

**Name of Faculty: Dr Juhi Jain**

**Student Name: Survepalli Sreekruti**

**Roll No: A2305220266**

**AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY**

**AMITY UNIVERSITY UTTAR PRADESH**

**NOIDA**



**Predicting tech workers mental health condition using machine learning techniques**



**Declaration**

I, **Survepalli Sreekruti**, student of B.Tech (6CSE1-X) hereby declare that the research paper on **“Predicting tech workers mental health condition using machine learning algorithms”**,which is submitted by me to **Department of Computer Science & Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh**, in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering, & has not been previously submitted for the basis of the award in any other degree, diploma or any other recognition or title

DATE:

Survepalli Sreekruti

A2305220266

5CSE1X (2020-24)

**certificate**

This is to certify that **Ms. Survepalli Sreekruti**, student of B. Tech in Computer Science & Engineering, has carried out work on project entitled “**Predicting tech workers mental health condition using machine learning algorithms**” as a part of her 6th year Program of Bachelor of Technology in Computer Science & Engineering from Amity University Noida,under my supervision.

I Dr. Juhi Jain

iDepartment of Computer Science & iEngineering

iASET

iNoida, Uttar Pradesh

**ACHNOWLEDGEMENT**

The completion of this project would be incomplete without the mention of people whose constant guidance has to be awarded with my success.

I would like to express my deep and sincere gratitude to my research supervisor, **Dr Juhi Jain,**for giving me the opportunity to do research on topic **“Predicting tech workers mental health condition using machine learning algorithms”** and providing invaluable guidance throughout this research. It was a great privilege and honour to work and study under her guidance. I would like to thank her for solving my queries and also guide me through out.

**index**

|  |  |  |
| --- | --- | --- |
| **S. NO.** | **CONTENT** | **PAGE NO.** |
| 1. | **INTRODUCTION**   * 1. Abstract   2. Data Ethics Statement   3. Introduction   4. Scientific & Social Relevance   5. Research Questions |  |
| 2. | **CONCEPTUAL FRAMEWORK**  2.1 Background  2.2 Literature Survey  2.3 Contribution  2.4 Feature Importance In Mental Health  Prediction |  |
| 3. | **METHODOLOGY**  3.1 Methods Used  3.2 Proposed System |  |
| 4. | Presentation of Data, Analysis and  Findings  Presentation of Data, Analysis and  Findings  **PRESENTATION OF DATA, ANALYSIS & FINDINGS**  4.1 Dataset Used  4.2 Data Cleaning & Preprocessing  4.3 Procedure  4.4 Evaluation Criteria  4.5 Software & Algorithms  GDP |  |
| 5. | **RESULTS** 5.1 Model Performance 5.2 Analysis  5.3 Feature Importance  Conclusion , Recommendations and  Bibliography  Conclusion , Recommendations and  Bibliography  Conclusion , Recommendations and  Bibliography |  |
| 6. | **DISCUSSION & CRITICAL ANALYSIS**  6.1 Questions & Answers  6.2 Future Work |  |
| 7. | **CONCLUSION & REFERENCES**  7.1 Conclusion  7.2 References |  |

**Introduction**

* 1. **ABSTRACT**

The state of one's mental health has a significant impact on their capacity to influence society. The evaluation of an employee's mental health in the workplace is a challenging task because it may find it difficult to tell their employers about their illnesses. The Open Sourcing Mental Illness (OSMI) data collection is used to examine how machine learning may be used to forecast mental health disorders among those working in the IT industry.

The rise in mental health issues and the demand for high-quality medical care have prompted research into the use of machine learning in mental health issues. In order to forecast mental health issues, this research presents a current thorough assessment of machine learning algorithms. We will also go over the difficulties, restrictions, and future directions for using machine learning in the field of mental health. By exploring dependable databases, we gather studies and publications on machine learning methods for forecasting mental health issues.

The performance of **six machine learning** **models** is examined using logistic regression as the baseline and the f1 score as the primary indicator. Comparing **XGBoost to Logistic Regression, Support Vector Machine, Decision Tree, ADABoost and Random Forest**, it is discovered that it displays the highest f1 score. Additionally, XGBoost exhibits the highest level of performance stability. The most important elements in predicting mental health disorders are identified through feature importance analysis.

The Results show that Work interference has the largest contribution. Whether the employee's mental health issues interfering with the work is the thing that the company should ask for its employees. Then Family history and care options(programs and benfits) provided by company is also influential in employees who want to get treatment.

For all the remaining features, there has been a little contribution.Noticing/knowing some of these features beforehand can even help support an individual who may be experiencing a mental health issues and and connect them with the appropriate employee resources.

Companies could use the models created from this thesis to get a glimpse of the conditions affecting their employees' mental health. Additionally, businesses could construct intervention and evaluation prioritisation based on feature importance analysis.

Additionally, we provide concrete recommendations on the potential future research and development of applying machine learning in the mental health field.

* 1. **DATA ETHICS STATEMENT**

There was no data collection on humans or animals used in this thesis. The author of this thesis admits they have no legal claim to the information.

* 1. **INTRODUCTION**

The state of a person's health is inextricably linked to how well they are able to perform on a daily basis. In addition to physical health, mental health is also important in a person's existence. According to the WHO (2008), "Mental health is a state of well-being in which a person realizes his or her abilities, can cope with everyday stresses, can work productively, and can contribute to his or her community." This remark suggests a strong connection between a person's ability to work and have an impact on society and their mental health.

