IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

INSIGHTS INTO COLLEGIATE MENTAL HEALTH: MACHINE LEARNING-BASED PREDICTION OF DEPRESSION RISK FACTORS AMONG TAMIL NADU STUDENTS

[1]A.S Janani, [2] K. Priyadharshini, [3]Dr. P. Meenakshi Sundari

[1]Post Graduate Student, [2]Post Graduate Student, [3]Assistant Professor,

[1]Department Of Computer Science

[1] Fatima College, Madurai, India

ABSTRACT: This paper focuses on using the Python Programming Language for Predicting the Emotional Wellbeing in the college students of Tamil Nadu. It characterizes the various prevalent symptoms with respect to depression found in the degree pursuers. Further, it discusses various Python packages and libraries, including NumPy, Pandas, and Matplotlib, that are commonly used in prediction methodologies. The paper highlights the potential of Python for Analysis of depression symptoms and most frequented symptom in the students who have risk of depression. Overall, this research paper demonstrates how Python can extract insights from College student's data efficiently and effectively.

Keywords: Depression, Mental Health, Emotional Well-being, anxiety, College students.

I. INTRODUCTION:

In today's world, there is a significant decline in mental health among people, with young individuals being particularly affected. Additionally, more than 20% of students have reported being diagnosed with depression. The issue of mental health among students is a growing concern in Tamil Nadu. Among adolescents, the most prevalent mental illnesses encompass anxiety, mood disorders, attention disorders, and behavioural problems. These mental health issues can impact multiple aspects of students' lives, such as diminishing their quality of life, academic performance, and physical well-being This paper is focused on analysing these issues using Machine learning. In the branch of artificial intelligence known as "machine learning," algorithms and models are created that can learn from data and generate predictions.

II. METHODOLOGIES:

The Machine Learning Models are computational algorithms or mathematical representations that enable a computer system to learn patterns and make predictions or decisions without being explicitly programmed. These models learn from data, discovering patterns, relationships, and insights that can be used for various tasks such as classification, regression, clustering, and more. The key components of Machine Learning Models are:

1. Input Data.
2. Model Architecture.
3. Training Process
4. Learned Parameters
5. Output or Predictions

2.1 INPUT DATA: The data utilized in this study were collected through a survey administered to college students from diverse academic institutions across Tamil Nadu. The survey aimed to assess various psychosocial factors potentially associated with depression risk among college students.

2.1.1 Survey Design: The survey instrument utilized in this study was meticulously constructed based on consultations with psychologists and psychiatrists, drawing upon their expertise in the field of mental health and depression risk assessment. It consisted of structured questions designed to gather information on demographics, academic performance, lifestyle habits, social support networks, stress levels, and mental health history.

2.1.2 Data Collection Procedure: The survey was distributed electronically using online survey platforms accessible to college students. Participation in the survey was voluntary and anonymous, with informed consent obtained from each participant prior to data collection. The survey remained open for a period of twelve weeks to allow for maximum participation.

2.1.3 Data Variables: The survey collected information on a range of variables potentially related to depression risk, including but not limited to:

- **Demographic Information**: Age, gender, ethnicity, year of study, and socioeconomic status.
- **Academic Performance**: Grade point average (GPA), academic workload, perceived academic pressure.
- **Lifestyle Habits**: Sleep patterns, exercise frequency, dietary habits, substance use (e.g., alcohol, tobacco).
- **Stress Levels**: Perceived stressors (e.g., academic, financial, interpersonal).

2.1.4 Data Quality Assurance: To ensure the validity and reliability of the data collected, measures were implemented to minimize response bias and data entry errors. These measures included clear instructions provided to participants, validation checks in the survey instrument, and regular monitoring of survey responses for completeness and consistency.

Ultimately, the input data underwent refinement by excluding negative responses from each question, enhancing the accuracy of the data collection process.

2.2 MODEL ARCHITECTURE: Model architecture embodies the foundational design and structure of a predictive model or machine learning algorithm, encompassing its type, component arrangement, parameters, activation functions, and optimization techniques.

