

International Conference on Machine Learning and Data Engineering

Detection and classification of anxiety in university students through the application of machine learning

Shaurya Bhatnagar^a, Jyoti Agarwal^{b*}, Ojasvi Rajeev Sharma^{a,b}^aDept of CSE, Amity University Uttar Pradesh, Sector 125, Noida, Uttar Pradesh 201313, India^bDept of CSE, Graphic Era Deemed to be University, Dehradun, Uttarakhand 248002, India^{a,b}Dept of CSE, Amity University Uttar Pradesh, Sector 125, Noida, Uttar Pradesh 201313, India

Abstract

Mental Health has recently transformed into a domain that caused interest in almost every field and has garnered attention in recent years, millions of people suffer due to mental illnesses that are often left unaccounted for, a vast majority of these people are youngsters, a large chunk being students belonging to various universities. This research's aim was to pinpoint the extent of anxiety, in conjunction, its effects, as noticed in Indian university students. The dataset was gathered using a questionnaire, that was matching with Likert scale measurement criteria, which consisted of university engineering students. This questionnaire was distributed amongst 127 engineering students, as a result, the level of anxiety was quantified, and its causes and effects were identified, a series of statistical reliability and validity tests were performed on the dataset. Machine learning algorithms are applied in the end to classify the anxiety level based on the effects of anxiety after being trained on pre-existing data points. It was found that the Cronbach's alpha value for the entire dataset was 0.723 and Pearson's correlation coefficient was 0.823, the accuracy for the naïve bayes, decision tree, random forest and support vector machine algorithms were 71.05%, 71.05%, 78.9% and 75.5% respectively.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering

Keywords: Anxiety; classification algorithms; dataset; machine learning; mental health; statistical analysis

1. Introduction

Anxiety happens to be a disorder which causes worry and nervousness so much that it begins to interfere with one's daily life. There can be several contributing factors that may trigger an individual and make them become anxious for prolonged times, which can leave behind crippling and devastating after-effects, for e.g., feeling worthless, self-hatred, hopeless etc. University students constitute a large portion of the sufferers, bearing numerous adverse effects caused by crippling anxiety. There exist several questionnaires that are specifically tailored to conclude whether an individual suffers from anxiety and its effects, such as the DASS-21 (Depression, Anxiety and Stress Scale), GAD-7 (Generalized Anxiety Disorder). The causes and effects of anxiety vary from person to person, however certain circumstances can generalize the causes as well as the effects of anxiety, this generalization can be attributed to several university students in India that is subject to copious amounts of stress, this may be academic, social, or intrinsic stress. This paper has dived into the possibility of pinpointing the plausible and common causes of anxiety, as well as linking said causes to certain effects that may surface in an average Indian university student. A self-created question form was used to gather the dataset for this purpose, following which the dataset gathered was subjected to certain statistical analyses to determine its reliability and validity, finally machine learning classification algorithms were applied on the dataset to classify new data points into mild, moderate, or severe categories in relevance to anxiety levels calculated using the identified causes. It has been noticed that there has been a significant rise in suicides in India especially during the past couple of years [1]. A study also found that 37 percent of University's students in India suffer from moderate depression [2]. This can be correlated with severe pressure that university students are often labouring under. Several factors contribute towards the origin of this stress and are often disregarded assuming that the student will simply get over it, however, this ordeal may render an individual crippled and unmotivated to extreme degrees. The dataset collection included 127 engineering students that participated in this research, the questionnaire for this purpose targeted numerous factors that is relatable to the causes and effects of anxiety that may present itself in an average Indian university student. The students that participated in this research belong to an engineering background, however the questionnaire designed can be used with students belonging to any study field at any respective University local to a certain region. The Questionnaire consists of 40 questions that is divided into following categories:

