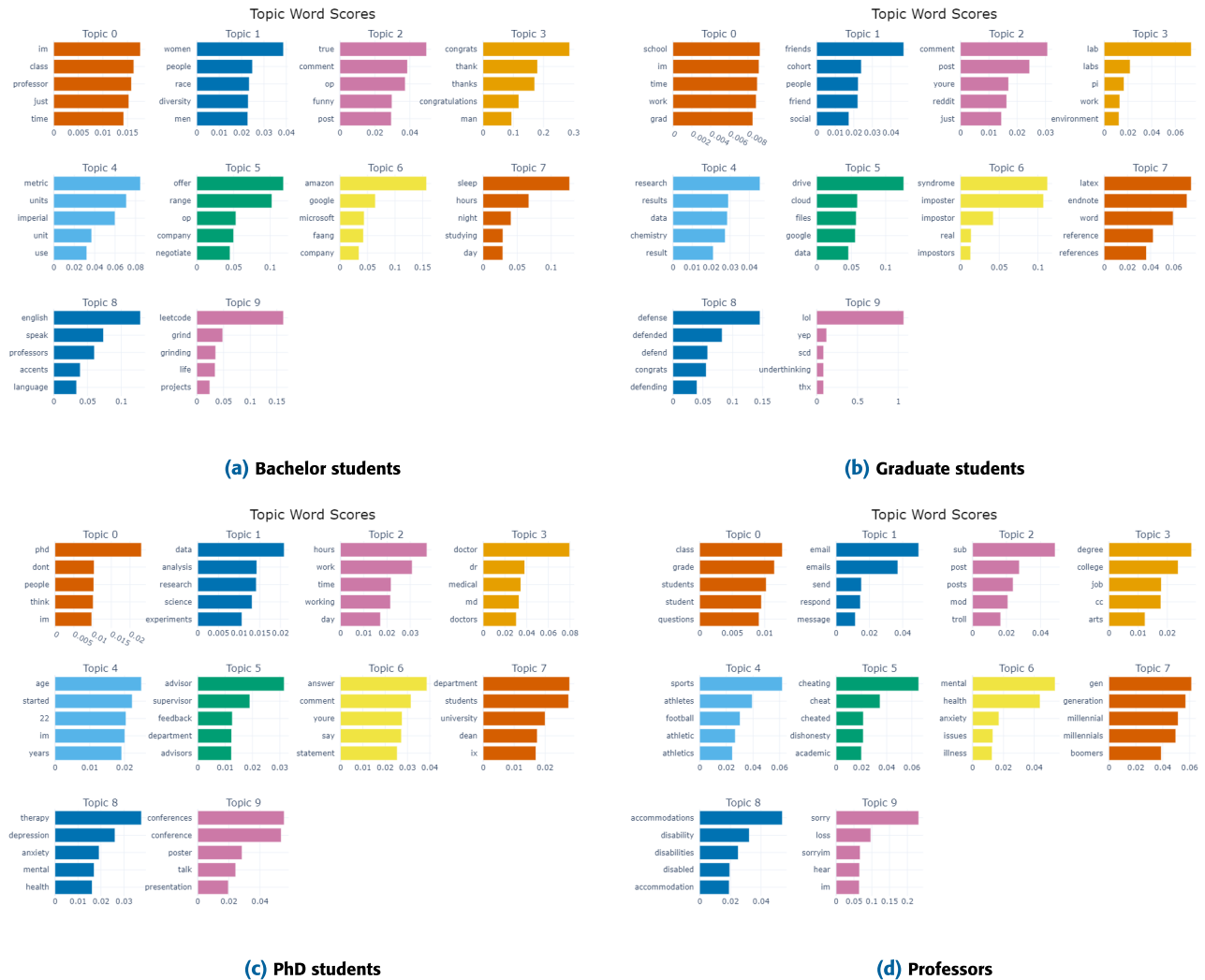


FIGURE 13. Topics with top 5 words using BERT.