Undoubtedly, a person's emotions, intellect, and ability to communicate with others are all affected by mental illness, which is a health issue. These problems have demonstrated that mental illness has major societal repercussions and necessitates novel prevention and therapeutic measures. Early mental health detection is a crucial step in implementing these techniques. According to Miner et al, medical predictive analytics will fundamentally alter the healthcare industry. The typical method for diagnosing mental disease is based on the patient's self-report, which calls for the use of questionnaires created to identify particular emotional or social interaction patterns. Many people with mental illness or an emotional condition should be able to recover with the right care and treatment.

Major mental health disorders, like chronic illnesses, bipolar disorder, and schizophrenia, can develop gradually over time and exhibit signs that can be identified in the beginning stages. Such conditions could be prevented or better regulated. If anomalous mental states are identified early on in the disease's progression, additional attention and therapy can be given. Therefore, interpreting someone's mental state based on their appearance or behavior requires a complex psychological science that has not yet been mechanized.

The number of people with mental disorders has increased, especially among those who are employed, and their conditions have a negative impact on both their health and work performance (Awal & Rao, 2021). It follows that when a disruption occurs in an employee's life, it has an impact on both their personal and professional lives. Therefore, it is important to address the research on employee mental health since doing so would be advantageous for both the individual and the business.

By providing various forms of support and benefits, many businesses have begun to pay more attention to the mental health or overall wellbeing of their employees. For instance, Google has provided mental health-related movies, counselling, psychologist help, and remote working options (Elias, 2020).To effectively address the problem, a business or organization must first determine the state of its members' or employees' mental health.

Additionally, recognising signs of declining mental health is one of the ways managers may help their staff members reach their full potential (Dimoff & Kelloway, 2019).

In situations like this, machine learning would be extremely useful since it could be used to create models that could forecast employees' mental health conditions without asking them directly.

This would be the thesis' primary goal. Additionally, machine learning algorithms could determine which elements are most crucial.

Using sophisticated statistical and probabilistic techniques, machine learning tries to create systems that can learn from experience. It is thought to be a very helpful tool for predicting mental health. It is enabling a number of academics to derive crucial knowledge from the data, offer individualized experiences, and create automated intelligent systems.

The widely used algorithms in the field of machine learning such as support vector machine, random forest, and artificial neural networks have been utilized to forecast and categorize the future events.

The computer business is recognized for its high levels of pressure and arbitrary deadlines, hence this thesis will concentrate on the employees who work in it. Additionally, the number of technological firms has been increasing rapidly. A number of research have been conducted to forecast the mental health status of IT employees. However, the way in which researchers defined mental health issues and the algorithms they employed varied across studies.

This essay will also suggest potential directions for further study on the subject. Additionally, it would highlight the difficulties and restrictions associated with using machine learning techniques in this context. In addition, the future directions and gaps in this area of study will be highlighted. As a result, this research will advance the state of the art by doing a thorough literature assessment of the machine learning methods used to forecast mental health issues. In order to help researchers learn more about the techniques and applications of big data in the disciplines of mental health, this publication presents a critical summary and future research directions.

For instance, Awal and Rao (2021) used “Whether an employee had an attempt in seeking help from a mental health professional” as the definition of mental health, and algorithms that were used were AdaBoost Classifier (ADA), Decision Tree, and Extreme Gradient Boost Classifier (XGB). Another study (Dharma et al., 2021) used “Do you currently have a mental health disorder?” as the ground truth and used Support Vector Machine (SVM) and Fully Connected Neural Network (FCNN).

* 1. **SCIENTIFIC AND SOCIETAL RELEVANCE**

The performance of various machine learning algorithms and the most important variables in predicting mental health conditions will be examined from a scientific standpoint in this thesis. These variables have never been used or implemented in the previous study in predicting the corresponding target attribute. This dissertation would contribute to the rapidly expanding field of mental health.

From a societal standpoint, this thesis would assist businesses and organizations in getting a general sense of the mental health of their staff and taking decisions based on that evidence. Additionally, workers might learn about the variables influencing their situations from the study's findings.

* 1. **RESEARCH QUESTIONS**

The following research question captures the major goal of this thesis:

***‘To what extent can we predict tech workers’ mental health condition using machine learning techniques?’***

Two sub-questions will be used to address this research question:

***RQ1 “Which machine learning technique(s) provide good performance in predicting tech workers’ mental health conditions?”***

***RQ2 “What are the contributing factors in predicting tech workers’ mental health conditions?”***

Algorithms that had demonstrated strong performance in predicting similar target properties in earlier studies will be utilized to address the first sub-question, and their performance will be compared.

Later, in the section on related work and methodology, the findings of earlier studies and the models that were picked will be covered.

Using the top-performing model from sub question one, a feature importance analysis will be performed to provide a response to the second sub-question.

**CONCEPTUAL FRAMEWORK**

**2.1 BACKGROUND**

Figure below illustrates the conventional procedure for a systematic literature review, which is followed in this review work. The planning stage of this review article is where the research topics or objectives are looked into and established. During the planning stage, the data sources are chosen, and then the topic-related search phrases are employed in the data sources. In doing the evaluation, a number of factors need to be given higher priority.

