A PROJECT REPORT

on

"STUDENT MENTAL HEALTH ANALYSIS"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN

COMPUTER SCIENCE

BY

ANUSHKA SINGH	21052144
RIYA SINGH	21052269
ANANYA THAKUR	21052225
TANMAY PANDEY	21052231
BIBEK SAHOO	21052244

UNDER THE GUIDANCE OF
Professor AJIT KUMAR PASAYAT

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

April 2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is to certify that the project entitled

"STUDENT MENTAL HEALTH ANALYSIS"

submitted by

ANUSHKA SINGH	21052144
RIYA SINGH	21052269
ANANYA THAKUR	21052225
TANMAY PANDEY	21052231
BIBEK SAHOO	21052244

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2023-2024, under our guidance.

Date:13/4/2024

(Ajit Kumar Pasayat)
Project Guide

Acknowledgements

We are profoundly grateful to **MR AJIT PASAYAT** of **KIIT Deemed to be University** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

ANUSHKA SINGH
RIYA SINGH
ANANYA THAKUR
TANMAY PANDEY
BIBEK SAHOO

ABSTRACT

The "Student Mental Health Analysis" project aims to analyse students' mental health using efficient machine learning techniques, as mental health among students is a critical concern related to their academic success and overall well-being. The project employs a robust tech stack, primarily using Python as the programming language and Jupyter Notebook as the development platform. The Frontend is implemented using HTML and CSS, providing an interactive interface for users. The core of this analysis primarily involves implementing numerous machine learning models to help in diagnosing student's mental health.

The system works by enabling users to fill a form asking questions relevant to make predictions. Upon filling the form, the Backend processes the data, extracting relevant information (such as…to be added). Python Libraries, including Pandas are instrumental in performing data manipulation and preprocessing. Matplotlib and Seaborn generate insightful visualizations. Sklearn, Keras and TensorFlow have also been used for model building and training. The Frontend developed with HTML and CSS, offers an intuitive dashboard where users can interact with system easily. Flask is used in Backend to render HTML pages and helps give a better user experience. The selected model is Random Forest, which predominantly has the better result than other models that have also been trained, fine-tuned and evaluated using appropriate metrics to assess their better performances.

The Findings of this project helps diagnose different factors which effect students' mental health with the efficient use of Random Forest. The project helps underlining the importance of proactive measures to support the student's well-being and their better academic growth. This project also serves as a valuable tool to demonstrate the potential of machine learning techniques in addressing and analysing challenges faced by the students regarding their mental health.

Keywords: Student, Mental Health, Diagnose, Python, Jupyter, HTML, CSS, Data Visualization, Text Analysis, Pandas, Matplotlib, Sklearn, Keras, User Interface.

Contents

Chapter 1

Introduction:

Student life is a dynamic and changing time marked by development, learning, and discovery. It includes not just academic achievements, but also social contacts, personal growth, and the formation of lifetime relationships. These years are critical for determining future efforts and laying the groundwork for a successful and rewarding life after college. With so much impact and importance, student life is also filled with many challenges and pressure that cause lifelong imprints.

Students face a diverse range of problems, including academic stress, social pressures, and personal struggles, which can have a significant influence on their mental health. While institutions promote individual progress, they may also create conditions that generate stress and anxiety as a result of heavy workload and social expectations. These pressures frequently lead to academic performance concerns and general emotional suffering among students, eventually leading to poor mental health conditions.

Poor mental health can have a very adverse impact on various aspects of a student’s life . It can cause poor performance, disengagement, and, in extreme circumstances, major mental health crises such as anxiety disorders and depression. Furthermore, students with mental health disorders may struggle to form good connections and deal with daily obstacles.

Our model aims to delve into the intricacies of mental health and accurately predict the possibility of depression among students to prevent these disorders from getting severe. Our project offers an exhaustive approach to assessing and evaluating student mental health. By using specialized surveys and innovative techniques, we want to capture students' varied experiences and unique needs. This methodology provides a refined and successful method for assessing mental health status, identifying significant risk factors, and implementing focused interventions to promote student well-being. Finally, we hope to contribute to a comprehensive strategy that incorporates mental health support into the educational framework, ensuring that students have the resources and assistance they require to prosper academically and personally.

Chapter 2

Literature Review

1. In the study by (Aleem et al., 2022) machine learning techniques were employed to predict mental health issues across three distinct categories: classification, deep learning, and ensemble methods. This study underscores the significance of machine learning models in healthcare, particularly in analyzing vast datasets and furnishing valuable insights for mental health assessment.

2. In a study conducted by (Sumathi & B., 2016), they investigated the prediction of mental health issues among children using different machine learning methods. They observed various factors, symptoms, and psychological assessments conducted by professionals. Among the methods considered, Multilayer Perceptron (MLP) exhibited the highest level of accuracy, reaching 78%
3. In Ahmed et al.'s 2019 study, various machine learning models were tested on a dataset related to depression and anxiety. The Convolutional Neural Network (CNN) outperformed others, achieving remarkable accuracies of 96.0% for anxiety and 96.8% for depression prediction. 10.1109/CSDE50874.2020.9411642

4. In Jerry et al.'s 2019 study, depression was analyzed using text and audio datasets. Random forests achieved the highest accuracy of 0.73 on the text data, while XGBoost performed best with an accuracy of 0.50 on the audio data. Xu, A. J., Flannery, M. A., Gao, Y., & Wu, Y. (2019). Machine learning for mental health detection.

5. Chung & Teo (2022) conducted a thorough review of machine learning's role in diagnosing mental health conditions, including schizophrenia and childhood issues. They identified challenges such as diverse data types and model interpretation. To maximize machine learning's potential in mental health diagnostics, future efforts should focus on refining models, improving data quality, and addressing ethical considerations.

6. Sajja 2021 explores machine learning algorithms for predicting anxiety and depression using speech data. Preprocessing enhances data consistency. SVM outperforms Random Forest and Naïve Bayes, achieving 94% accuracy. The study underscores the efficacy of machine learning in mental health prediction, with SVM demonstrating superior performance.

7. Chekroud et al. (2016) evaluated gradient boosting's performance in detecting depression in 1949 patients with level 1 depression, yielding a 64.6% accuracy. This suggests its potential for precise depression screening in clinical contexts.

8. Sau and Bhakta (2019) explore machine learning's efficacy in screening anxiety and depression among seafarers. CatBoost emerges as the top performer, with 82.6% accuracy and 84.1% precision, indicating its potential for automated mental health screening in maritime settings.

9. In a study by Lahiri Mallik et al. (2023) depression's prevalence among college students is highlighted. It adversely affects academic performance, social relationships, and mental health, emphasizing the critical need for university support services.10.5281/zenodo.8255225.

10. Alharbi et al. (2019) surveyed 1245 high school students in Qassim Region, using PHQ-9 and GAD-7 questionnaires. Results showed: 26.0% not depressed, 34.0% mildly, 24.6% moderately, 10.4% moderately severe, and 5.0% severely depressed. Anxiety levels: 36.5% without anxiety, 34.1% mild, 19.5% moderate, and 9.8% severe anxiety.

Chapter 3

Basic Concepts

3.1 Machine Learning Model: Random Forest

Random forest is a powerful ensemble Machine Learning technique used for both classifying and regressing tasks. It works by creating several decision trees during the training phase where each tree is constructed using a random subset of the data set. In prediction, the algorithm aggregates the results of all trees either by voting (for classification) or by averaging prediction (for regression).

Concept Description:

Ensemble Learning: Random Forest is a type of Ensemble Learning where multiple individual models are combined to produce a stronger model. It borrows the principle "wisdom of the crowd", where aggregating predictions from multiple models benefits the total yield.
Decision Trees: A decision tree is one of the most powerful tools of supervised learning algorithms. It builds a flow-chart-like structure where each of the internal nodes, branch, and leaf nodes represents the feature, decision rule, and outcome.
Bootstrap Aggregating (Bagging): It's an ensemble learning model, where multiple-week models are trained with different subsets of training data. Here each subset is sampled with the help of replacement and the predictions are made by averaging the weak models for the regression problem and considering the majority vote for the classification problem.
Random Feature Selection: It's an embedded method, where the methods combine the qualities of filter and wrapper methods. Using this method, Random Forest only considers a random subset of features for splitting at each node, rather than all features. Due to this it also prevents from overfitting.

3.2 Natural Language Processing: Sentiment Analysis

Sentiment Analysis is a component of Natural Language Processing (NLP) that seeks to understand the emotional tone embedded in textual content. It endeavours to ascertain whether the sentiment conveyed is positive, negative, or neutral.