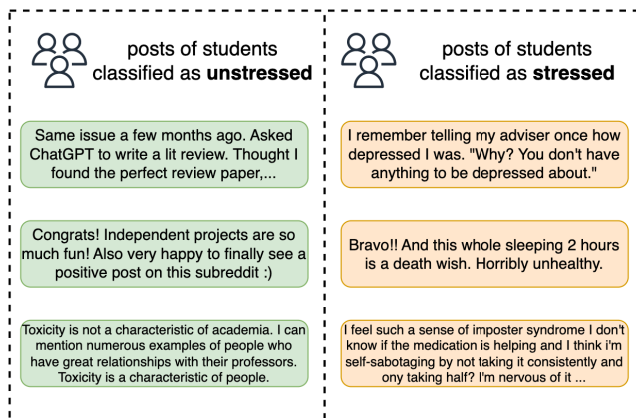


FIGURE 14. Examples of students' posts (or comments) classified as stressed and unstressed.

like news and posts on the Internet that do not evoke negative emotions.

We challenge the findings of [11], asserting that lower-grade students experience more stress than sixth-year medical students. Our research findings indicate that students in higher grades, specifically those pursuing Master's degrees (the same age as medical students in the 5th–6th grades), demonstrate the highest stress levels. This is likely attributed to increased responsibilities, the demand for studying complex topics, completing a Master's dissertation, and the challenges of combining academic work with other commitments.

As we see, previously, similar research had been conducted, but on much broader topics such as mental health, depression, disorders, etc. We narrowed our focus and studied the stress levels in academia and the associated emotions and topics. We also analyze the stress levels of various academic representatives and do not rely on self-reporting (through questionnaires). Studies investigating such broad concepts as stress based on questionnaires are limited [59].

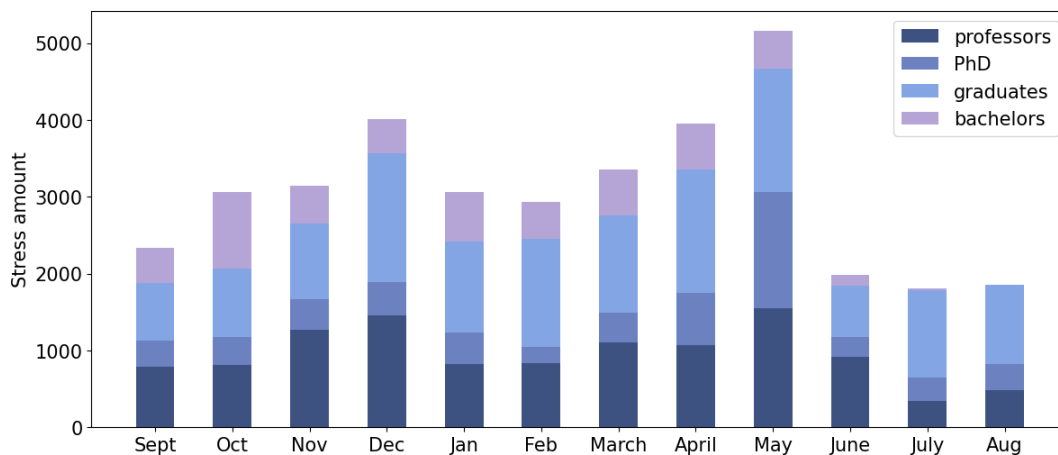


FIGURE 15. Distribution of posts/comments classified as “stressed” by months. The stacked bar charts illustrate the amount of stressed Reddit posts/comments from different levels of academia.