- Demographic
- Probable Causes of Anxiety
- Probable Effects of Anxiety

1.1 Understanding Anxiety

Anxiety can be described as a feeling composed of worrisome thoughts, tension, and physical manifestations in the form of high blood pressure and several other effects. There can be several reasons behind the feeling anxious, it is the human body's natural response to stressful situations, such as starting a new job, or one's first day at college/university [3]. It is essential to understand the difference between ordinary anxiety that manifests as a natural response, and anxiety disorders that can adversely affect an individual's life [4]. Ordinary anxiety is temporary, it is sporadic and intermittent, usually kicking in when an individual is placed under stressful circumstances, for example, delivering a presentation before an audience, or taking a huge test, the feeling of anxiety that accompanies these situations usually subsides, and the individual returns to their original state of mind. However, anxiety disorders are perpetual. They can severely incapacitate an individual, to an extent of rendering them incapable of carrying out everyday tasks. There are several categories of anxiety disorders, as listed below [4]:

Panic Disorder: This causes an individual to experience a panic attack at any point in time, this may be triggered by a situation and usually leaves the person incapacitated.

Phobia: This is characterized by an intense fear of a certain situation or object, for example, claustrophobia is the fear of being stuck in tight spaces, such as a small elevator.

- Social Anxiety Disorder: Characterized by the fear of social interactions, usually owing to the fear of judgement by others.
- Obsessive Compulsive Disorder: Characterized by an individual compulsively and repetitively engaging in a particular set of activities due to their intrinsic fears.
- Separation anxiety: Characterized by the fear of having to be away from loved ones.

- Hypochondria: Being perpetually fearful of one's health.
- PTSD (Post-traumatic stress disorder): Characterized by a feeling of anxiety that follows an extremely traumatic event.

Anxiety manifests itself in the form of physical alterations in the body as well, a person experiencing about of anxiety may exhibit the physical symptoms A feeling of dizziness, one may feel as though they are about to faint, shortness of breath, dryness of mouth, increased blood pressure and heart palpitations, feeling nauseous etc[4]. Anxiety disorders can affect a huge chunk of an individual's life, in 2017, 44.9 million people in India were suffering from anxiety disorders [5].

1.2 Understanding Factors Affecting Anxiety

The feeling of worrying thoughts, mental pressure and accompanying physical signs and symptoms akin to cold sweating, blood pressure, dry mouth and several other implications constitute anxiety. This feeling of anxiety can present itself due to several reasons and situations, such as a job interview, or an exam [3]. Nevertheless, anxiety can present a problem when it becomes perpetual and recurrent for an individual, therefore it becomes necessary to draw out a distinction between 'usual' anxiety and disorders pertaining to anxiety that can be caused concern [4]. Anxiety disorders can be devastating, considering that any individual suffering from its effects will find it extremely hard to go about their daily routine without entering the vicious cycle of intrusive and worrying thoughts this may lead to adverse and long and lasting effects on mental health.

The diagnosis of anxiety disorders does not narrow itself down to one single method, in this domain, questionnaires present themselves as useful tools to judge the state of mind of an individual and their level of anxiety, they consist of questions that directly highlight symptoms that an individual may be manifesting, this is determining level of anxiety of an individual, GAD-7(General Anxiety Disorder), DASS-21 (Depression Anxiety And Stress Scale), PTSD questionnaire(Post-traumatic Stress Disorder), Social Anxiety Disorder questionnaire are some of many questionnaires that are for this.

The objective of this paper was to quantify anxiety in University students and classify it into the mild, moderate, and severe category using machine learning algorithms by considering just the effects observed. The main contribution this research aims to make is to target the causes and effects of anxiety in university students and classify it to produce awareness about the potentially immense mental pressure faced by them.

This research paper has been organized into numerous sections, section 2 elaborates on the work that has been done regarding past research related to the topic of this research, section 3 describes the research methodology that has been used to achieve the objective of this paper. Section 4 consists of the result analysis followed by conclusion and future scope in section 5.

2. Related Work

The objective of this paper was to review how different ML algorithms are used to diagnose different mental illnesses and how they are employed. The algorithms reviewed are Support Vector Machine (SVM), Gradient Boosting Machine (GBM), Naïve Bayes, Random Forest and K-Nearest Neighbors since they were in frequent use. It begins by defining different types of machine learning algorithms into broad categories, i.e., supervised machine learning and unsupervised machine learning, it then touches upon different algorithms being talked about throughout the research and their subtleties and nuances. A research search engine is used to search for articles related to mental illnesses involving the use of machine learning. The articles are then categorized in accordance with mental illness type, ML techniques, accuracy, and sample size. Mental illnesses such as PTSD (Post Traumatic Stress Disorder), Schizophrenia, depression and autism are considered. Different ML algorithms are applied in numerous research papers using various datasets that yield respective accuracies. It was found out that most of the SVM classifiers yielded an accuracy of more than 75% in the papers that it was used [6]. Srividya, Mohanavalli and Bhalaji employ numerous machine learning algorithms, such as SVM (Support Vector Machine), KNN (K-Nearest Neighbor), Logistic Regression, Naive Bayes Classifier to identify the state of mental health in a certain target group. The questionnaire