Concept Description:

Text Preprocessing: Text processing is one of the essential steps which is being performed in NLP, because it helps in cleaning and transforming the raw data into a suitable format which can be then used for analysis or modelling. Some steps like tokenization, lowercasing, removing stop words, and stemming or lemmatization are performed to clean and standardized the data.

Sentiment Lexicons: Sentiment Lexicons serve as a method within NLP for discerning the sentiment conveyed within a text. It relies on a carefully selected list of words and phrases linked to various emotions to label the text and determine the sentiments conveyed.
Machine Learning Approaches: In sentiment analysis, researchers try to use some sophisticated methods like Logistic Regression, Support Vector Machine (SVM), and Neural Networks. These methods work like smart algorithms that learn from examples. By feeding them lots of text data with known sentiments (positive, negative, or neutral), they learn to recognize patterns in the text. Once trained, they can predict the sentiment of new text they've never seen before. It's like teaching a computer to understand feelings by showing it many examples of how people express different emotions in writing.

Chapter 4

Problem Statement / Requirement Specifications

The problem addressed in this project revolves around predicting the mental well-being and stress status of students based on various factors like lifestyle, demographics, and sentiment analysis of responses provided by students in textual format. Mental wellness among students is a critical concern, as there is an increase in stress, anxiety, and depression due to academic or by any means. Identifying early indications of mental health issues and providing guidance can promote student well-being and academic success.

4.1 Project Planning

The success of this project hinges well on a structured and well-defined project planning phase. Initially the project objective was identified, the planning involved defining the project scope and features. The tam identified key functionalities such as form distribution, data collection, model implementations, diverse data visualizations which ensured a clear roadmap for development.

The foremost aspect of planning was to identify the selection of appropriate technologies. Python was chosen as the language due to its versatility in machine learning, while Jupyter served as the perfect IDE due to its robust environment.

The timeline for the project was carefully defined, considering the implementation of model then development of backend and the integration of frontend interfaces. This structured planning led us to the systematic and efficient development of this project which meticulously also address any unforeseen challenges.

4.2 Project Analysis

The analysis phase of this project goes deeper into visualizations and processing. The very initial step involved data prepossessing. The initial step involved manipulation of collected data to be used by the Pandas library and functions to convert data useful for the use in project.

In terms of data visualization, the project uses Matplotlib and Seaborn to accurately generate a diverse range of charts and graphs. These visualizations were instrumental in representation of complex data in an accessible and accurate manner.

Additionally, the project also integrated external libraries like Sklearn to train and implement machine learning model. Also, Keras and TensorFlow helped in implementation of Neural networks.

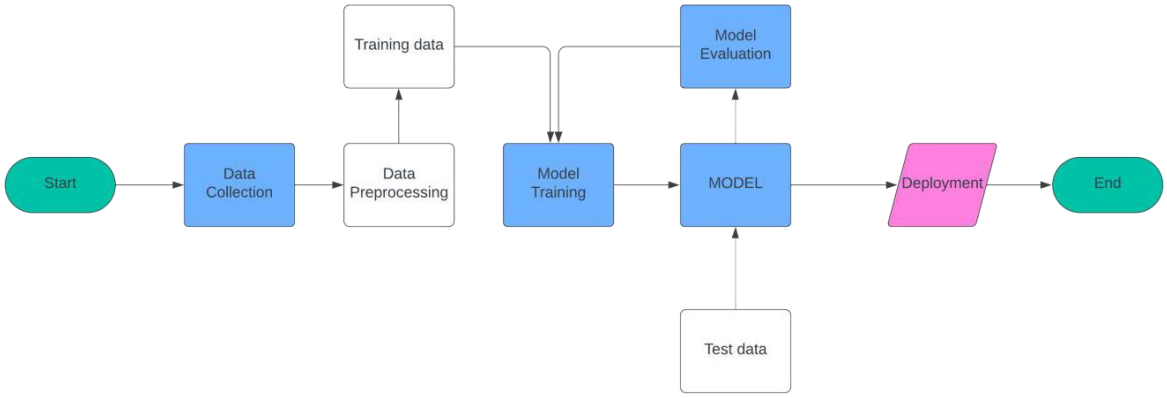
The “Student Mental Health Analysis” project’s analysis phase thus helped user to easily fill form and get a look on their mental status. The combination of robust data processing and effective model implementation gives a comprehensive analysis framework that aligns with the project’s scope and goals.

Chapter 5

Implementation

5.1 Methodology

System Design



For this project, following listed methodologies are opted:

Data Collection:

Data collection is one of the crucial steps to be taken care of while implementing a machine learning model. HTML is used to create a Google form and is implied as our tool to collect responses from students. Responses collected are then stored in a CSV file for processing.

Categorized distribution of questions in the form:

Demographic Information: This category includes questions aimed to gather basic demographic information about the respondents. It also provides the context of respondents' backgrounds. Analyzing those demographic factors can help in understanding how stress and emotional well-being can differ among different groups.

1. Age
2. Which state are you from?
3. Gender
4. Grade Level/Year in School?
5. What are you currently studying?

Emotional Well-being: These questions focus on assessing various aspects of emotional well-being and mental health. They inquire about the frequency and intensity of feelings such as sadness, anxiety, and hopelessness over the past two weeks. Additionally, they address physical symptoms, changes in appetite or weight, sleep disturbances, difficulty concentrating, and thoughts of self-harm. By probing into these areas, the questions aim to gauge the respondents' overall emotional health and identify potential signs of stress or mental health issues

1. Do you think you have depression or anxiety?
2. Have you ever been diagnosed with depression or anxiety by a healthcare professional?

Data Preprocessing:

The raw data was cleaned, encoded, and transformed so that it could be analyzed for model training. This involved the collection of data from students using Google Forms, handling missing values, encoding categorical variables, and performing sentiment analysis on textual data.

Training Data:

Training data refers to the portion of the dataset that is used to train a machine-learning model. Training data consists of cleaned and preprocessed data that includes various features related to student's mental health, demographic information, emotional well-being, etc. Training data would typically be divided into two main components:
Features: These are the variables or attributes that serve as inputs to the model.
Labels or Target Variable: In supervised learning, the training data would also include the target variables or labels that the model aims to predict.
Once the training data is prepared, it is used to train the machine learning model through an iterative process where the model learns to identify patterns and relationships between the input features and the target variable.

Model Training:

Model training is the process of teaching a machine learning algorithm to recognize patterns and make predictions based on input data. Here's a breakdown of the steps involved in model training:

Choose a Model: The first step is to select an appropriate machine learning model for the task at hand. This choice depends on factors such as the nature of the problem (e.g., classification, regression), the size and complexity of the dataset, and the computational resources available. For our model, we have chosen Random Forest Classifier, Support Vector Machines, and Neural Networks.

Prepare the Training Data: As mentioned earlier, the training data should be preprocessed and cleaned to ensure that it's in a suitable format for model training. This involves handling missing values, encoding categorical variables, and processing textual data if applicable. The training data is typically split into features (inputs) and labels (outputs) before feeding it into the model.

Split Data into Training and Validation Sets: Before training the model, it's important to split the training data into two subsets: one for actual training and the other for validation. The training set is used to train the model, while the validation set is used to evaluate its performance and make adjustments to prevent overfitting.

Train the Model: With the prepared training data, the selected machine learning model is trained using an optimization algorithm to minimize the difference between the predicted outputs and the actual labels. During training, the model adjusts its internal parameters based on the training data to improve its predictive accuracy. This process is repeated for multiple iterations or epochs until the model's performance on the validation set converges or reaches a satisfactory level.

Throughout the model training process, it's important to maintain good practices such as monitoring performance metrics, experimenting with different algorithms and hyperparameters, and documenting the steps taken to ensure the reproducibility and reliability of the trained model.

Model:

We have used a Random Forest Classifier, Support Vector Machine, and Neural Networks as our models.

- Random Forest Classifier: A Random Forest classifier was trained on the basis of pre-processed data to predict the mental wellness and stress levels of students based on various features.

Advantages:

1. Robustness: Random Forest is crafted to resist overfitting, unlike individual decision trees. By amalgamating multiple trees and employing random feature selection, it maintains a balanced predictive performance.

2. High Accuracy: Esteemed for its precision in both classification and regression tasks, Random Forest consistently delivers commendable accuracy.

3. Implicit Feature Selection: One of its remarkable traits lies in its knack for automatically pinpointing crucial features within the dataset. By discerning features frequently employed in tree splitting, it unearths the most informative attributes.