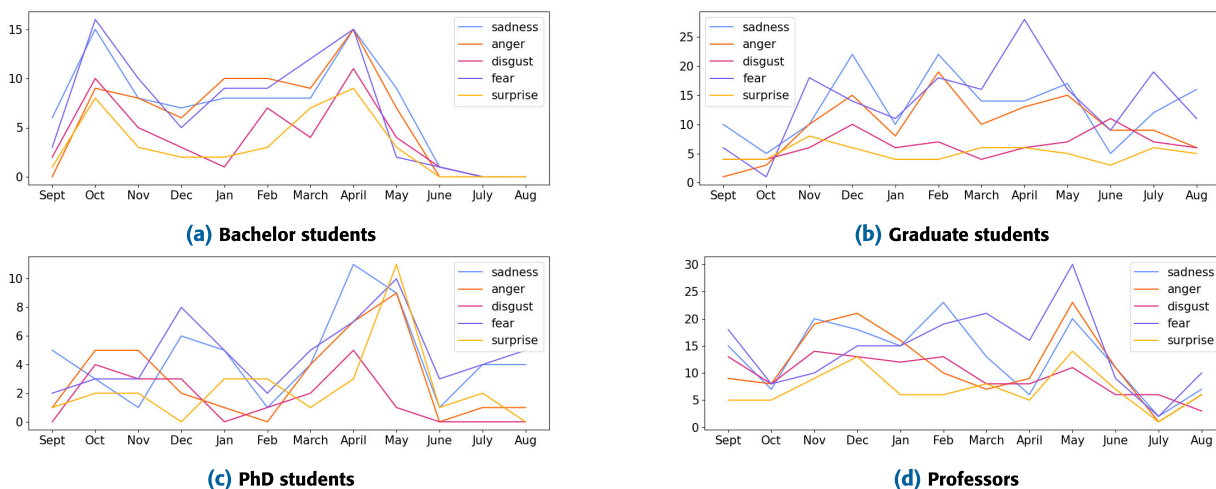


FIGURE 16. Prevailing emotions in Reddit posts/comments classified as “stressed” by months in different levels of academia.

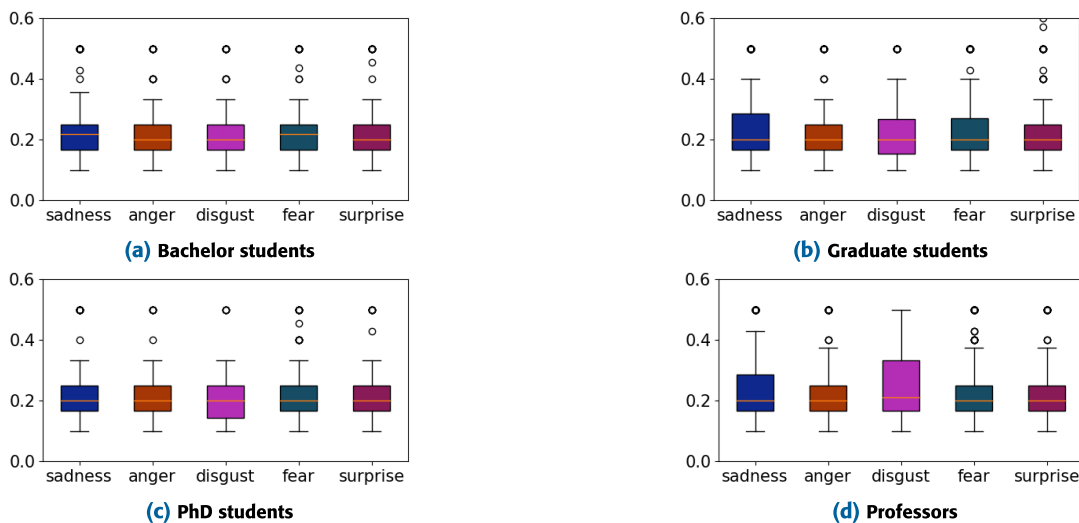


FIGURE 17. Summarising emotions in posts/comments classified as “stressed”. The whisker plots show the distribution of detected emotion values in stressed Reddit posts/comments of different levels of academia.

VI. CONCLUSION

This research uses ML and NLP techniques to recognize and evaluate stress-related posts and comments in Reddit's academic communities.

Automatic stress detection is crucial in mitigating stress risks and improving mental health outcomes. By enabling early detection, addressing hesitancy to seek help, and providing a preliminary assessment, such systems can revolutionize stress management and positively impact individuals' well-being. Stress detection from textual data in social media platforms presents a promising solution, considering its effectiveness and ease of access to data.

We achieved the highest predictive performance on the Dreaddit dataset with Bag of Words and Logistic regression, with an accuracy of 77.78% and an F1 score of 0.79. To validate our model's applicability in academia, we conducted a supplementary experiment by manually annotating 100 posts from academic subreddits, achieving a 72% accuracy rate.

Our findings suggest that the overall stress level in academic texts is 29%, with the key emotions being sadness and fear. Moreover, the key factors contributing to this stress are:

- For bachelor students, these include studying classes, professors and their English skills, major IT companies, internships, and sleep.
- For graduate students - balancing job and education, research, and defense of a master's dissertation.
- For PhD students, it is research and their supervisors, mental health and therapy, and talking to big audiences at conferences.
- For professors - students, and teachers, responding to questions and emails.

Our findings can assist universities in building better-targeted interventions and support systems by identifying and quantifying stress levels at different academic levels early. This tailored approach to managing academic stress-related issues can improve students' and faculty's mental health and well-being. Administrators and educators can use the current study's findings to adopt policies and practices that reduce stress and foster a better learning and working environment.