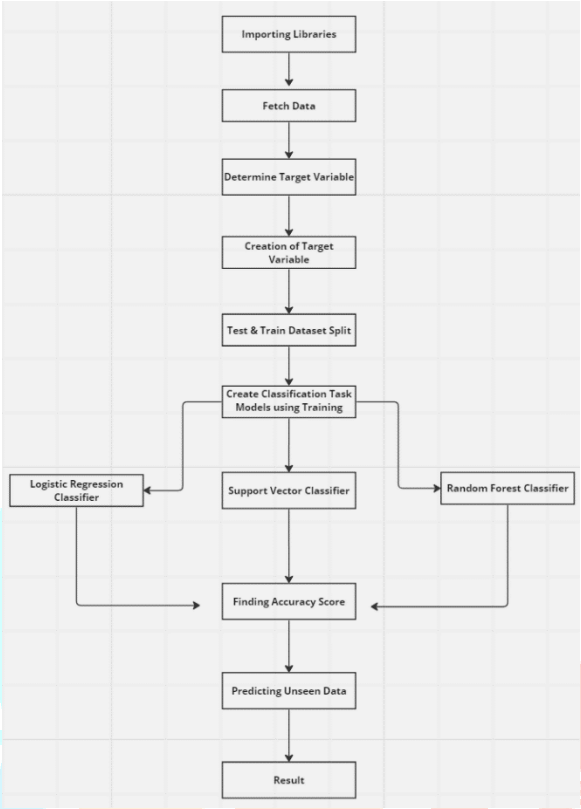


Fig 2.1 Data Flow Diagram for program implementation

The Model Architecture for this paper is developed by,

1. **Collect Input Data:** We sought guidance from psychologists and psychiatrists to aid in crafting the foundational information essential for assessing the likelihood of depression risk. Utilizing this guidance, we formulated a survey comprising questionnaires targeting symptoms associated with depression. This survey was disseminated across colleges in Tamil Nadu, facilitating the collection of relevant data.
2. **Feature Engineering:** The collected data was pre-processed by handling null values and missing values. Along with this quantile transformation is carried out to normal probability distribution, Handle Outliers etc.,
3. **Model Selection:** For predicting the depression risk in college students, classification task algorithms are used. They are logistic regression, Support Vector Machine and Random Forest Classifier.
4. **Model Training:** The selected model is trained by splitting the data into train and test which is then cross validated by the rest of the data. The cross validation for optimizing the performance of the model.

Thus, Model Architecture has been incorporated into this project for selecting model and training the model.

2.3 TRAINING PROCESS:

1. **Data Splitting**: The dataset is split into training and testing sets using the train_test_split function from sklearn.model_selection. This function divides the data into 60% for training and 40% for testing.

2. **Training:** Three classification models are trained: Logistic Regression, Support Vector Classifier (SVC), and Random Forest Classifier (RFC).

3. **Hyperparameters Tuning:** Hyperparameter tuning is performed for each model using GridSearchCV to find the best combination of hyperparameters that optimize a chosen metric in this case, accuracy.

For example, for Logistic Regression:

- A logistic regression model is instantiated (LogisticRegression(solver='liblinear')).
- A grid of hyperparameters is defined (grid_lr), including the solver, penalty type, and C values.
- Stratified K-Fold cross-validation is performed with 100 splits (StratifiedKFold(n_splits=100, random_state=10, shuffle=True)).
- GridSearchCV is applied to search for the best hyperparameters (GridSearchCV(estimator=model, param_grid=grid_lr, cv=cross_val, scoring='accuracy', error_score=0)).

Evaluation: The best models obtained from hyperparameter tuning are evaluated using various metrics such as accuracy, precision, recall, and F1-score.Classification reports are generated to analyse the precision, recall, and F1-score for each class. Predictions are made on the test dataset using the trained models, and the classification reports are printed.

2.4 LEARNED PARAMETERS:

- **Logistic Regression**: In logistic regression, the learned parameters include the coefficients (weights) assigned to each feature. These coefficients determine the importance of each feature in predicting the target variable. After training, the logistic regression model (lr_result) contains these learned coefficients, which can be accessed using lr_result.coef_.
- **Support Vector Classifier (SVC)**: For an SVC model, the learned parameters include the support vectors and the coefficients associated with them. After training, the SVC model (result) contains information about the support vectors and coefficients, which can be accessed through attributes like result.support_vectors_ and result.dual_coef_.
- **Random Forest Classifier (RFC)**: In a random forest classifier, the learned parameters include the feature importances, which represent the importance of each feature in making predictions. After training, the RFC model (result) contains information about feature importances, which can be accessed using result.feature_importances_.

2.5 OUTPUT & PREDICTIONS:

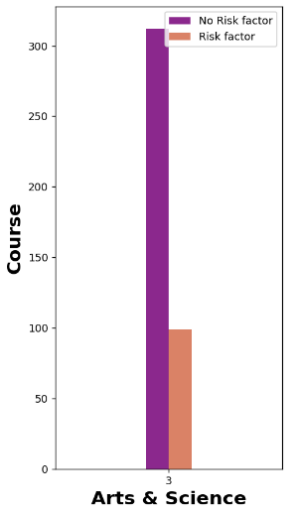


Fig 2.5.2 Arts & Science students have lower risk factors of depression.