- Support Vector Machine (SVM): An SVM classifier with a sigmoid kernel was trained on the values of features or variables with similar scales to predict mental wellness.

Advantages:

1. Effectiveness in High-Dimensional Spaces: SVM showcases prowess even in sprawling high-dimensional spaces, an invaluable trait for intricate tasks like text classification, image recognition, and bioinformatics.

2. Versatility: SVM boasts a repertoire of kernel functions ranging from linear to radial basis function (RBF), affording it the adaptability to tackle diverse datasets.

3. Optimization Objectives: SVM adheres to a convex objective function, ensuring the discovery of globally optimal solutions. This strategic approach steers clear of entrapment in local optima during training, a testament to its reliability.

- Neural Network: A Feedforward Neural Network (FNN) with multiple hidden layers was developed using some open-sourced platforms like TensorFlow and Keras.

Advantages:

1. Non-Linearity: The allure of neural networks lies in their adeptness at capturing intricate non-linear relationships between inputs and outputs. This makes them indispensable for tasks where conventional linear models falter.

2. Feature Learning: Neural networks exhibit a remarkable talent for autonomously gleaning pertinent features from raw data. This negates the arduous manual feature engineering process, particularly beneficial when dealing with complex, high-dimensional data.

3. Scalability: Neural networks possess the remarkable ability to scale seamlessly with burgeoning data volumes and computational resources. Leveraging techniques like distributed training and hardware accelerators such as GPUs and TPUs, they effortlessly tackle mammoth datasets and intricate models.

Model Evaluation:

Model evolution refers to the process of iteratively improving a machine-learning model over time. It involves refining the model architecture, adjusting hyperparameters, and updating the model's parameters based on new data or insights gained from previous iterations. The model is evaluated using metrics relevant to the problem at hand like accuracy, precision, recall, F1-score, etc. The evaluation helps in identifying areas where the model can improve.

Cross Validation: The accuracies of all machine learning models were compared with a set of the data not used in the training called cross validation set. The accuracies came out to be:
SVM:93.85%
Random forest:95.38%
Neural Network:92.31%

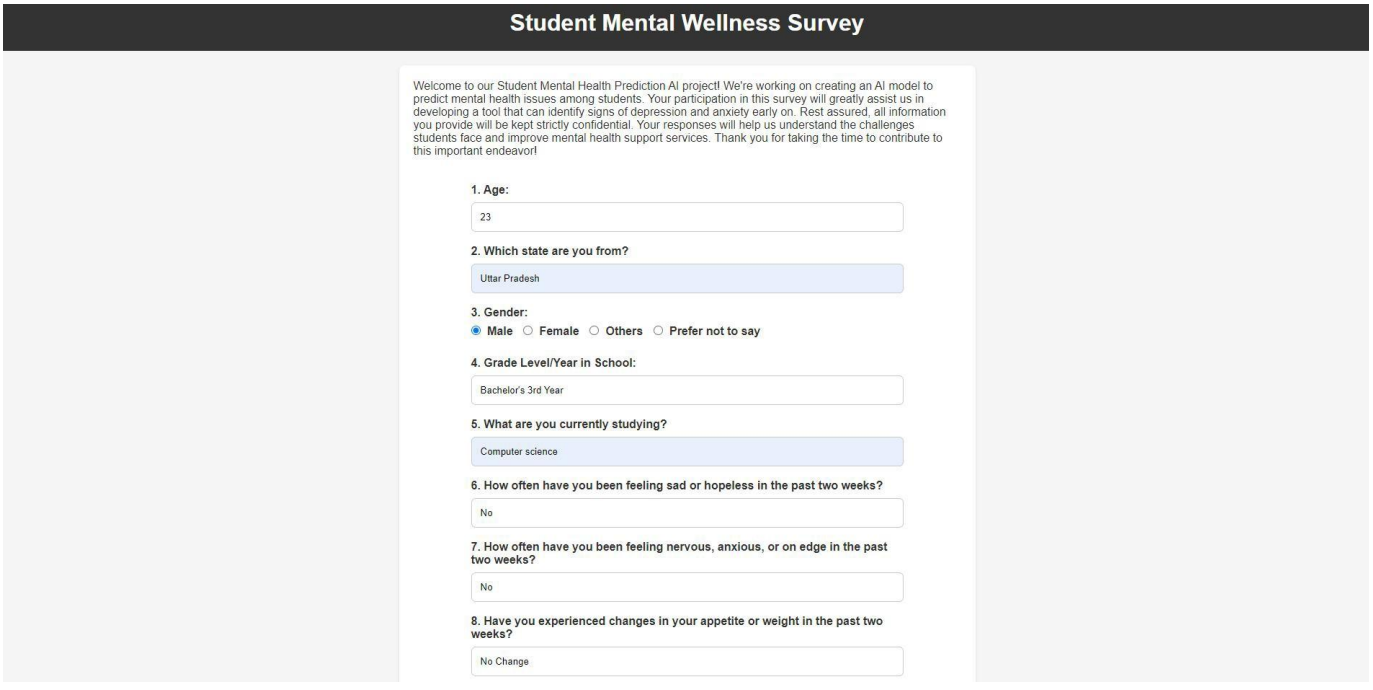
As we can see that the accuracy of random forest for this set of data and on the cross validation set came out to be the highest, we used random forest model to make our app and our model of choice.

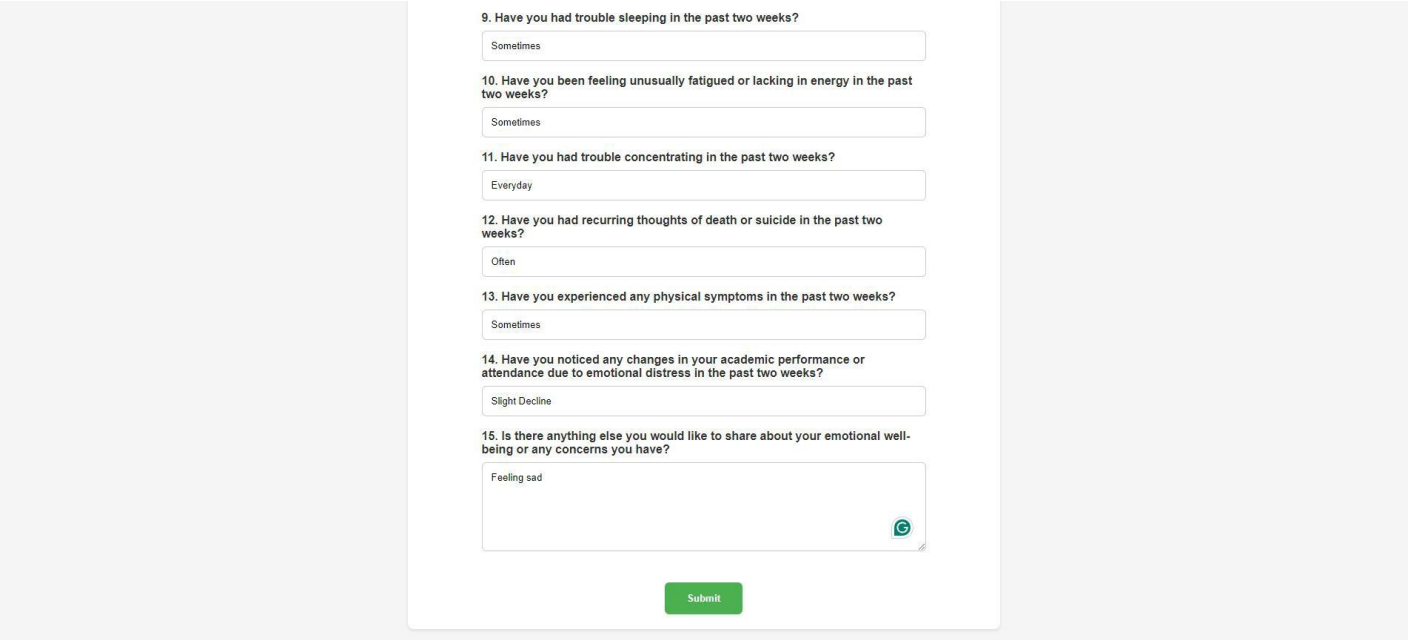
Deployment:

Deploying a model involves making it accessible and operational in a production environment, where it can serve prediction or perform its intended task. For this model, we have used Flask as our platform to deploy and test the model locally.