The study has some limitations. Estimates of academic stress levels were based solely on one social network, Reddit. The other limitation is the relatively small dataset size. These are issues that need to be addressed in future work. We also plan to experiment with other Machine Learning techniques (Bert, Random Forest, Decision Tree, XGBoost) and Computing with Words technique [60], [61] to increase the accuracy.

The majority of approaches for stress identification focus on physical aspects only or require a psychological test, which can lead to inaccurate results. Stress can manifest through various modalities, including text, speech, and physiological signals. Integrating these diverse data sources to improve detection accuracy is a complex but important challenge we plan to explore.

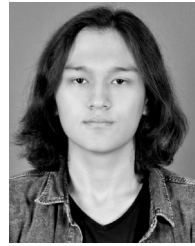
REFERENCES

- [1] M. Milyavskaya, M. Saffran, N. Hope, and R. Koestner, "Fear of missing out: Prevalence, dynamics, and consequences of experiencing FOMO," *Motivat. Emotion*, vol. 42, no. 5, pp. 725–737, Oct. 2018.
- [2] L. Tomczyk and E. Selmanagic-Lizde, "Fear of missing out (FOMO) among youth in Bosnia and Herzegovina—Scale and selected mechanisms," *Children Youth Services Rev.*, vol. 88, pp. 541–549, May 2018.
- [3] (2019). *Gen Z Report Key*. [Online]. Available: <https://brand-news.it/wp-content/uploads/2019/10/2019-Gen-Z-Report.pdf>
- [4] S. Mahmud, S. Hossain, A. Mueyed, M. M. Islam, and M. Mohsin, "The global prevalence of depression, anxiety, stress, and insomnia and its changes among health professionals during COVID-19 pandemic: A rapid systematic review and meta-analysis," *Heliyon*, vol. 7, no. 7, Jul. 2021, Art. no. e07393. <https://www.sciencedirect.com/science/article/pii/S2405844021014961>
- [5] M. H. Akhand, "Factors contributing to mental health problems in Kazakhstan: Literature review," B.S. thesis, Dept. Appl. Sci., Turku Univ., Turku, Finland, 2019.
- [6] P. W. Corrigan and A. C. Watson, "Understanding the impact of stigma on people with mental illness," *Proc. Comput. Sci.*, vol. 1, pp. 16–20, 2022.
- [7] H. J. Lim, L. Moxham, C. Patterson, D. Perlman, V. Lopez, and Y. S. Goh, "Students' mental health clinical placements, clinical confidence and stigma surrounding mental illness: A correlational study," *Nurse Educ. Today*, vol. 84, Jan. 2020, Art. no. 104219.
- [8] K. Masood, B. Ahmed, J. Choi, and R. Gutierrez-Osuna, "Consistency and validity of self-reporting scores in stress measurement surveys," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 4895–4898, doi: 10.1109/EMBC.2012.6347091.
- [9] T. W. Wong, Y. Gao, and W. W. S. Tam, "Anxiety among university students during the SARS epidemic in Hong Kong," *Stress Health*, vol. 23, no. 1, pp. 31–35, Feb. 2007.
- [10] K. Shamsuddin, F. Fadzil, W. S. W. Ismail, S. A. Shah, K. Omar, N. A. Muhammad, A. Jaffar, A. Ismail, and R. Mahadevan, "Correlates of depression, anxiety and stress among Malaysian University students," *Asian J. Psychiatry*, vol. 6, no. 4, pp. 318–323, Aug. 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1876201813000592>
- [11] M. M. Al-Shahrani, B. S. Alasmri, R. M. Al-Shahrani, N. M. Al-Moalwi, A. A. Al Qahtani, and A. F. Siddiqui, "The prevalence and associated factors of academic stress among medical students of King Khalid University: An analytical cross-sectional study," *Healthcare*, vol. 11, no. 14, p. 2029, Jul. 2023.
- [12] T. McCloud, S. Kamenov, C. Callender, G. Lewis, and G. Lewis, "The association between higher education attendance and common mental health problems among young people in England: Evidence from two population-based cohorts," *Lancet Public Health*, vol. 8, no. 10, pp. e811–e819, Oct. 2023.
- [13] J. Friedrich, A. Bareis, M. Bross, Z. Bürger, Á. Cortés Rodríguez, N. Effenberger, M. Kleinhansl, F. Kremer, and C. Schröder, "How is your thesis going? Ph.D. students' perspectives on mental health and stress in academia," *PLoS ONE*, vol. 18, Jul. 2023, Art. no. e0288103.
- [14] (2023). *Digital 2023: Global Overview Report*. [Online]. Available: <https://datareportal.com/reports/digital-2023-global-overview-report>
- [15] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in Reddit social media forum," *IEEE Access*, vol. 7, pp. 44883–44893, 2019.
- [16] H. Almerakhi, H. Kwak, and B. J. Jansen, "Investigating toxicity changes of cross-community redditors from 2 billion posts and comments," *PeerJ Comput. Sci.*, vol. 8, p. e1059, Aug. 2022, doi: 10.7717/peerj-cs.1059.
- [17] R. Shounak, S. Roy, V. Kumar, and V. Tiwari, "Reddit comment toxicity score prediction through BERT via transformer based architecture," in *Proc. IEEE 13th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2022, pp. 0353–0358.
- [18] M. Chauhan, S. V. Vora, and D. Dabhi, "Effective stress detection using physiological parameters," in *Proc. International Conference on Innovations in Information, Embedded and Communication Systems*, vol. 2017, pp. 1–6.
- [19] G. Shanmugasundaram, S. Yazhini, E. Hemapratha, and S. Nithya, "A comprehensive review on stress detection techniques," in *Proc. IEEE Int. Conf. Syst., Comput., Autom. Netw. (ICSCAN)*, Mar. 2019, pp. 1–6.
- [20] S. Gedam and S. Paul, "A review on mental stress detection using wearable sensors and machine learning techniques," *IEEE Access*, vol. 9, pp. 84045–84066, 2021.