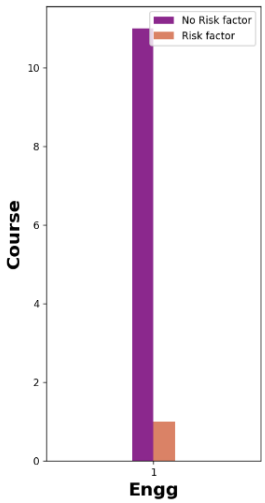


Fig 2.5.2 Engineering students have lower risk factors of depression.

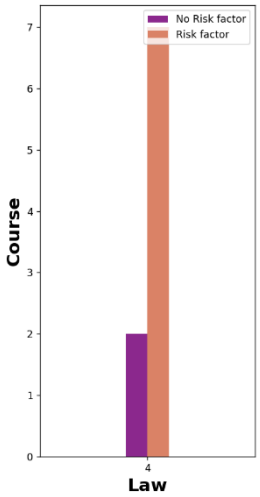


Fig 2.5.3 shows Law students have higher risk factors of depression.

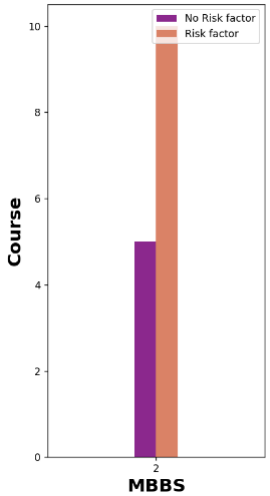


Fig 2.5.4 shows Mbbs students have high risk factors.

Course-Specific Risk Level: The graph indicates that students in MBBS and Law programs have a higher risk of depression compared to those in Arts & Science and Engineering disciplines. This suggests that the demanding nature of medical and legal education might contribute to increased stress levels and susceptibility to depression among students in these fields. Conversely, students in Arts & Science and Engineering programs may experience lower stress levels, possibly due to differences in curriculum structure and support systems. Understanding these differences is essential for implementing targeted interventions to support student mental health across all academic domains.

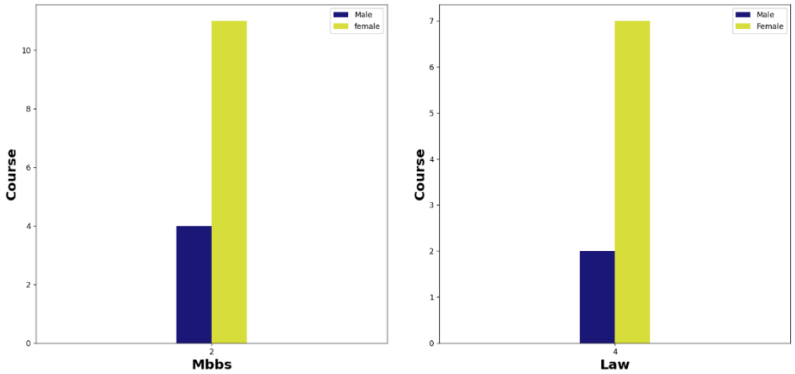


Fig 2.5.5 shows the gender count of the MBBS & Law courses is plotted.

Gender Disparity in MBBS & Law: Female students in MBBS and Law courses face a higher risk of depression compared to males, indicating potential gender-specific stressors. Factors such as societal expectations and workplace dynamics may disproportionately affect females, emphasizing the need for tailored support to address mental health disparities in these fields. Understanding and addressing these differences are crucial for promoting the well-being of all students within these demanding academic disciplines.

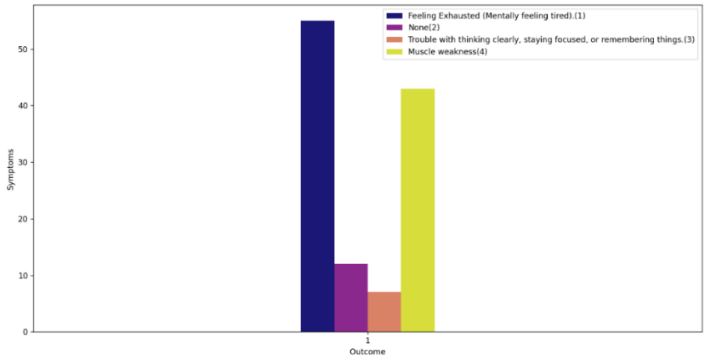
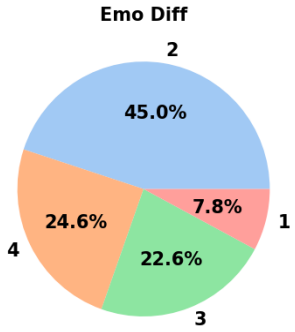
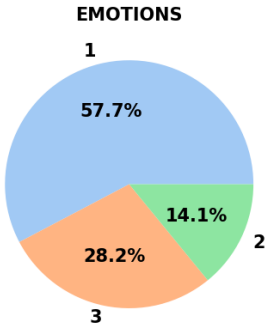
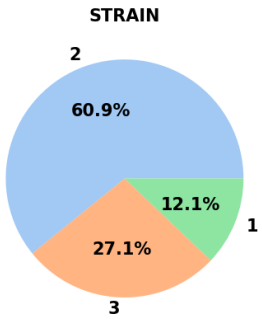