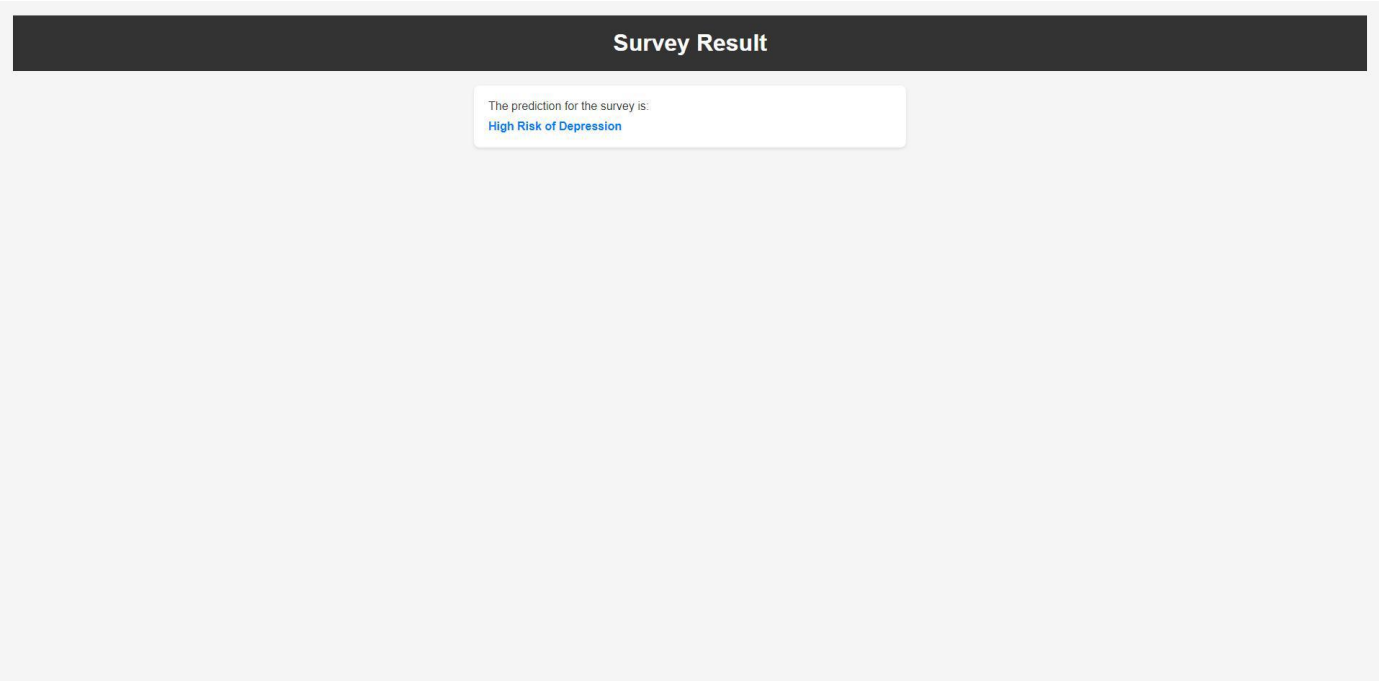
Web Application Development: An interface was created for the users to input data using Flask as a web application, which would be processed to make prediction using the trained models and to display the results produced.

Main Page




9. Have you had trouble sleeping in the past two weeks?
Sometimes
10. Have you been feeling unusually fatigued or lacking in energy in the past two weeks?
Sometimes
11. Have you had trouble concentrating in the past two weeks?
Everyday
12. Have you had recurring thoughts of death or suicide in the past two weeks?
Often
13. Have you experienced any physical symptoms in the past two weeks?
Sometimes
14. Have you noticed any changes in your academic performance or attendance due to emotional distress in the past two weeks?
Slight Decline
15. Is there anything else you would like to share about your emotional well-being or any concerns you have?
Feeling sad
Submit

Result:


Survey Result
The prediction for the survey is:
High Risk of Depression

5.2 Testing OR Verification Plan

Test cases were developed to verify the performance of the implemented models. Each test case includes a Test ID, Test Case Title, Test Condition, System Behaviour, and Expected Result.

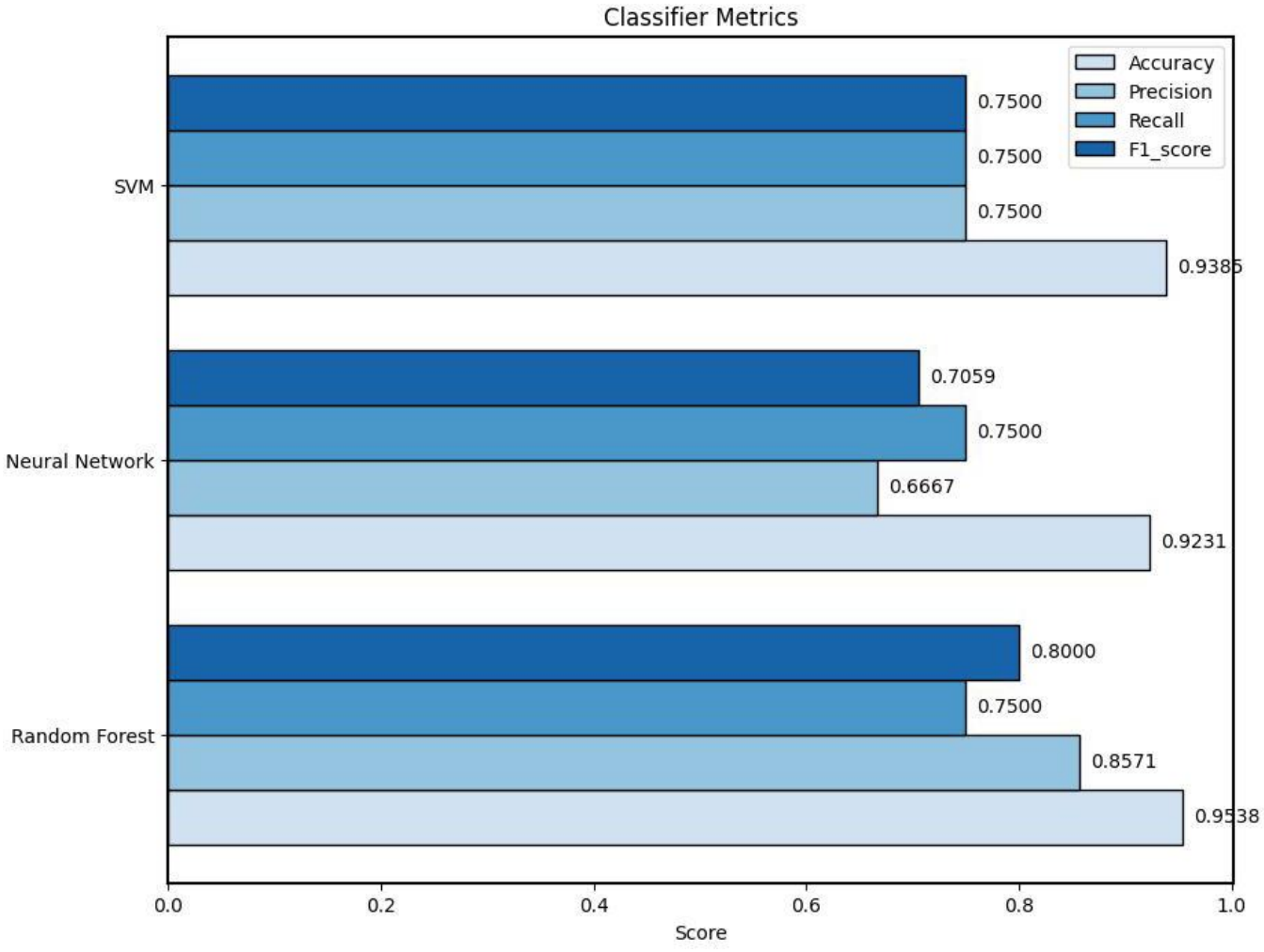
Test ID	Test Case Title	Test Condition	System Behaviour	Expected Result
T01	Data Preprocessing Test	Valid Input Data	Data is cleaned and encoded properly	Pre-processed data is ready for model training
T02	Model Training Test	Training data available	Models are trained successfully	Trained models are ready for prediction
T03	Web Application Test	Input data provided through web interface	Predictions are generated correctly	Predictions are displayed to the user