consisted of a scoring system constituting a range 0-100, where 0-30 was labelled as mentally distressed, 31-60 as neutral, and 61-100 as happy. All the algorithms performed almost equally well, and the use of ensemble machine learning yielded an accuracy of 90% [7]. T. Wyatt, S. Oswalt and Y. Ochoa have performed the examination of differences in the diagnosis related to mental health and the academic impact related to the same, considering the year of college as a classification factor, they have done a secondary data analysis here where responses from US Undergraduate students from the American College Health Association-National College Health Assessment II were considered, the sample size was 66,159 students and regression analysis was performed. The results of this paper outlined the distinction in diagnoses of mental health, found by classification, where it was reported that first year students exhibited a higher rate of injury to self and serious consideration of suicide. The rest of the upper students a heavier academic impact due to factors relating to mental health [8]. E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K. Oedegaard and J. Tørresen perform surveys related to recent research in the field of Mental Health Monitoring Systems(MHMS), which uses Machine learning and sensor data, the focus has been placed on numerous mental conditions such as anxiety, depression, stress, bipolar disorder etc, a classification structure has been proposed to direct the audit of the related work and the research challenges have been discussed. [9] Similarly, N. Mohd Shafiee and S. Mutalib review mental health problems amongst higher education students in Malaysia, it talks about the higher education system in Malaysia and the factors affecting the students' mental health, they have compared different research papers in order to determine research gaps. The most common factors affecting the mental health were identified as lack of social support, the learning environment and financial problems, it was found that related research uses the SVM (Support Vector Machine) algorithm the most, and it yields an accuracy of 70% to 90% [10]. Another review performed by A. Le Glaz summarizes the studies in which machine learning and Natural Language processing was utilized for the purpose of mental health, in addition to this, the potential use of these methods was taken into consideration for clinical practice. 58 articles out of a total of 327 were included in the review, it was observed that population studies could be widely grouped into 3 categories: social media users, patients included in medical databases and patients coming into emergency. Python was used most frequently as the language [11]. P. Tiffin and L. Paton explore the benefits of applying machine learning to mental healthcare in addition to conventional statistical modelling, the challenges and limitations have been discussed as well. They have described machine learning as a tool in the box of several others that can pave the way to real time practice [12].

M. Dobias, M. Sugarman, M. Mullarkey and J. Schleider find out if adolescents battling depression had access to mental health treatments. They used symptoms of depression and sociodemographic variables for this purpose. They performed a secondary data analysis; the dataset was from the National Survey of Drug Use and Health. 4 cross-validated random forest models were used to predict whether depressed adolescents with symptoms that were elevated had accessed treatment in the prior 12 months, it was found that 53.38% of them had accessed some form of treatment, it was found that neither the symptoms of depression, nor the socioeconomic variables could predict the access of treatment, they also go on to highlight the limitations of the research [13]. Y. Liu, J. Hankey, B. Cao and P. Chokka screen for MDD (Major Depressive Disorder) in tertiary care environments. They use machine learning to test their composite screening application, which involves early detection and life history factors, the primary aim in this study was to improve specificity in a tertiary setting. There was a total of 955 participants in this study that went through an interview with a psychiatrist, here, machine learning was used to make more accurate predictions on an individual level, the accuracy of their tool was 72%, the sensitivity was 74.2%, and the specificity was 69.8% [14]. A. Sheik Abdullah, R. Abirami, A. Gitwina and C. Varthana examine the psychotherapeutic support offered to students to overcome distress during the COVID-19 pandemic, it does so by using machine learning techniques. It evaluates the effectiveness of the academic performance of the students during the situation of the pandemic for this purpose. They have taken a hybrid approach using a generic algorithm and upon statistical evaluation, they have used an artificial neural network. The three key factors taken into consideration are cognitivism, behaviorism and social constructivism. The dataset included 775 students. The data was processed by the algorithm and then classified using the neural network. The hybrid approach yielded an accuracy of 88.18%, precision 94.69%, recall 92.94%. The P value for the Fisher's F-Statistical test was $P < 0.001$ [15]. F.M. Albagmi, A. Alansari, D.S.A Shawan, H.Y. AlNujaidi and S.O. Olatunji leverage the power of machine learning by classifying two-class and three-class anxiety issues at an early stage during the covid-19 pandemic, the dataset was collected via an online survey and consisted of 3017 participants, they identified the factors affecting anxiety levels, followed by responses from the GAD-7 questionnaire, following which they used the Support Vector Machine classifier which yielded promising results [16]. K.-S. Na, S.-E. Cho, and S.-J. Cho follow an approach based on machine learning, they aim to draw out a distinction between panic disorder and other anxiety disorders by taking