- [21] A. Triantafyllopoulos, S. Zänkert, A. Baird, J. Konzok, B. M. Kudielka, and B. W. Schuller, "Insights on modelling physiological, appraisal, and affective indicators of stress using audio features," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 2619–2622.
- [22] P. Chyan, A. Achmad, I. Nurtanio, and I. S. Areni, "A deep learning approach for stress detection through speech with audio feature analysis," in *Proc. 6th Int. Conf. Inf. Technol., Inf. Syst. Electr. Eng. (ICITISEE)*, Dec. 2022, pp. 1–5.
- [23] N. Raichur, N. Lonakadi, and P. Mural, "Detection of stress using image processing and machine learning techniques," *Int. J. Eng. Technol.*, vol. 9, no. 3S, pp. 1–8, Jul. 2017.
- [24] E.-H. Lee, "Review of the psychometric evidence of the perceived stress scale," *Asian Nursing Res.*, vol. 6, no. 4, pp. 121–127, Dec. 2012.
- [25] G. Singh, D. Kaur, and H. Kaur, "Presumptive stressful life events scale (PSLES)—A new stressful life events scale for use in India," *Indian J. Psychiatry*, vol. 26, pp. 14–107, Apr. 1984.
- [26] A. D. Kanner, J. C. Coyne, C. Schaefer, and R. S. Lazarus, "Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events," *J. Behav. Med.*, vol. 4, no. 1, pp. 1–39, Mar. 1981.
- [27] T. Nijhawan, G. Attigeri, and T. Ananthkrishna, "Stress detection using natural language processing and machine learning over social interactions," *J. Big Data*, vol. 9, no. 1, pp. 1–24, Dec. 2022.
- [28] S. Jadhav, A. Machale, P. Mharnur, P. Munot, and S. Math, "Text based stress detection techniques analysis using social media," in *Proc. 5th Int. Conf. Comput., Commun., Control Autom. (ICCUBEA)*, Sep. 2019, pp. 1–5.
- [29] E. Turcan and K. McKeown, "Dreaddit: A Reddit dataset for stress analysis in social media," in *Proc. 10th Int. Workshop Health Text Mining Inf. Anal. (LOUHI)*, 2019, pp. 97–107. [Online]. Available: <https://aclanthology.org/D19-6213>
- [30] S. Muñoz and C. A. Iglesias, "A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations," *Inf. Process. Manage.*, vol. 59, no. 5, Sep. 2022, Art. no. 103011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457322001212>
- [31] D. William and D. Suhartono, "Text-based depression detection on social media posts: A systematic literature review," *Proc. Comput. Sci.*, vol. 179, pp. 582–589, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921000491>
- [32] Y. Zheng, Y. Li, Z. Wen, B. Liu, and J. Tao, "Text-based sentential stress prediction using continuous lexical embedding for Mandarin speech synthesis," in *Proc. 10th Int. Symp. Chin. Spoken Lang. Process. (ISCSLP)*, Oct. 2016, pp. 1–5.
- [33] K. Yang, T. Zhang, and S. Ananiadou, "A mental state knowledge-aware and contrastive network for early stress and depression detection on social media," *Inf. Process. Manage.*, vol. 59, no. 4, Jul. 2022, Art. no. 102961. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457322000796>
- [34] T. Ghosh, M. H. A. Banna, M. J. A. Nahian, M. N. Uddin, M. S. Kaiser, and M. Mahmud, "An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417422020255>
- [35] M. R. Febriansyah, Nicholas, R. Yunanda, and D. Suhartono, "Stress detection system for social media users," *Proc. Comput. Sci.*, vol. 216, pp. 672–681, Jan. 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S187705092202261X>
- [36] M. Thelwall, "TensiStrength: Stress and relaxation magnitude detection for social media texts," *Inf. Process. Manage.*, vol. 53, no. 1, pp. 106–121, Jan. 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457316302321>
- [37] E. Turcan, S. Muresan, and K. McKeown, "Emotion-infused models for explainable psychological stress detection," in *Proc. Conf. North Amer. Chapter ACL*, 2021, pp. 2895–2909.
- [38] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, "Machine learning driven mental stress detection on Reddit posts using natural language processing," *Human-Centric Intell. Syst.*, vol. 3, no. 2, pp. 80–91, Mar. 2023.
- [39] S. Jadhav, A. Machale, P. Mharnur, P. Munot, and S. Math, "Text based stress detection techniques analysis using social media," 2019, pp. 1–5, doi: [10.1109/ICCUBEA47591.2019.9129201](https://doi.org/10.1109/ICCUBEA47591.2019.9129201).