5.3 Result Analysis:

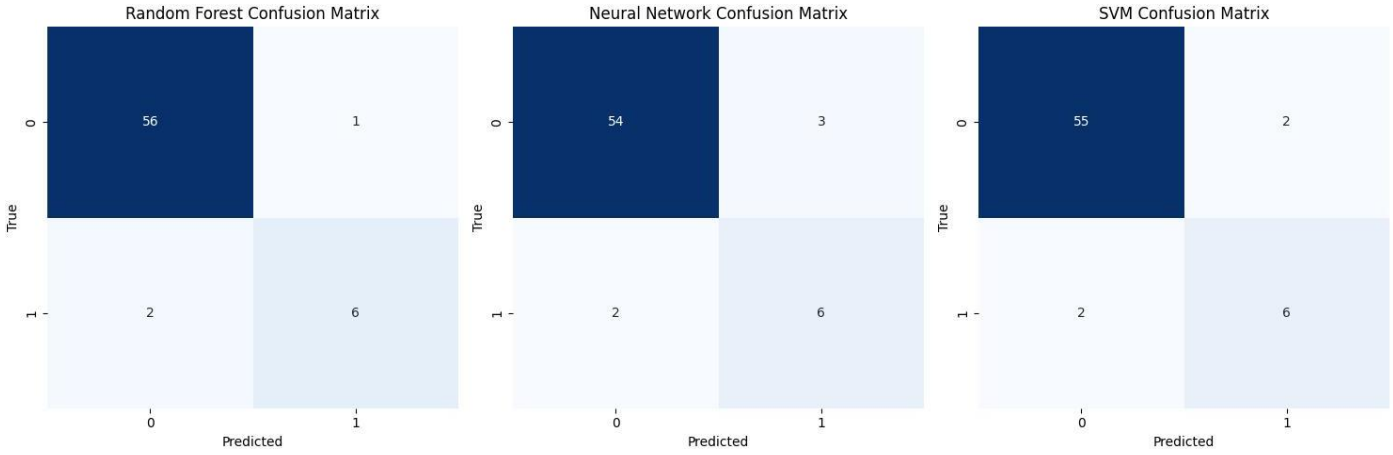
- Accuracy: This represents the overall correctness of the machine learning model's predictions. This indicates the percentage of instances which were classified correctly. The higher the accuracy the more reliable the prediction of depression risk.
- Precision: This measures the ratio of true positive predictions among all the present positive predictions. It indicates the model's ability to correctly identify the instances which are at true risk. The higher the precision the fewer the false positive predictions will be, which thereby reduces the chance of misclassifying the instances.
- Recall: Also known as sensitivity, which measures the ratio of true positive predictions to those with the actual positive instances in the dataset. It indicates the model's ability to capture all instances who are at true risk and thereby minimise the chances of false negatives.
- F1 Score: This is the harmonic mean between precision and recall. It helps in providing a balanced measure of implemented model's performance. It therefore considers both

false positives and false negatives and is very useful when the distribution of class is imbalanced. It thereby reflects the overall effectiveness in identifying the instances which are at risk and those which are not at risk by considering the threshold between precision and recall.

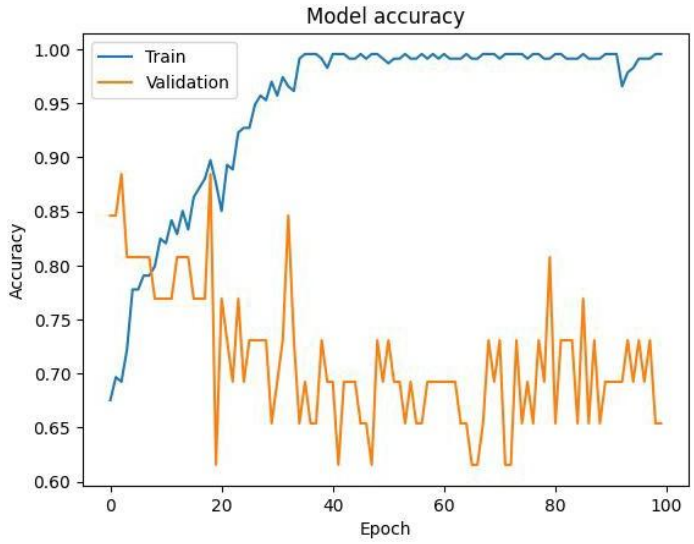
In this project Random Forest model presents by far the most superior performance over all the metrics and features present in the collected data. We can say that it more accurately predicts the depression risk compared to SVM and Neural Network Models implemented on same dataset. However, the choice of correct and best possible model also depends upon different requirements, constraints applied and priority level. For instance, if maximizing the precision i.e. reducing false positives is critical to avoid unnecessary interventions or treatments, then a model with high precision might be preferred, even if it sacrifices some recall. Conversely, if identifying as many individuals at-risk as possible i.e. high recall is therefore will be needed to ensure comprehensive screening and support, then a model with high recall might be favoured, even if it leads to more false positives.