HRV(Heart Rate Variability) into consideration, the dataset consisted of 60 patients diagnosed with panic disorder and 61 others diagnosed with other anxiety disorders, it was seen that logistic regression performed best with an accuracy of 78.4% [17]. W. A. Bokma et al., focus on the recovery period of anxiety, and predict it using machine learning algorithms, they used a dataset consisting of 887 patients with different forms of anxiety disorders, several baseline predictors were used to predict the recovery process, the random forest algorithm was used in this case and it was found that 484 patients had recovered from their anxiety disorders [18]. V. Khullar, R. G. Tiwari, A. K. Agarwal, and S. Dutta focus on GAD(Generalized Anxiety Disorder) and enumerate an automatic intelligent system to discern anxiety based off of physiological symptoms. The dataset was collected and pre-processed a comparison was drawn out between ML algorithms and ensemble learning [19]. I. C. Passos et al., zero in on the probability of suicide attempts amongst patients with mood disorders, the dataset included a total of 144 patients that were diagnosed with mood disorders, the machine learning model was trained on the basis of demographic variables, the subjects were then categorized into suicide attempters or non-attempters, the accuracies lied between 65% to 72% [20]. L. Jani Anbarasi, M. Jawahar, V. Ravi, S. M. Cherian, S. Shreenidhi & H. Sharen aim to analyze sleeping disorders and anxiety disorders amongst a wide variety of individuals belonging to different strata of life in view of the recent COVID-19 pandemic, the study also aims to predict symptoms in association with stress induced due to the pandemic. A total of 740 participants were included in the research, GAD-7 and PSQI index were used to measure stress. The study highlighted the connection between a lower social status and mental health problems as well. [21]

3. Experimental Setup

The following software was used for the purpose of this study:

1. IBM SPSS: A software developed by IBM for the purpose of data analysis.
2. Jupyter Notebook: A server-client application that helps the user to run and edit notebook documents through a web browser.

In addition to the above, the following python libraries were used:

3. NumPy: A powerful python library that is used for the purpose of working with multi-dimensional arrays and matrices.
4. Matplotlib: A cross-platform library that is used for the purpose of data visualization and the plotting of graphs, usually used in conjunction with NumPy.
5. Pandas: A software library used for the purpose of manipulation and analysis of data in a tabular form.
4. Scikit-learn: A robust software library that offers a wide variety of tools for the purpose of implementing machine learning algorithms such as regression and classification algorithms.

4. Research Methodology

From the related work, it can be observed that numerous statistical analysis techniques as well as machine learning algorithms have been used to deduce meaningful results in terms detection, prediction and diagnosis of mental health issues. The complete procedure that was followed for the purpose of this study is mentioned in Fig. 1.

Step 1: Identify factors that affect the mental health and well-being of Indian engineering students: In order to achieve this task, an in-depth literature review was done, and numerous factors that affect the mental health of university students were deduced and taken into consideration, each individual research work was scanned, and various pre-existing questionnaires such as the GAD-7(General Anxiety Disorder), DASS-21 (Depression Anxiety And Stress Scale) were referred to as well. In addition to this, a psychologist was also consulted for the purpose of the identification of the factors.