- [40] M. L. Mauriello, T. Lincoln, G. Hon, D. Simon, D. Jurafsky, and P. Paredes, "SAD: A stress annotated dataset for recognizing everyday stressors in SMS-like conversational systems," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, May 2021, pp. 1–7.
- [41] R. A. Rahman, K. Omar, S. A. Mohd Noah, M. S. N. M. Danuri, and M. A. Al-Garadi, "Application of machine learning methods in mental health detection: A systematic review," *IEEE Access*, vol. 8, pp. 183952–183964, 2020.
- [42] (2022). *Why Do We Need Feature Extraction?* [Online]. Available: <https://www.analyticsvidhya.com/blog/2022/01/nlp-tutorials-part-ii-feature-extraction/>
- [43] Y. Ding, X. Chen, Q. Fu, and S. Zhong, "A depression recognition method for college students using deep integrated support vector algorithm," *IEEE Access*, vol. 8, pp. 75616–75629, 2020.
- [44] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2020, pp. 51–57.
- [45] D. Montgomery, E. Peck, and G. Vining, *Introduction To Linear Regression Analysis* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2013.
- [46] S. Ghosh and T. Anwar, "Depression intensity estimation via social media: A deep learning approach," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 6, pp. 1465–1474, Dec. 2021.
- [47] M. Awais, M. Raza, N. Singh, K. Bashir, U. Manzoor, S. U. Islam, and J. J. P. C. Rodrigues, "LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16863–16871, Dec. 2021.
- [48] A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes, "SemEval-2015 task 11: Sentiment analysis of figurative language in Twitter," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, 2015, pp. 470–478.
- [49] E. Shamo, A. Turdybay, P. Shamo, I. Akhmetov, A. Jaxylykova, and A. Pak, "Sentiment analysis of vegan related tweets using mutual information for feature selection," *PeerJ Comput. Sci.*, vol. 8, p. e1149, Dec. 2022, doi: [10.7717/peerj-cs.1149](https://doi.org/10.7717/peerj-cs.1149).
- [50] Statology. (2021). *F1 Score vs. Accuracy: Which Should You Use?* [Online]. Available: <https://www.statology.org/f1-score-vs-accuracy/>
- [51] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, Nov. 1971.
- [52] E. Ferrara and Z. Yang, "Quantifying the effect of sentiment on information diffusion in social media," *PeerJ Comput. Sci.*, vol. 1, p. e26, Sep. 2015, doi: [10.7717/peerj-cs.26](https://doi.org/10.7717/peerj-cs.26).
- [53] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, 2013.
- [54] S. Mohammad and P. Turney, "Emotions evoked by common words and phrases: Using mechanical Turk to create an emotion lexicon," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*. Los Angeles, CA, USA: Association for Computational Linguistics, Jun. 2010, pp. 26–34. [Online]. Available: <https://aclanthology.org/W10-0204>
- [55] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [56] R. Ahuja and A. Banga, "Mental stress detection in university students using machine learning algorithms," in *Proc. Comput. Sci.*, vol. 152, Jan. 2019, pp. 349–353. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050919306581>
- [57] M. Gupta and S. Vaikole, "Recognition of human mental stress using machine learning paradigms," in *Proc. 3rd Int. Conf. Adv. Sci. Technol. (ICAST)*, Apr. 2020, pp. 1–8. [Online]. Available: <https://ssrn.com/abstract=3571754>
- [58] L. D'Cruz, V. Dubey, and P. Thakur, "Depression prediction from combined Reddit and Twitter data using machine learning," in *Proc. 2nd Int. Conf. Innov. Technol. (INOCON)*, Mar. 2023, pp. 1–5.
- [59] G. Barbayannis, M. Bandari, X. Zheng, H. Baquerizo, K. W. Pecor, and X. Ming, "Academic stress and mental well-being in college students: Correlations, affected groups, and COVID-19," *Frontiers Psychol.*, vol. 13, May 2022, Art. no. 886344, doi: [10.3389/fpsyg.2022.886344](https://doi.org/10.3389/fpsyg.2022.886344).
- [60] P. Shamo, A. Inoue, "Computing with words for direct marketing support system," in *Proc. 23rd Midwest Artif. Intell. Cogn. Sci. Conf. (MAICS)*, 2012, pp. 19–26.
- [61] A. Kali, P. Shamo, Y. Zhangbyrbayev, and A. Zhandaulet, "Computing with words for industrial applications," in *Intelligent Systems and Applications*. Cham, Switzerland: Springer, Sep. 2022, pp. 257–271, doi: [10.1007/978-3-031-16075-2_17](https://doi.org/10.1007/978-3-031-16075-2_17).