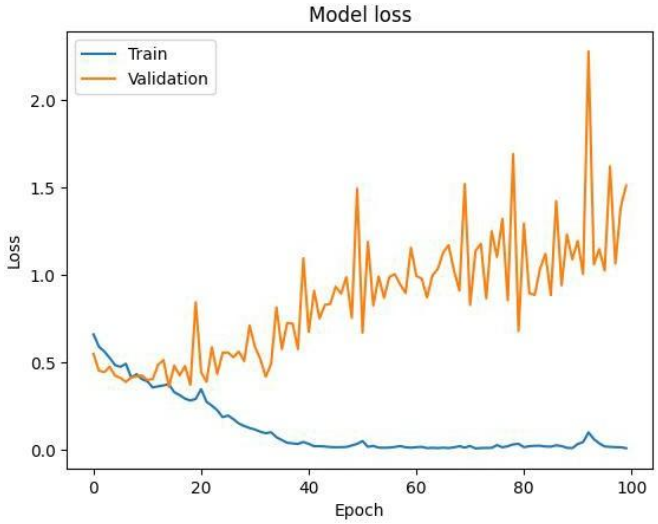


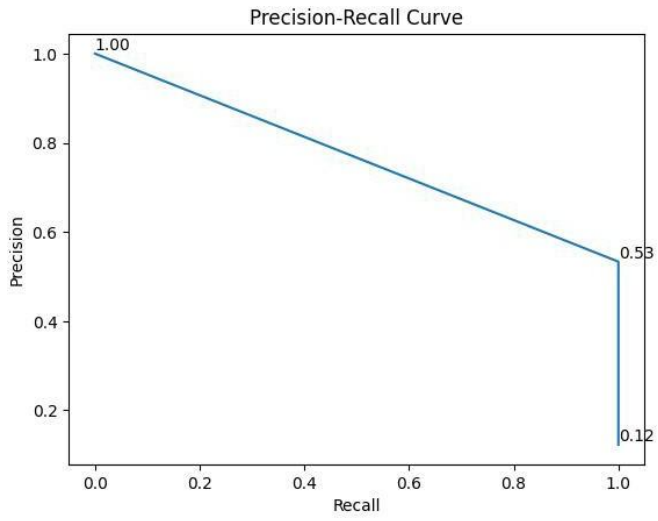
Confusion Matrix of all Model

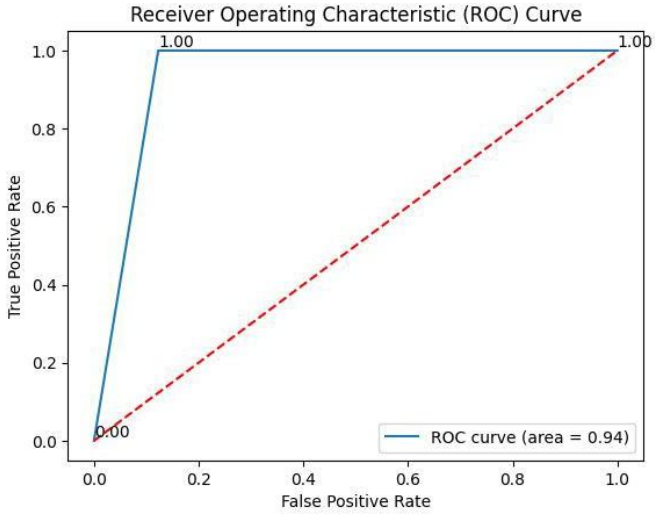


Neural Network Performance

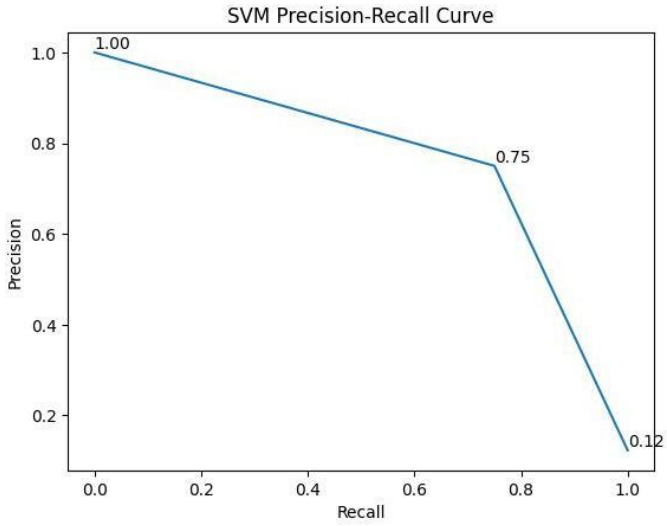


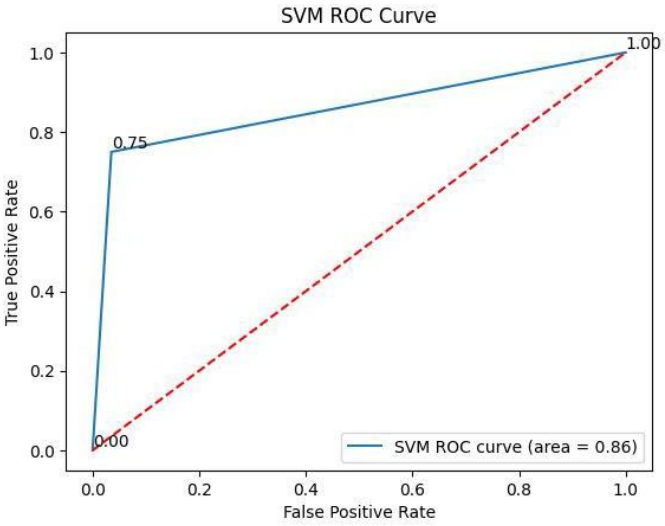




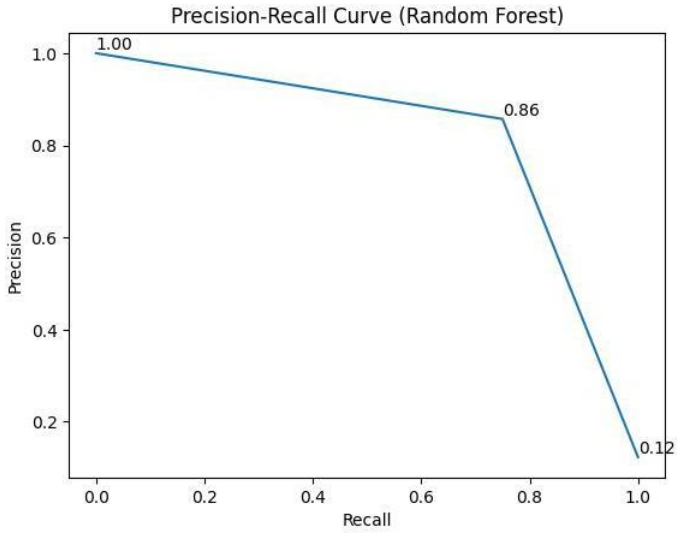


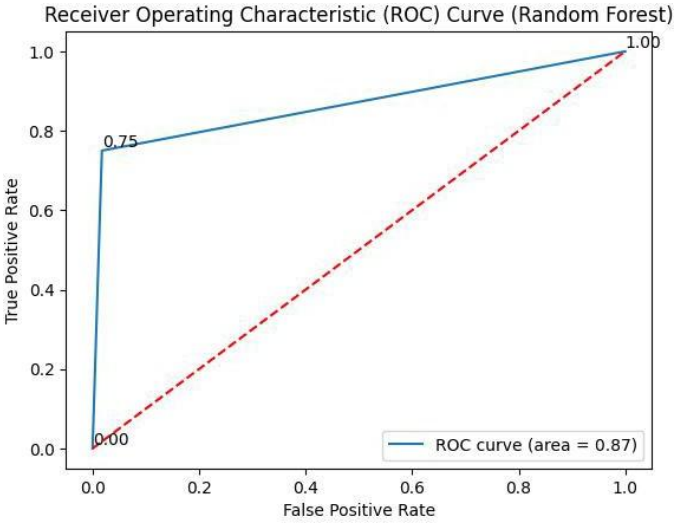
SVM Performance





Random Forest Performance





Chapter 6

Standards Adopted

6.1 Coding Standards

Here are some coding standards followed in this project:

1. Naming Conventions: All variables, classes and functions were named using clear names which describe their use to make the code more readable.

2. Indentation: Proper Indentation used to make the code more readable and to clearly divide and identify the code blocks.

3. Comments: To explain the purpose of the code Comments were added throughout the code and to make it easier for other developer to easily understand the code.

4. Code Reusability: Useful reusable block of codes was identified and grouped into separate classes or functions to reduce redundant code and increase modularity.

5. Error Handling: Suitable error handling mechanisms were implemented throughout the code to handle any error or exception that might occur at runtime.

6. Code Optimization: To reduce excess memory usage and execution time the code is optimized wherever possible.

Chapter 7

Discussion

7.1 CONCLUSION:

To summarize, the model developed underscores the importance of understanding and addressing student mental well-being. By leveraging machine learning techniques and sentiment analysis, we've crafted a predictive framework to examine survey data and identify patterns influencing students' mental health.

We have initiated the process by first circulating a comprehensive survey form among students, inviting them to candidly share their experiences, emotions and thoughts and choose the best options which define their current mental state. Through the implementation of Random Forest we have been successful in establishing a robust framework for our project which is capable of diagnosing the students mental health with much higher accuracy. This approach has not only ensures the authenticity and richness of the data collected but also captures the diverse array of perspectives and insights that might go unseen.

Our Utilization of different technologies for both backend and frontend design gives a seamless experience to users who can access and interact with the said diagnostic tool. We have meticulously employed preprocessing techniques in addition to feature engineering, with these we have been able to enhance the model’s reliability and efficacy in processing our surveyed data.

The model's ability to correctly analyze the survey responses and predicting mental wellness outcomes holds great promise for educational institutes and mental health practitioners. Moreover the interdisciplinary nature of our approach underscores the importance of collaboration and cooperation between data scientists, policymakers and psychologists. By identifying individuals who may be at risk and understanding the factors that influence their mental health problems, interventions can be targeted more precisely, resulting in improved support and resources for students.

The incorporation of sentiment analysis into our project also helps us achieve a pivotal advancement in understanding the nuanced emotions and sentiments of the users. By integrating Natural Language processing techniques we have been able successfully unlock the deeper layers of insights into student’s mental states, which thereby transcends the limitation of traditional survey analysis. With this Deep Learning Implementation we will be able to discern not the explicit content within the filled responses but also the underlying emotions and attitudes of one’s behavior.

This project is an important tool for helping students with their mental well-being. It shows how technology can help address problems in education and mental health. By persistently researching and refining our methods, we can foster a more nurturing environment where

students can thrive academically, emotionally, and socially. By maintaining the required transparency and accountability in our practices we will be able to uphold the integrity of our model and reaffirm our commitment towards advancing the student's mental health with integrity and empathy.

7.2 LIMITATIONS

The journey towards enhancing the students mental health will be an ongoing endeavor that will require the continuous integration and adaptation with newer and better practices. Therefore it is crucial that we acknowledge the possible underlying limitations:

- Bias : The data collected from the survey form may be subjected to some sampling bias due to the voluntary participation of some participants. This could hence lead to the problem of underrepresented demographic groups or even over representation of some others which will surely impact the generalizability of our findings.
- Quality : The accuracy and required reliability of the data may be influenced by some factors such as social desirability bias or memory recall issues of the participants. Also some of them may be reluctant to give such sensitive information. This will then potentially introduce the noise or accuracy issues into our model's predictions.
- Limited scope of features: Although we have employed the required preprocessing methods and feature engineering techniques, there still may be some additional variables or other factors that could not be captured from the dataset which could possibly influence in diagnosing the students mental health status.
- Interpretation done by sentiment analysis: The algorithm used for sentiment analysis may not always be accurate to capture the emotions and context expressed in the text data. Ambiguities present in different languages, or sarcasm used or the other cultural aspects may lead to misinterpretation of the sentiment analysis result.
- Ethical Considerations**:** There are many ethical and general considerations surrounding the use of predictive model for analysis which might include privacy concerns, potential stigmatization of individual participants who have been identified having certain level of risk and the possible responsibility to provide appropriate support and resources to those in need.
- Generalization**:** The model's performance may vary when applied to a different population or institution or any cultural contexts. Factors such as policies being followed in a particular institution, availability of support services and other socio-cultural factors which influence the effectiveness of our model.
- Dynamic Nature of Mental Health: Mental health is inherently dynamic and multifaceted, influenced by a myriad of individual, interpersonal, and environmental factors. Our model's static analysis of survey data may overlook the complex interplay of these factors over time, limiting its ability to capture the full spectrum of students' mental health experiences and trajectories

7.3 FUTURE SCOPE

Looking forward into the horizon of this project, it surely reveals a landscape full of possibilities and opportunities for advancement. As we move ahead we envision that our understanding of the student's well being will be continually pushed in future which will pave the way for more innovative solutions and transformative interventions.