A red hexagons with white text

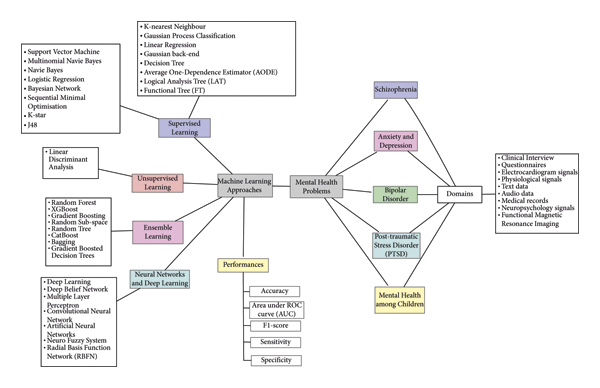
Description automatically generated

For instance, studies that answer research questions and studies connected to the topic will be picked from publications of research articles or papers. In addition, the evaluation phase will start with data extraction from the selected research publications or papers. The information or proof from the chosen publications and papers will next be subject to additional investigation. There will be discussion and investigation of the research's tendencies related to the subject. The discussion and conclusion make up the final step of the procedure. In this section, the research's constraints, shortcomings, and gaps will be highlighted and examined. In addition, potential directions for the research will be looked into and identified. A conclusion will be offered in light of the research's findings.

The categorization and classification of the comprehensive review of the literature on this subject are shown in Figure below. The applications of machine learning to issues with mental health are being researched and examined. Schizophrenia, anxiety and depression, bipolar illness, post-traumatic stress disorder (PTSD), and mental health issues in children are the five categories into which mental health issues fall.

Additionally, in order to demonstrate the effectiveness of machine learning techniques in the field of mental health, the machine learning model's performances will be provided in this study.

For instance, the performances such as accuracy, the area under the ROC curve (AUC), *F*1-score, sensitivity, or specificity will be specified and mentioned in this review paper to provide further analysis.



This thesis is built on prior studies that predicted mental health issues using data science methodologies. The outcomes of pertinent studies will be discussed in this section.

This section will concentrate on previous studies that made use of the Open Source Mental Illness (OSMI) dataset's Mental Health in Tech Survey.

The dataset is frequently used by researchers, and this thesis will also use it exactly. The method section will include a full explanation of this dataset's specifics.

This dataset is from a 2014 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace.

**2.2 literature survey**

A synthesis is a re organization and rearranging of material, whereas a summary is a rehash of the key points from the source. It could provide a fresh interpretation of previous research, blend fresh and old views, or chart the field's intellectual development across time, including key controversies.

The literature review may assess the sources and advise the reader on which are most topical or relevant, depending on the circumstances.

1. According to **Hatoon AlSagri** and **Mourad Ykhlef's** article, ***"Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features,"*** they want to use machine learning to identify potential depressed users.Based on tweets and network behaviour, a Twitter user. They developed and tested classifiers to do this, using information taken from a person's network activities and tweets to determine if the user is depressed or not. The findings indicated that the accuracy and F-measure scores in identifying depressed users increased with the number of features included. This strategy uses data to make predictions about the likelihood of detecting depression or other mental diseases in their early stages.
2. ***“Predicting the Use of Mental Health Treatment with Different Machine Learning Algorithms”*** by Meera Sharma, Sonak Mahapatra and Adeethyia Shankar, stated that the widespread and pervasive stigma associated with taking care of one's mental health has prevented modern society from providing adequate diagnosis and treatment for those who suffer from mental health illnesses. In recent years, attempts to forecast suicidality (as well as other diagnostic methods) in the data science community have failed miserably.
3. “***Prediction of Mental Disorder for employees in the IT Industry”*** by Sandhya P, Mahek Kantesaria, claimed that they have collected the dataset of the questions that were given to workers in the IT sector. The conclusion is reached based on their responses. The result will be whether or not the person need attention in this case. To obtain the results, various machine learning approaches are applied. This projection also demonstrates how critical it is for IT professionals to receive frequent mental health screenings in order to monitor their wellbeing. The employers are responsible for providing medical care within their workplace and for providing benefits to the impacted employees.
4. ***“Prediction of Mental Health Problems Among Children Using Machine Learning Techniques”*** by Ms. Sumathi M.R, Dr. B. Poorna noted that 25 attributes had been determined to be essential for assessing the problems in the reports. By using feature selection techniques on the entire data set of attributes, the attributes have been reduced. The effectiveness of various machine learning approaches on both the entire attribute set and the chosen attribute set has been examined. The end findings clearly show that the three classifiers—Multilayer Perceptron, Multiclass Classifier, and LAD Tree—bring about more effective results, and there is hardly any difference in their performances throughout the entire collection of attributes and a particular set of attributes.
5. Seven machine learning models were used by ***Reddy, Thota, and Dharun (2018)*** to forecast mental health issues among IT employees. Boosting, Random Forest, Logistic Regression, KNN, Decision Tree, and Bagging were the selected algorithms. The study's main goal was to predict stress.Therefore, the results of the question "Have you ever sought treatment for a mental health disorder from a mental health professional?" specifically for stress-related disorders were compared with the predictions produced from the models. The accuracy score was selected as the primary metric. In the prediction, fourteen final features were applied. Boosting was deemed to be the best model after achieving the greatest accuracy score of 0.75. Bagging had the lowest accuracy score out of all the models, at 0.69. In their research, additional indicators were utilised to assess the model's performance. For instance, cross-validated AUC, precision, and false-positive rate. The cross-validated AUC score for Random Forest, which fared better than the other models (with a score of 0.79), was used by the researcher to support his claim that it is the model with the best level of stability. In comparison, the KNN classifier was deemed inaccurate because it had a 0.40 false-positive rate.
6. Similar machine learning techniques were employed by ***Awal and Rao (2021)***, a different team of researchers, to forecast mental health disorders. AdaBoost Classifier (ADA), Logistic Regression, K-nearest Neighbour (KNN), Bernoulli Naive-Bayes (BNB), Extreme Gradient Boost Classifier (XGBoost), Decision Tree, and Gaussian Process Classifier (GPC) were among the seven machine learning algorithms employed. The question "Has being identified as a person with a mental health issue affected your career?" and 63 final predictors served as the foundation for the target attribute. The outcomes were displayed using various evaluation indicators. For instance, its geometric mean values, accuracy score, balanced score, f1 weighted score, and so forth. Based on these metrics, the researchers concluded that the Decision Tree is the model that performs the best. Decision Tree outperformed other algorithms and displayed the highest score across a variety of evaluation measures. The AdaBoost classifier and XGBoost were the next most promising models. These three models perform significantly better than other models, with a f1 score of over 0.90 compared to 0.70-0.80 for other models.
7. ***Dharma et al (2021)*** used Support Vector Machine (SVM) and Fully Connected Neural Networks (FCNN) to predict mental health disorders, in contrast to the two research that were previously discussed. Do you currently have a mental health disorder? is the basis for the definition of a mental health condition. 42 predictors made up the models. The accuracy score—the study's primary metric—was computed by the researchers using a confusion matrix to assess the model's performance.Both SVM and FCNN demonstrated satisfactory results, according to the researchers, with accuracy scores of 0.77 and 0.76. However, based on how well it performs across labels, SVM is regarded as the best-performing model.