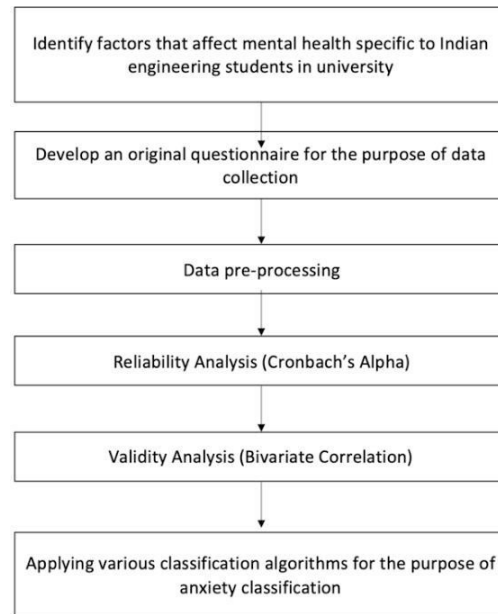


Fig. 1. Research Workflow

Step 2: Develop an original questionnaire for the purpose of data collection: After taking various factors that affect the mental health of engineering university students into consideration, an original questionnaire was developed which has 40 questions targeting the potential causes as well as effects of anxiety experienced by Indian engineering students, it carefully considers all aspects surrounding university life and narrows down potential sources of stress to a few common parameters that affect a large chunk of students, such as, social life, academics, family life, financial situation etc. This questionnaire was distributed amongst students and garnered 127 responses. The questions on the survey follow a cause and effect relationship, which consists of questions targeting the potential causes of anxiety, such as, “I get distressed/anxious due to my friends and/or their behaviour.”, this questions suggests that a student may be anxious due to the company they are keeping, similarly there are questions that target the potential effects of this anxiety that may manifest in a student, such as, “I can’t seem to enjoy things which I used to before such as TV, games, movies, music, etc”, these questions help us understand the impact that this anxiety is making.

Step 3: Data pre-processing: This is one of the most crucial steps to follow, the data that was gathered consisted of a small number of missing values that were not extremely significant to the research, therefore those data points were not considered. In addition to this, data transformation has been done that converts the string data responses to numeric data adhering to the Likert Scale, for example, “Strongly Agree” was assigned a numeric value of “5”, and “Strongly Disagree” was assigned a numeric value of “1”. In addition to this, for some questions, an inversion of scales had to be implemented for them to adhere to the pattern of anxiety level calculation. The dataset was then split into the training set and test set, where 70% of the data points were considered for the training set and 30% of the data points were considered for the test set. Standardization was also done for the process of feature scaling.

Step 4: Reliability Analysis: The Cronbach’s alpha test was performed for the purpose of reliability analysis of the questionnaire. This was done to check whether the questionnaire was consistently able to measure the unidimensional factor of anxiety levels. The Cronbach’s alpha test happens to be one of the tests which is used to determine inter-item reliability and consistency, especially given that the survey been put by masses of the analysis and subjected to test has been tailored to Likert Scale measurement. There is a mathematics involved with the explanation of the Cronbach’s alpha test, it is characterized by the formula given below:

$$a = (k * c) / (v + ((k - 1) * c)) \quad (1)$$

where c is average of all covariances between items, k is the number of scale items and v denotes the average variance of each item. A score for every one item of the scale is calculated, and a comparison is draw out with the total score for each observation that the scale consists of in to calculate the Cronbach’s alpha, this is then compared to the variance & the individual item scores. To interpret the Cronbach’s alpha analysis result, the purpose of some parameters needs to be understood. The alpha coefficient falls within the range of 0 to 1, 0 shows that the inter-item relatability

is absent and that the items on that scale are completely independent of one another, which also signifies that there is no reliability. On the hand, if the alpha value deduced were to approach 1, then that would mean that the inter-item reliability is high, which shows that the consistency of the results is high, and that the scale is highly reliable [20]. Alpha approaches one if the number of items on the scale approach infinite. Another way of looking at it is, if the extent to which the items on the questionnaire are measuring the same variable needs to be known, in which case, '1' would mean that all the items on the scale consistently measure the same thing and '0' would mean all the items are independent and measure completely different things altogether. However, 0 and 1 are ideal values, and the real-world values of alpha lie between 0 and 1, a "good" alpha value completely depends on the scale being used and what it aims to measure however, it is said that a good alpha value should lie within the range of 0.65 to 0.8 (or higher). If the scale that is being analyzed is unidimensional then any values below 0.5 should be immediately rejected. It should be noted that the alpha value is a function of both the number of items on the scale and the covariances, this means that the alpha value can be increased by simply increasing the number of items present on the scale, therefore it should be noted that a very high alpha value is subject to redundancy in the items. For this research, IBM SPSS software was used in order to perform both the reliability and validity analysis.