NAZZERE ORYNGOZHA received the B.S. degree in systems of information security from International Information Technology University, Almaty, Kazakhstan, in 2022. She is currently pursuing the M.S. degree in software engineering with the School of Information Technology and Engineering, Kazakh-British Technical University, Almaty. Since 2022, she has been a Tutor with the Cybersecurity Department, International Information Technology University. She is also a Backend Developer with Python.



AYAN IGALI is currently pursuing the bachelor's degree in information systems with the School of Information Technology and Engineering, Kazakh-British Technical University, Almaty, Kazakhstan. His research interests include machine learning, NLP, emotion detection, and human-friendly systems.

...



PAKIZAR SHAMOI received the B.S. and M.S. degrees in information systems from Kazakh-British Technical University, Almaty, Kazakhstan, in 2011 and 2013, respectively, and the Ph.D. degree in engineering from Mie University, Tsu, Japan, in 2019. In the academic journey, she has held various teaching and research positions with Kazakh-British Technical University, where she has been a Professor with the School of Information Technology and

Engineering, since August 2020. Her commitment to education and contributions to the field have been recognized with the prestigious Best University Teacher Award in Kazakhstan, in 2022. She is the author of one book and more than 28 scientific publications. Her research interests include artificial intelligence and machine learning, focusing on fuzzy sets and logic, soft computing, representing and processing colors in computer systems, natural language processing, computational aesthetics, and human-friendly computing and systems. Awards for the best paper at conferences were received four times. She took part in the organization and worked in the org. committee (as the Head of the Session and Responsible for Special Sessions) of several international conferences, such as IFSA-SCIS 2017, Otsu, Japan; SCIS-ISIS 2022, Mie, Japan; and EUSPN 2023, Almaty, Kazakhstan. She served as a Reviewer at several international conferences, including IEEE: SMC 2022, SCIS-ISIS 2022, SMC 2020, ICIEV-IVPR 2019, and ICIEV-IVPR 2018.