**2.3 contribution**

In their investigations, researchers have used a range of data science methodologies and varied classifications of mental health problems. As in **Dharma et al. (2021)'s** study, this thesis will employ the question ***"Do you currently have a mental health disorder?"*** as the basis for prediction. Only about two machine learning methods' performances were known for this target attribute. FCNN and SVM were the two models, with SVM coming out on top. Decision Tree, AdaBoost, and XGBoost models, which performed well in prior experiments, had never been utilised to forecast the selected target characteristic.

This thesis intends to assess if these top-performing algorithms from previous research will also perform well in predicting whether or not a person now has a mental condition. SVM, which was previously employed for this objective attribute, Random Forest, AdaBoost Classifier, and XGBoost are the methods selected for this thesis to achieve this goal. The starting point will be based on logistic regression.

**2.4 Feature importance**

A feature importance analysis can be done in addition to the main goal to find out how each feature affected the models that were previously created. With feature importance, it is possible to pinpoint the variables that have the most impact on forecasting a target property.The ensemble approaches were utilised by ***HENRY & ISA, 2022,*** to forecast mental health issues among IT professionals.

The researchers carried out feature importance analysis in addition to evaluating the performance of various machine learning techniques. The research's goal attribute is if the employee needs to receive therapy for their mental health.According to their research, "bad feedback to mental interference in work" had the biggest impact on the model. The employee's mental disease diagnosis is another factor that is important to the model. Both old and new diagnoses were subject to this problem.The view of mental health illnesses among coworkers and whether the firm placed a high priority on physical health were the next two standout characteristics.

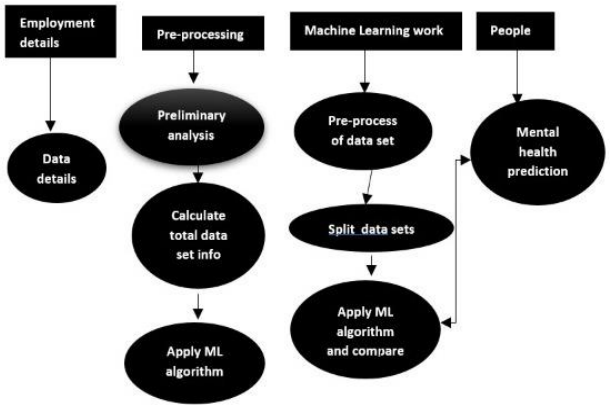
In a study aimed at predicting stress in IT workers, ***Reddy et al. (2018)*** employed Decision Trees to carry out feature significance analysis. The characteristics that contributed most to the model were gender and a family history of mental illnesses. Additionally, a number of employer-related characteristics were discovered to be crucial components. For instance, the employer's mental health benefits and how much emphasis is placed on both physical and mental health.

In previous research, feature importance analysis was used to determine the key variables that predicted stress and the need for treatment in IT workers. The goal attribute of this thesis, which is determining whether or not employees now have a mental health condition, had never been done. With the use of the machine learning method that produces the best model performance, this thesis will examine the significance of features in predicting mental health disorders.

**Methodology**

**3.1 methods used:**

This thesis will use ***Five*** machine learning techniques, which are **Logistic Regression, Decision Tree, Support Vector Machine (SVM), Random Forest,ADABoost and Extreme Gradient Boost Classifier (XGBoost)**.

****

1. ***Introduction to Decision Tree:***

The Random Forest Model, AdaBoost, and XGBoost are three decision tree-based models used in this thesis. The way the decision tree models were trained accounts for the difference between these two models. A tree-based method with decision nodes and leaf nodes is called a decision tree. These decision nodes have two or more branches that correspond to the values of the characteristics.

Leaf nodes display decision targets as numbers.