Step 5: Validity Analysis: This analysis was performed between 2 variables, the anxiety calculated using the cause-based questions and the overall effects of anxiety. In general, a link between two items can be identified as the correlation between them, the value denoting that correlation is said to be within the range of -1 to +1, this is also known to be the Pearson's correlation coefficient (r) when measured between two continuous variables, a direct relationship is characterized by the positive value of the coefficient and an inverse relationship of any two variables is characterized by a negative value, however, the value 0 is indicative of the fact that there is absolutely no linear relationship that is there between the two variables being subjected to the test, although useful, this should not be construed in a sense that other forms of non-linear relationships do not exist between the two variables.

Step 6: Applying various classification algorithms for the purpose of anxiety classification: The dataset was separated in 3 classes of anxiety levels, i.e., mild, moderate, and severe. This was done according to the following mean range of anxiety calculated using the cause-based questions:

- Mild: 1 to 2.3
- Moderate: 2.3 to 3.6
- Severe: 3.6 to 5

After this, the new data-points were attempted to be classed given the values of effects-based questions, that is, the severity of the effects of anxiety was used as training data to correctly class the new data-points. The questions that target the effects of anxiety were split into the training and test set in a 70% and 30% ratio. The algorithms applied were the Naïve Bayes, Random Forest, Decision Trees, and SVM (Support Vector Machine).

5. Result Analysis

The reliability and validity analyses followed by the application of Machine Learning algorithms which were applied to dataset yielded certain significant results which are discussed in this section. While performing a reliability analysis on the whole dataset, an alpha value of 0.723 was achieved. Furthermore, a reliability test on the questions that pinpoint the causes of anxiety yielded an alpha value of 0.871, similarly, questions that pinpoint the effects of anxiety yielded an alpha value of 0.843. This is shown in Table 1. This provided preliminary evidence of the following fact: The questions on the questionnaire reliably and consistently measures anxiety levels, since our Cronbach's alpha value falls within the range of 0.7 to 0.9. The Cronbach's alpha test consistently yielded a value above 0.7, which is indicative of the fact that this questionnaire consistently measures the underlying factor of anxiety calculated using the cause-based questions, in addition to this, a bivariate correlation (validity analysis) is performed between levels of anxiety deduced from the cause variables and overall effects of anxiety (average). The validity analysis yielded a Pearson correlation value of 0.823. SIG 2 tailed value for the validity analysis is 0.001 (< 0.05) which indicates a significant directly proportional relationship. The validity test, namely, the Pearson's correlation analysis proved that the relationship between average anxiety pertaining to all data-points and overall effects of anxiety holds a positive, directly proportional relationship, that is, if the causes rise, the effects of anxiety will also rise. The application of ML algorithms can be interpreted as observed by the accuracy levels. The accuracies obtained are shown in Table 2. Fig. 2 depicts the accuracies yielded by different machine learning algorithms applied. The naïve bayes classification algorithm yielded an accuracy of 71.05 %, that means 71.05 % of

the data-points were correctly classed after being trained on the data provided, which consisted of the effects of anxiety questions and the level of anxiety calculated using the questions on causes, and the random forest algorithm, decision tree & SVM algorithms yielded accuracies 78.9 %, 71.05%, 75.5%. The random forest algorithm appears to have performed the best. The computation time for the SVM, Random Forest, Naïve Bayes and Decision trees algorithm was noted as 0.016s, 0.048s, 0.028s and 0.013s respectively.

Table 1. Alpha values for different variables.

Variables Considered	Alpha Value
All Variables	0.723
Cause-based variables	0.871
Effect-based variables	0.843

Table 2. Accuracies for classification algorithms.