One of the very promising avenues for future exploration lies in the significant refinement and continuous development and integration of predictive models. By honing the novel techniques of models which are equipped with advanced machine learning methods and ensemble approaches could potentially offer more targeted support to those who are in need.

Moreover the integration of diverse data represents a frontier for further explorations in this field. Beyond traditional survey responses, researchers are also exploring the integration of multimodal data, including social media activities, smartphone usage patterns, biometric indicators, and academic performance metrics. By weaving together these diverse strands of information, researchers are hoping to paint a better picture of students' mental well-being, capturing the dynamic interplay of factors that influence their mental health trajectories.

In parallel, the development of real-time monitoring systems also offers the promise of proactive support. Using wearable devices, smartphone applications, and new sensing technologies, educators and mental health professionals may gain timely insights into students' well-being and successfully deliver personalized interventions in real-time which can potentially be helpful in averting crises and fostering resilience.

Personalization also emerges as a key theme in future directions for this project. By tailoring interventions and support resources to the individual needs and preferences of students, institutions can thereby enhance engagement and effectiveness. Machine learning-driven recommendation systems hold the potential to deliver targeted resources and services uniquely designed for each student, which then amplifies the impact of interventions.

In the journey ahead, collaborative partnerships between researchers, educators, policymakers, and technology developers will be crucial for driving innovation and translating research findings into actionable strategies. By fostering these interdisciplinary collaboration and also community engagement, we can collectively work towards a future where every student has the opportunity to thrive academically, emotionally, and socially.

References

[1] Aleem S, Huda Nu, Amin R, Khalid S, Alshamrani SS, Alshehri A. Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions. Electronics. 2022; 11(7):1111.https://doi.org/10.3390/electronics11071111

[2] Sumathi, M., & B., D. (2016). Prediction of Mental Health Problems Among Children Using Machine Learning Techniques. International Journal of Advanced Computer Science and Applications, 7(1). https://doi.org/10.14569/ijacsa.2016.070176

[3] Ahmed, A., Sultana, R., Ullas, M. T. R., Begom, M., Rahi, M. M. I., & Alam, M. A. (2020, December 16). A Machine Learning Approach to detect Depression and Anxiety using Supervised Learning. 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). https://doi.org/10.1109/csde50874.2020.9411642

[4] Gao, Y., Flannery, M., Assan, J., Wu, Y., & Resom, A. (2019). Machine Learning for Mental Health Detection. : Worcester Polytechnic Institute.

[5] Chung, J., & Teo, J. (2022, January 5). Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. Applied Computational Intelligence and Soft Computing, 2022, 1–19. https://doi.org/10.1155/2022/9970363

[6] Sajja, G. S. (2021, December 22). Machine Learning based Detection of Depression and Anxiety. International Journal of Computer Applications, 183(45), 20–23.

[7] Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016, March). Cross-trial prediction of treatment outcome in depression: a machine learning approach. The Lancet Psychiatry, 3(3), 243–250. https://doi.org/10.1016/s2215-0366(15)00471-x

[8] Sau, A., & Bhakta, I. (2019). Screening of anxiety and depression among seafarers using machine learning technology. Informatics in Medicine Unlocked, 16, 100228. https://doi.org/10.1016/j.imu.2019.100228

[9] Lahiri Mallik, Soma & Rathore, Nirmala Singh & Jagawat, Dr. (2023). DEPRESSION AMONG COLLEGE GOING STUDENTS: AN EVALUATIVE STUDY. Seybold Report. 18. 420-430. 10.5281/zenodo.8255225.

[10] Alharbi, R., Alsuhaibani, K., Almarshad, A., & Alyahya, A. (2019). Depression and anxiety among high school student at Qassim Region. Journal of Family Medicine and Primary Care, 8(2), 504. https://doi.org/10.4103/jfmpc.jfmpc_383_18

[11] , E., & Buscema, M. (2007, December). Introduction to artificial neural networks. European Journal of Gastroenterology & Hepatology, 19(12), 1046–1054. https://doi.org/10.1097/meg.0b013e3282f198a0

[12] Hearst, M., Dumais, S., Osuna, E., Platt, J., & Scholkopf, B. (1998, July). Support vector machines. IEEE Intelligent Systems and Their Applications, 13(4), 18–28. https://doi.org/10.1109/5254.708428

[13] Hosameldin Ahmed; Asoke K. Nandi, "Decision Trees and Random Forests," in Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines , IEEE, 2019, pp.199-224, doi: 10.1002/9781119544678.ch10

INDIVIDUAL CONTRIBUTION

STUDENT MENTAL HEALTH ANALYSIS

RIYA SINGH
21052269

Abstract: Python, Jupyter Notebook, HTML, and CSS are used in the "Student Mental Health Analysis" project to effectively analyze the mental health of students. For data processing, visualization, and model training, it uses Random Forest and other machine learning models that are integrated with Pandas, Matplotlib, and Sklearn. Users can enter information via a form on the interactive frontend, which is Flask-powered, and receive predictions about their mental health. This research serves as an example of how machine learning may be used to better understand and support students' mental health issues so they can succeed academically.

Individual contribution and findings: I collaborated with Ananya and helped with drafting the Methodology/Proposal of the project. I also participated actively in the testing and training of the model, testing various techniques, including random forest. Models were optimized for performance, and their accuracy was assessed. provided information that demonstrated the efficacy of our method in the project report's chapters on model training and evaluation.

Individual contribution to project report preparation: Played an important role in creating the project report, with an emphasis on the model training and evaluation portions, specifically working on the data collection segment. I documented the data collection and model selection criteria for our proposed model.

Individual contribution for project presentation and demonstration: Contributed to the project presentation by demonstrating the model training and evaluation process, emphasizing the effectiveness of Random Forest in mental health analysis. Demonstrated how model selection and fine-tuning contribute to accurate mental health predictions.

Full Signature of Supervisor:

…….

Full signature of the student:

……..

INDIVIDUAL CONTRIBUTION

STUDENT MENTAL HEALTH ANALYSIS

ANANYA THAKUR
21052225

Abstract: Python, Jupyter Notebook, HTML, and CSS are used in the "Student Mental Health Analysis" project to effectively analyze the mental health of students. For data processing, visualization, and model training, it uses Random Forest and other machine learning models that are integrated with Pandas, Matplotlib, and Sklearn. Users can enter information via a form on the interactive frontend, which is Flask-powered, and receive predictions about their mental health. This research serves as an example of how machine learning may be used to better understand and support students' mental health issues so they can succeed academically.

Individual contribution and findings: I led model training and testing operations of the project, with an emphasis on parameter optimisation and performance evaluation. Conducted thorough studies to compare the effectiveness of several machine learning algorithms, including Random Forest. Contributed to the Random Forest model's selection and fine-tuning using experimental and validation data.

Individual contribution to project report preparation: I contributed considerably to the project report by detailing the model training and testing procedures employed, with a particular emphasis on the conclusion and future scope parts. Provided insights on potential future developments, such as adding real-time data streams or increasing the dataset to improve predictive skills.

Individual contribution for project presentation and demonstration: Contributed to the project presentation by demonstrating model training and testing techniques, with a focus on model selection and implementation. I explained the technical components of model evaluation and demonstrated the performance of our predictive system.

Full Signature of Supervisor: Full signature of the student:

……. ……..

INDIVIDUAL CONTRIBUTION

STUDENT MENTAL HEALTH ANALYSIS

ANUSHKA SINGH
21052144

Abstract: Python, Jupyter Notebook, HTML, and CSS are used in the "Student Mental Health Analysis" project to effectively analyse the mental health of students. For data processing, visualization, and model training, it uses Random Forest and other machine learning models that are integrated with Pandas, Matplotlib, and Sklearn. Users can enter information via a form on the interactive frontend, which is Flask-powered, and receive predictions about their mental health. This research serves as an example of how machine learning may be used to better understand and support students' mental health issues so they can succeed academically.