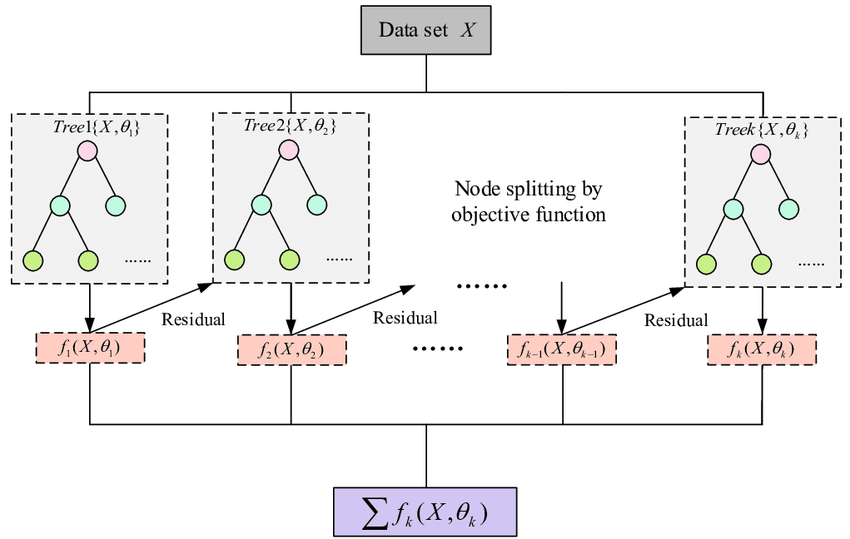
1. ***Introduction to Random Forest:***

The decision tree model is the foundation of the random forest model. However, this model is made up of a large number of decision trees that were built using a bootstrap sample obtained from the earlier bagging method. Calculated is the overall decision tree's average prediction score. Because strongly connected basis models would result in identical findings, this model heavily rely on whether the base models are diverse enough **(Jiang, 2021).**



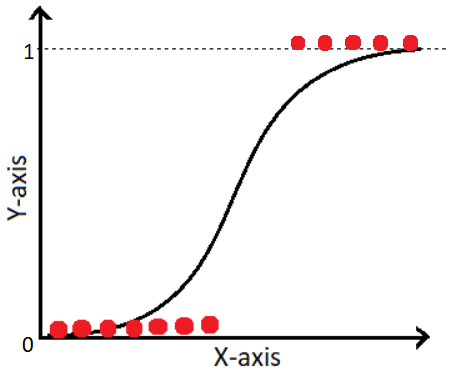
1. ***Extreme Gradient Boost Classifier (XGBoost):***

The underlying model for XGBoost is gradient boosting for the decision tree.XGBoost and AdaBoost are pretty comparable when comparing the methods used to train the decision tree models in each model. XGBoost begins by creating a single leaf rather than a tree or stump, whereas AdaBoost grows on stumps with subsequent stumps dependent on mistakes made by the prior stump. This leaf represents a first estimation of the sample weights. The tree for the leaf will continue to be made by XGBoost. However, XGBoost differs from AdaBoost in that the tree can be greater than the stumps, as opposed to AdaBoost where the tree solely comprises of stumps **(Chen & Guestrin, 2016).** XGBoost is heavily used due its robustness and scalability. XGBoost could be used for solving regression, classification, and ranking problems.



1. ***Introduction to Logistic Regression:***

A statistical method called logistic regression can be used to predict the likelihood of discrete outcomes given an input variable. This method is comparable to linear regression. However, it uses an S-Curve that uses a sigmoid function rather than linear lines to connect data points, as shown in Figure.The disadvantage of logistic regression is that complicated models cannot be handled by it, and it is quickly outpaced by competing methods.



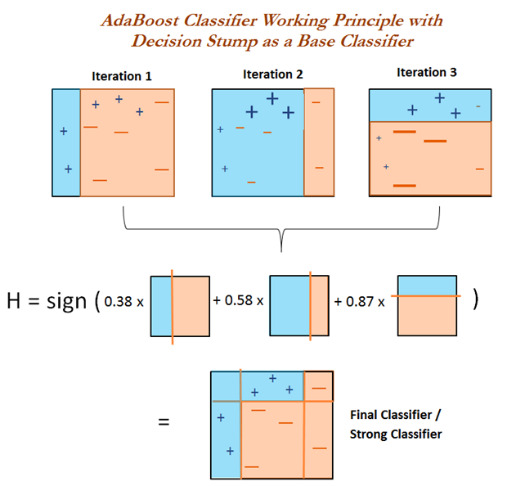
1. ***Support Vector Machine (SVM):***

The Support Vector Machine (SVM) is a supervised machine learning technique that searches an N-dimensional space for hyperplanes that can clearly identify data points. Figure below provides an illustration of a case involving two-dimensional data points, where the hyperplane is shown as the red line dividing the yellow class from the blue class. The algorithm will determine the ideal value of W, which stands for the separation between various data points.



1. ***ADABoost Classifier (ADA):***

A decision tree serves as the foundation of the AdaBoost model, which stands for adaptable boosting, however most of its algorithms use "stumps," or decision trees with only one split. Each stump is built by considering the flaws of the preceding one, and incorrectly identified stumps are given more weights and higher priority (Jiang, 2021). This model will continue to run until a little amount of error is produced. To sum up, AdaBoost is a supervised machine learning algorithm that uses an iterative method to strengthen weak classifiers by picking up from its errors in previous performance.



**3.2 proposed system:**

***A. Data Pre-manipulation :***

This specific section of the report will collect the data, assess the quality of the data, and then trim and clean the provided dataset for analysis as indicated.

***B. EDA:***

Analyze and investigate dataset and summarize their main characteristics, employing data visualization methods.

***C. DATA GRAPHING:***

The data set that was gathered to forecast the provided data is divided into a training set and a testing set. Typically, the Training set and Test set are divided into 7:3 ratios. Based on the correctness of the test results, the Data Model that was developed using machine learning techniques is applied to the Training set.

***D. LABEL ENCODING AND TREATING MISSING VALUES:***

Convert categorical columns into numerical ones so that they can be fitted by machine learning models which only take numerical data. It is an important pre-processing step in a machine-learning project.