Algorithm	Accuracy
Decision Trees	71.05%
Naïve Bayes	71.05%
Random Forest	78.9%
Support Vector Machine	75.55%

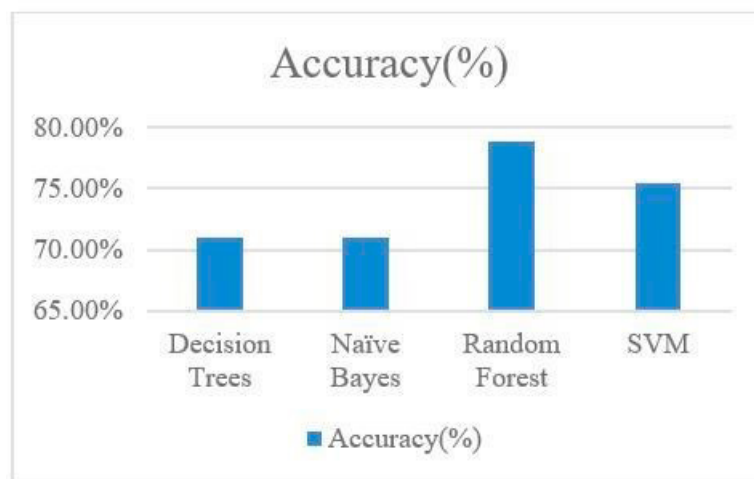


Fig. 2. Graph Representing Accuracy

6. Conclusion and Future Scope

The result of this study yields a wide variety of meaningful conclusions. The reliability and validity analyses provide a strong foundation for the correctness of the questionnaire and the factors they aim to analyze. The reliability analysis is indicative of consistency in the measurement of anxiety levels, that is, it consistently measures those levels of anxiety being calculated. In addition to this, the validity analysis indicates a positive relationship between the measured anxiety levels and overall effects of anxiety, the evidence here is clear: with the rise of causes of anxiety, the overall effects will also be aggravated, this is a powerful conclusion that can be drawn and used as a strong pillar for further research, causes of anxiety can be pinpointed and linked to the effects, which will eventually pave a path to mitigation. Apart from this, the machine learning algorithms performed almost equally well, averaging a 75% accuracy, however the random forest algorithm performed best, yielding an accuracy of 78.9%, which alludes to the fact that the severity of effects of anxiety can be a significant factor that can be used to classify anxiety into mild, moderate and severe categories, this information wields a lot of power, in that it can help pinpoint several students that may be at risk of a severe mental breakdown. In the future, this research can be extended to various other departments apart from the engineering and technology department

to help understand the causes and effects of anxiety amongst students who are studying different courses, moreover, the research can be extended to other locations in the country as well to understand what impact different regions have and what challenges students living in these regions to face, as well as how these challenges and anxiety triggers differ from one another, so as to cater to each subset of students in the most competent and effective manner.