Fig 2.5.6 shows that people with risk of depression have the symptom of Feeling exhausted.

Common Symptom: Feeling exhausted is a prevalent symptom among individuals at risk of depression. This fatigue often manifests as a persistent sense of physical and mental tiredness, regardless of the amount of rest or sleep obtained. It can significantly impact daily functioning, making even simple tasks feel overwhelming and draining. This exhaustion may stem from various factors associated with depression, such as disrupted sleep patterns, persistent negative thoughts, and decreased motivation. Moreover, the emotional toll of managing depressive symptoms can exacerbate feelings of fatigue, contributing to a cycle of exhaustion and decreased well-being. Recognizing and addressing this symptom is crucial in effectively managing depression and improving overall quality of life.







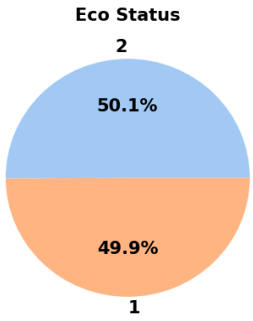


Fig 2.5.7 shows that that people only sometimes have Emotional difficulties.

Fig 2.5.8 shows that people feel completely stuck higher to like nothing gets better.

Fig 2.5.9 shows that people feel strained while building conversation only sometimes.

Fig 2.6.0 shows that students depression risk due economy

III. CONCLUSION

This research reveals that in Tamil Nadu, the risk factors of depression are primarily found in female rather than male and concerning academic disciplines, students enrolled in programs such as MBBS and Law exhibit a higher prevalence of depression compared to those in other courses. From the survey conducted, it is found that feeling exhausted is commonly detected in people with risk of depression. Indians tend to not treat mental health as a serious issue, and this lack of acknowledgment is where the risk factors of depression increases. Some of the strategies that found to be helpful in preventing and treating depression.

Prevention Strategies:

- Encourage regular exercise, balanced nutrition & adequate sleep.
- Foster positive relationships and social connections.
- Strengthen coping skills and resilience to better manage stress.

Home-Based Strategies:

- Promote healthy habits for mental well-being.
- Create supportive home environments.

While home-based strategies contribute to support, clinical treatment remains pivotal in addressing depression. In conclusion, recognizing and addressing the identified risk factors for depression in college students is pivotal for promoting mental well-being on campuses.

REFERENCES

[1] "Machine Learning in Python for Weather Forecast based on Freely Available Weather Data" by E. B. Abrahamsen, O. M. Brastein, B. Lie∗, published in The 59th Conference on Simulation and Modelling, September 28, 2018.+-.

[2] "A Few Useful Things to Know About Machine Learning" by Pedro Domingos, published in Communications of the ACM, October 2012.

[3] "Machine Learning: Algorithms, Real-World Applications and Research Directions" by Iqbal H. Sarker, published in Springer Nature Journal, January 27, 2021.

[4] "Machine Learning Methods for Predicting Postpartum Depression: Scoping" by Zahid Ahmad Butt, MBBS, MSc, PhD, published in JMIR Ment Health, November 24, 2021.

[5] "Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis" by Xibiao Ye, Sunita Bayyavarapu Bapuji, Shannon Elizabeth

Winters, Ashley Struthers, Melissa Raynard, Colleen Metge, Sara Adi Kreindler, Catherine Joan Charette, Jacqueline Angela Lemaire, Margaret Synyshyn, Karen Sutherland, published in BMC Health Services Research, July 2014.

[6] "Machine Learning in Python for Weather Forecast based on Freely Available Weather Data" by E. B. Abrahamsen, O. M. Brastein, B. Lie∗, published in the 59th Conference on Simulation and Modelling, September 28, 2018.

[7] https://www.javatpoint.com/machine-learning-models