***E. DESIGNING THE CLASSIFICATION MODEL:***

This classification model is made up of FIVE algorithms of Machine Learning such as Logistic Regression, Random Forest, Decision Tree Classifier, Support Vector Machine, ADABoost and XGBoost .

***F. FINE TUNING OF PROMISING MODELS TO OBTAIN MOST PROMISING MODEL:***

A mathematical model containing a number of parameters that must be learned from the data is referred to as a machine learning model. Hyperparameters, on the other hand, are those parameters that cannot be learned directly. Prior to starting training, humans frequently choose them based on some intuition or trial and error. By enhancing the model's functionality, such as increasing the model's complexity or learning rate, these parameters demonstrate their significance. Models may have a large number of hyper-parameters, and determining the ideal set of parameters can be approached as a search issue. Finding the best hyper-parameter for an **SVM** model is a highly challenging issue. These hyper-parameters include things like what C or gamma values to utilise. But it can be discovered by simply attempting all possible combinations and observing which inputs are most effective. Its main purpose is to build a grid of hyper-values and then to simply test every possible combination of those parameters. In order to train a model, **GridSearchCV** uses a dictionary that lists possible training-parameter combinations. The parameter grid is conceptualised as a dictionary, with the parameters serving as the keys and the settings to be tested as the values.

This article explains how to select the best hyper-parameters using the GridSearchCV searching technique, which will enhance the accuracy and prediction outcomes.

After finetuning, **XGboost** is generalizing well on our dataset. So XGBoost will be my selected model.

**PRESENTATION OF DATA, ANALYSIS & FINDINGS**

**4.1 DATASET**

The ***"Mental Health in Tech Survey"*** dataset will be used in this thesis. This dataset is made available via **OpenSource Mental Illness (OSMI).** For those who work in the IT industry and communities, OSMI works on education, raising awareness, and providing information relevant to mental health. An annual poll measuring perceptions of mental health and mental health disorders among those working in the tech industry has been undertaken by OSMI since 2014. On the website of OSMI, a questionnaire that may be completed online is used to collect the data. People who meet the requirements can take the survey globally.

This dataset is available on the **websites of OSMI and Kaggle** and is kept in.**csv** format. The number of records that are received from each survey varies annually, which has an impact on the size of the collection. For instance, there were 756 replies to the 2017 survey. However, there were just 131 responses to the 2021 survey.

Past researchers have used various subsets of all available datasets. For instance, Dharma et al. (2021) used this dataset for their research using 2016 to 2019 data, while Reddy et al. (2018) used only 2017 data. Questions used in the annual surveys each year had mainly remained constant. However, from 2017 until 2021, the questions were almost identical, with approximately 80 questions and 1500 responses. For this underlying reason, this thesis will use the "Mental Health in Tech Survey" dataset from **2014-2016**.

**4.2 DATa cleaning and preprocessing**

In this thesis, open-ended questions with lengthy responses in which the respondent describes circumstances are not included. Instead, using the label encoder function of the Sklearn library, multiple-answer questions are converted into numerical values. Later, the features are standardised with Sklearn's StandardScaler function.

The second problem is particularly connected to the age and gender inquiries.

The answers to these two questions are brief and open-ended.

This indicates that the respondent is not constrained to a set of possibilities and can provide a free response. This brings up various problems. For instance, several respondents stated that they were between the ages of "1" and "100". These responses are seen as abnormal. However, because there are so few replies that exhibit these types of responses, the dataset excludes the entirety of these observations.

A significant number of answer variants are seen for the gender question.

There are about fifty different answer categories. For the female gender, for instance, some responders say "women" "f" "female-ish". Another illustration involves the male category. While some responders simply state "Male," it is also common to see responses like "Man," "m," and "Identify as male."

The closest group is then used to encode these responses.

The final problem is a missing data. Imputation will be carried out in this thesis using the most popular method. A preventative technique is devised, though, to make sure that the bulk of the responses are not imputed data that could harm the model's performance. Only 16% of the responses had all of the required questions answered. The dataset excludes these observations.

This thesis is a **binary-class** classification problem. The target attributes have two classes which are "Yes" or "No". These target properties are eventually changed into numerical values to maintain consistency across all models because one of the models only accepts integer values as the Y values.

A close up of a word

Description automatically generated

**‘DISTRIBUTION OF TARGET ATTRIBUTE’**

Initially, the dataset has ***1260*** observations with ***27*** features. After the data cleaning process and dropping irrelevant features, the final dataset for developing the machine learning model process contains ***1241*** responses, ***23*** features, and one target attribute

(**TRAIN DATA**-***868*** responses,***24*** features **TEST DATA**-***373*** responses,***24*** features).

**4.3 procedure**

The complete flow of this thesis’ procedure is illustrated in figure below:

**A diagram of a research flow

Description automatically generated**

**RESEARCH FLOW**

We first merge the data. Raw data from 2014 to 2016 that have been compiled then went through the cleaning and preprocessing phase.

Out of the cleaned data, we divided into X (feature) and Y (target attribute), while ensuring that the target attributes would not be included as input.

The training set and test set are then separated from the dataset. 70% of the total dataset was used for the training set, while 30% was used for the test set. The effectiveness of each model will be assessed using their respective f1 scores.

These results are reported since they serve as the primary benchmark for performance comparison between the models.