References

- [1] Swain, Prafulla Kumar, Manas Ranjan Tripathy, Subhadra Priyadarshini, and Subhendu K. Acharya (2021) “Forecasting suicide rates in India: An empirical exposition”. *PLOS ONE* 16(7):1-21.
- [2] Sawhney, Manisha, Seth Kunen, and Ashum Gupta (2018) “Depressive Symptoms and Coping Strategies Among Indian University Students.” *Psychological Reports* 123(2): 266-280.
- [3] Kazdin (2000) “Encyclopedia of psychology”, Oxford University Press, New York.
- [4] Kessler, Ronald C, Ayelet Meron Ruscio, Katherine Shear, and Hans-Ulrich Wittchen (2009) “Epidemiology of Anxiety Disorders”. *Behavioral Neurobiology of Anxiety and Its Treatment*: 21-35.
- [5] India State-Level Disease Burden Initiative Mental Disorders Collaborators. (2020) “The burden of mental disorders across the states of India: the Global Burden of Disease Study 1990–2017.” *The Lancet Psychiatry* 7(2): 148-161
- [6] Cho, Gyeongcheol, Jinyeong Yim, Younyoung Choi, Jungmin Ko, and Seoung-Hwan Lee (2019) “Review of Machine Learning Algorithms for Diagnosing Mental Illness.” *Psychiatry Investigation* 16(4): 262-269.
- [7] Srividya, M, S Mohanavalli, and N Bhalaji (2018) “Behavioral Modeling for Mental Health using Machine Learning Algorithms.” *Journal Of Medical Systems* 42(5): 1-12.
- [8] Wyatt, Tammy Jordan, Sara B. Oswalt and Yesenia Ochoa (2017). “Mental Health and Academic Success of First-Year College Students. *International Journal of Higher Education* 6(3):178.
- [9] Garcia-Ceja, Enrique, M. Riegler, T. Nordgreen, P. Jakobsen, K. Oedegaard, and J. Tørresen (2018) “Mental health monitoring with multimodal sensing and machine learning: A survey.” *Pervasive And Mobile Computing* 51:1-26.
- [10] Mohd Shafiee, Nor Safika, and Sofianita Mutalib (2020) “Prediction of Mental Health Problems among Higher Education Student Using Machine Learning.” *International Journal of Education and Management Engineering* 10(6):1-9.
- [11] Glaz, Aziliz le, Yannis Haralambous, Deok-Hee Kim-Dufor, Philippe Lenca, Romain Billot, Taylor C Ryan, Jonathan Marsh, Jordan DeVlyder, Michel Walter, Sofian Berrouguet, and Christophe Lemey (2021) “Machine Learning and Natural Language Processing in Mental Health: Systematic Review”. *Journal of Medical Internet Research* 23(5):1-20.
- [12] Tiffin, Paul A, and Lewis W Paton. (2018) “Rise of the machines? Machine Learning Approaches and Mental Health: Opportunities and challenges.” *The British Journal of Psychiatry* 213(3): 509–510.
- [13] Dobias, Mallory L, Michael B Sugarman, Michael C Mullarkey, and Jessica L Schleider (2021) “Predicting mental health treatment access among adolescents with elevated depressive symptoms: Machine learning approaches.” *Administration and Policy in Mental Health and Mental Health Services Research* 49(1): 88–103.
- [14] Liu, Yang S., J. Hankey, Bo Cao, and P. Chokka (2021) “Screening for major depressive disorder in a tertiary mental health centre using Early Detect: A machine learning-based pilot study.” *Journal of Affective Disorders Reports* 3:100062.
- [15] Abdullah, Azman, R. Abirami, A. Gitwina, and C. Varthana (2021) “Assessment of academic performance with the E-mental Health Interventions in virtual learning environment using Machine Learning Techniques: A hybrid approach”. *Journal of Engineering Education Transformations* 34:79-85.
- [16] Albagmi, Faisal Mashel, Aisha Alansari, Deema Saad Al Shawan, Heba Yaagoub AlNujaidi, and Sunday O. Olatunji. (2022) “Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach.” *Informatics in Medicine Unlocked*, 28:100854.
- [17] Na, Kyoung-Sae, Seo-Eun Cho, and Seong-Jin Cho. (2021) “Machine learning-based discrimination of panic disorder from other anxiety disorders.” *Journal of Affective Disorders*, 278:1–4.
- [18] Bokma, Wicher A., Paul Zhutovsky, Erik J Giltay, Robert A Schoevers, Brenda W J H Penninx, Anton L J M van Balkom, Neeltje M Batelaan, and Guido A van Wingen. (2020) “Predicting the naturalistic course in anxiety disorders using clinical and biological markers: A machine learning approach.” *Psychological Medicine*, 52(1):57–67.
- [19] Khullar, Vikas, Raj G. Tiwari, Ambuj K. Agarwal, and Soumi Dutta. (2021) “Physiological Signals Based anxiety detection using ensemble machine learning”. *Cyber Intelligence and Information Retrieval*: 291: 597–608.
- [20] Rakesh, Gopalkumar. (2017) “Suicide prediction with machine learning.” *American Journal of Psychiatry Residents Journal*, 12(1):15–17.
- [21] Anbarasi, L. Jani, Malathy Jawahar, Vinayakumar Ravi, and Sherin M. Cherian. (2022) “Machine Learning Approach for anxiety and sleep disorders analysis during COVID-19 lockdown”, *Health and Technology* 12(4): 825–838.