Individual contribution and findings: Collaborated with Bibek on data preprocessing tasks, ensuring the dataset was fit for analysis and prediction. Implemented various techniques to handle missing values, performed One-Hot encoding, and performed feature extraction.

Individual contribution to project report preparation: Contributed significantly to the project report's preparation by working on parts such as data pre-processing, SRS (Software Requirements Specification) documentation, and report completion. Documented the data pre-processing procedures and their impact on model performance. Provided insights into the SRS paperwork, which outlined the project's needs and objectives.

Individual contribution for project presentation and demonstration: Contributed to project presentation by presenting scope and objectives based on SRS documentation, emphasising the importance of data pre-processing for accurate predictions in student mental health analysis. Highlighted key findings on data pre-processing and its impact on model training outcomes, emphasising the importance of clean data in developing a robust machine learning solution.

Full Signature of Supervisor: Full signature of the student:

……. ……..

INDIVIDUAL CONTRIBUTION

STUDENT MENTAL HEALTH ANALYSIS

TANMAY PANDEY
21052231

Abstract: Python, Jupyter Notebook, HTML, and CSS are used in the "Student Mental Health Analysis" project to effectively analyse the mental health of students. For data processing, visualization, and model training, it uses Random Forest and other machine learning models that are integrated with Pandas, Matplotlib, and Sklearn. Users can enter information via a form on the interactive frontend, which is Flask-powered, and receive predictions about their mental health. This research serves as an example of how machine learning may be used to better understand and support students' mental health issues so they can succeed academically.

Individual contribution and findings: I led the project's implementation and deployment phases, with an emphasis on combining machine learning models with Flask for backend development. Connected frontend interfaces with backend activities to provide a smooth user experience. While deploying the mental health analysis tool, emphasis was placed on accessibility and performance optimisation.

Individual contribution to project report preparation: My contribution to the project report includes documentation of the implementation and deployment methodologies, as well as insights into configuring the Flask server and optimising the system for user interaction. Described the technical elements of integrating machine learning models with web technologies, including the problems and solutions identified during implementation.

Individual contribution for project presentation and demonstration: I prepared and gave the project presentation, which showcased our system's live deployment. I demonstrated the frontend UI and discussed how users may interact with the system to get mental health forecasts.

Full Signature of Supervisor: Full signature of the student:

……. ……..

INDIVIDUAL CONTRIBUTION

STUDENT MENTAL HEALTH ANALYSIS

BIBEK RANJAN SAHOO
21052244

Abstract: Python, Jupyter Notebook, HTML, and CSS are used in the "Student Mental Health Analysis" project to effectively analyse the mental health of students. For data processing, visualization, and model training, it uses Random Forest and other machine learning models that are integrated with Pandas, Matplotlib, and Sklearn. Users can enter information via a form on the interactive frontend, which is Flask-powered, and receive predictions about their mental health. This research serves as an example of how machine learning may be used to better understand and support students' mental health issues so they can succeed academically.

Individual contribution and findings: Collaborated with Anushka to perform data pre-processing, including organizing and cleaning the dataset for analysis. To prepare the dataset for model training, data manipulation techniques were used, along with Python libraries. Contributed to analysing several algorithms for model selection, stressing the efficiency of Random Forest for student mental health analysis.

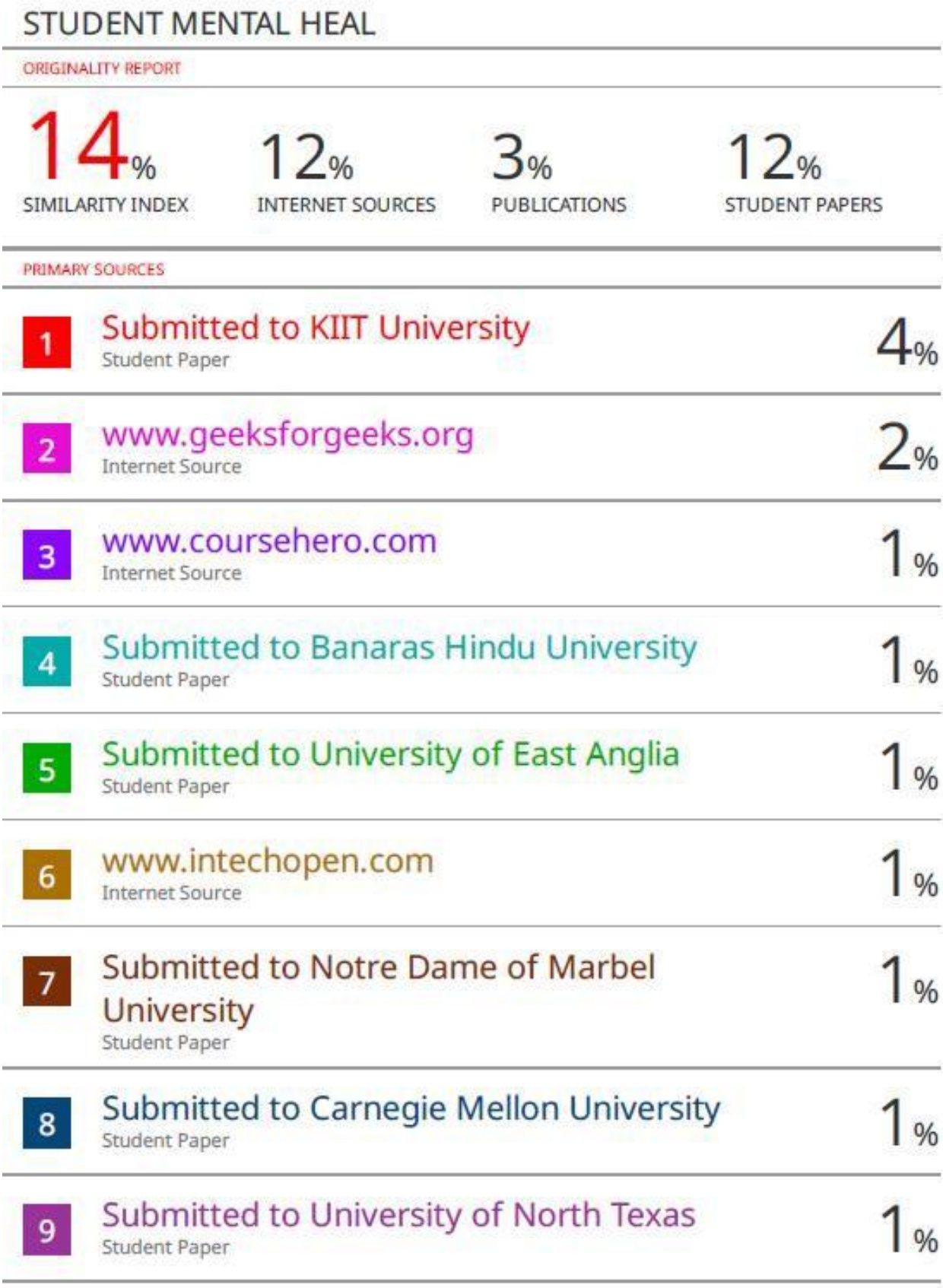
Individual contribution to project report preparation: Helped in documenting data pre-processing approaches and their influence on model performance in the project report. Provided extensive descriptions of the data cleaning and transformation methods, demonstrating their importance in increasing model correctness. Contributed insights to the comparative study of several algorithms, as well as the reasoning for using Random Forest as the key predictive model for student mental health analysis.

Individual contribution for project presentation and demonstration: Presented the data pre-processing methodologies used during the project presentation, emphasising the significance of data quality in generating consistent machine learning results. Described the precise approaches utilised to address data preparation issues and optimise the dataset for model training.

Full Signature of Supervisor: Full signature of the student:

……. ……..

Plagiarism Report



STUDENT MENTAL HEAL

ORIGINALITY REPORT

14%	12%	3%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to KIIT University Student Paper	4%
2	www.geeksforgeeks.org Internet Source	2%
3	www.coursehero.com Internet Source	1%
4	Submitted to Banaras Hindu University Student Paper	1%
5	Submitted to University of East Anglia Student Paper	1%
6	www.intechopen.com Internet Source	1%
7	Submitted to Notre Dame of Marbel University Student Paper	1%
8	Submitted to Carnegie Mellon University Student Paper	1%
9	Submitted to University of North Texas Student Paper	1%

10	Submitted to Universiteit van Amsterdam Student Paper	1%
11	www.testmagzine.biz Internet Source	<1%
12	"Applications of Computational Methods in Manufacturing and Product Design", Springer Science and Business Media LLC, 2022 Publication	<1%
13	www.frontiersin.org Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		