The performance of each model will be enhanced by hyperparameter tuning. Given that there are numerous models for which parameter optimisation will be carried out, Random Search is chosen as the technique to save processing time. The initial 50 models for each technique are produced after the base models have undergone crossvalidation with five folds and 10 iterations each. The ideal model is chosen among those that perform the best in the cross-validation.

Utilising these ideal models to make predictions on previously unobserved data, specifically the test set, is the penultimate step of this training phase. To respond to the first subquestion of this inquiry, this stage would provide the final f1 scores. As a result, one model will be selected as the top performer and utilised to conduct a feature importance analysis to respond to the second sub-question of this thesis.

**4.4 evaluation criteria**

The dataset's response to the question ***"Do you currently have a mental health disorder?"*** serves as the ground truth, and will be compared to the predictions made by the models.

The primary metric used in this thesis to assess how well the model performed at predicting mental health disorders is the f1 score. It is often used to assess how well machine learning algorithms perform while solving categorization challenges. Precision and recall are equally important in this thesis. As a result, f1 is selected.

The model performance in the cross-validation procedure and while evaluating the out-of-sample performance will be assessed in this thesis using the f1 score. For each model, the test performance and the f1 cross-validation score will be compared. This assessment is intended to evaluate the model's performance using hypothetical data. The model can be classified as having generally steady performance and generalised if the difference between two performances is not too large. On the other hand, there is a chance that the trained model is overfitting if the performance in test data is significantly lower than cross-validation performance.

For error analysis, the model's performance in predicting across labels will also be examined in addition to its prediction accuracy. A machine learning model may perform well in the key metrics that were selected, but it may not perform well in a particular class.

As a result, it must be looked into to see if the performance is constant among labels. To deal with this, a confusion matrix table will be created.

**4.5 software and algorithms**

The entire process of putting this thesis into practice—from planning to wrapping up—was carried out in **Python 3.8.8**, with the help of a number of integrated third-party libraries.

The conclusions in this thesis are supported by a number of visualisations created with **matplotlib 3.3.4**. Furthermore, **pandas 1.2.4** is used to process the dataset as a Data Frame. Last but not least, practically all machine learning and preprocessing techniques are derived from **Scikit Learn version 0.24.1.**

**Results**

**5.1 model performance**

The 5-fold cross-validation procedure is utilised in this thesis, with 10 model candidates in each step. The best model is chosen after performing hyperparameter tuning and evaluating its performance during the validation phase. All selected algorithms (Logistic Regression, SVM, Decision Tree, Random Forest, AdaBoost, and XGBoost) are subjected to this procedure.

1. LOGISTIC REGRESSION-

A close up of a number

Description automatically generated

A screenshot of a computer

Description automatically generated

1. DECISION TREE-

A close up of a number

Description automatically generated

A screenshot of a test

Description automatically generated

1. SUPPORT VECTOR MACHINE-

A screenshot of a computer

Description automatically generated

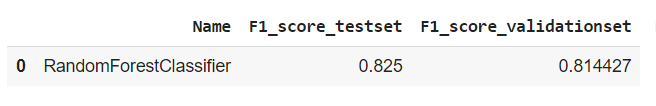
A screenshot of a test

Description automatically generated

1. RANDOM FOREST-

A screenshot of a computer

Description automatically generated



1. ADABOOST-

A screen shot of a computer

Description automatically generated

A screen shot of a test

Description automatically generated

1. XGBOOST-

A screenshot of a computer

Description automatically generated

A screenshot of a test

Description automatically generated

The prediction result on out-of-sample performance for the test performance column is used to calculate the f1 score.

Compared to logistic regression, which serves as the baseline in this thesis, all of the selected models had higher f1 scores. Therefore, it can be said that all sophisticated algorithms performed better than the traditional approach overall. The f1 scores on the test data drop for all models when comparing within models.

The lowest discrepancy between cross-validation and test performance is observed in XGBoost.

A graph with pink bars

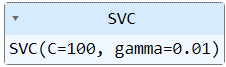
Description automatically generated

Now, we shortlisted the models and performed hyper parameter tuning on them-

1. **SUPPORT VECTOR MACHINE-**

A close-up of numbers

Description automatically generated



A screenshot of a computer code

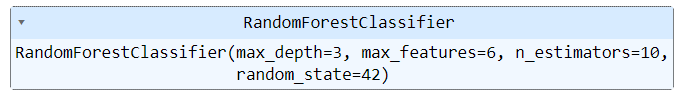
Description automatically generated

A close up of numbers

Description automatically generated

1. **RANDOM FOREST-**





A close up of a number

Description automatically generated

A close-up of a test

Description automatically generated

1. **XGBOOST-**



A screenshot of a computer code

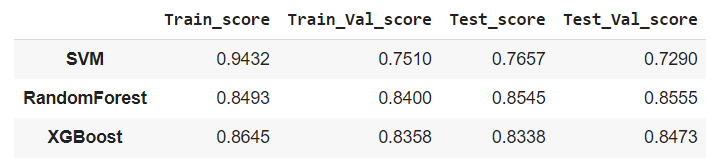
Description automatically generated

A screenshot of a computer

Description automatically generated

A close up of a number

Description automatically generated



All selected algorithms outperform the baseline model (Logistic Regression) in this thesis in terms of f1 scores. In comparison to other algorithm techniques, XGBoost has the highest f1 score and the smallest difference between cross-validation performance and out-of-sample performance. This outcome is in line with research by Reddy et al. (2018), who found that the boosting method works better than other algorithms for predicting mental health disorders.

It is the most sophisticated extension from the Decision Tree algorithm, specifically for XGBoost. In comparison to previous models, the model contains a large number of variables that may be tested, leading to model flexibility—possibly XGBoost's strongest asset.

Decision Tree (the foundational method of Random Forest), XGBoost, and AdaBoost were the top three most promising models as determined by several evaluation metrics. It is possible to conclude that boosting algorithms are promising algorithms for making predictions about mental wellness.

According to this thesis, Random Forest outperforms AdaBoost in terms of f1 score. The theoretically more advanced variant of Random Forest is AdaBoost. As a result, it is anticipated that AdaBoost will perform better.

But occasionally, a simpler version of the algorithm may perform better than a more complex one. Awal and Rao's investigation (2021) also found evidence of this occurrence. In their testing, Decision Tree performs better than XGBoost and AdaBoost, the extended version of the algorithm. The f1 score difference amongst the top three models (XGBoost, Random Forest, and AdaBoost) in this thesis is minimal.

**5.2 error analysis**

These models are designed to address the problem of binary-class classification.

To examine these models' performance in-depth, confusion matrices are built.

For all algorithms, confusion measures are created in order to assess model performance across classes. Almost all models predict "Yes" and "No" labels with promising outcomes. One of the potential explanations is that the employees in the "Yes" and "No" classes might have distinguishing qualities.

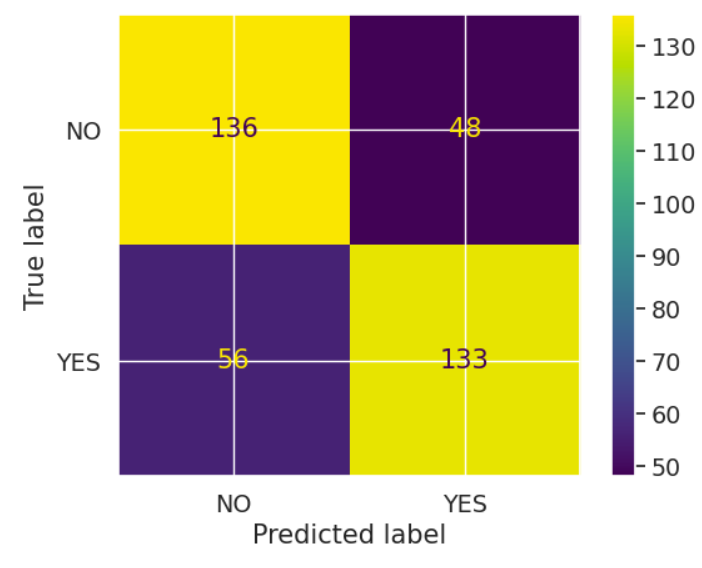
A blue squares on a white background

Description automatically generated

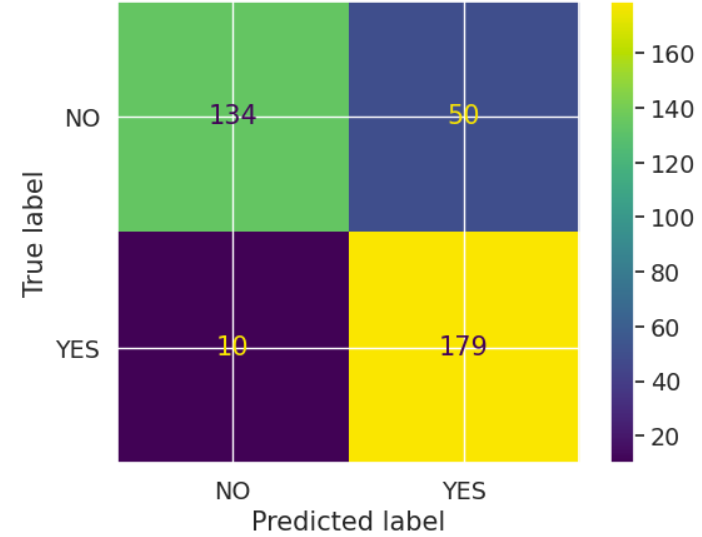
A screenshot of a computer

Description automatically generated

1. LOGISTIC REGRESSION-

****

1. DECISION TREE-



1. SUPPORT VECTOR MACHINE-

A chart of different colors

Description automatically generated

1. RANDOM FOREST-

A chart of different colors

Description automatically generated

1. ADABOOST-

A yellow and purple squares with black text

Description automatically generated

1. XGBOOST-

A chart of different colors

Description automatically generated

**5.3 feature importance**

The model with the greatest f1 score out of the five models selected for this thesis, XGBoost, is used in feature importance analysis to respond to the second sub-question of this thesis.

A graph showing a number of key features

Description automatically generated

Out of 35 features, 14 are internal factors that are either demographic questions or personal inquiries, 21 of which are external factors connected to coworkers or employers.

The biggest factor is work interference. The question the business should be asking its employees is whether their mental health concerns are interfering with their ability to do their jobs.

Employees who wish to receive treatment may also be influenced by their family history and the care alternatives (programmes and benefits) offered by their employer.

There has been a small contribution for each of the remaining features.

It may even be possible to support someone who may be struggling with mental health concerns by spotting or anticipating some of these characteristics and connecting them with the right employee resources.

This observation is comparable to the problem raised in the introduction section, which is that people's health may indirectly interfere with their ability to execute their jobs in some circumstances. In order to prevent mental health issues that could be hurting their employees' performance, it is vital for employers to periodically and appropriately check their mental health.

According to **Reddy et al. (2018)**, gender is the most effective predictor of mental health disorders. However, according to the findings of this study, gender is not a significant predictive factor for mental health. The fact that this thesis's focus on mental health disorders is more broad than the previous one's, which simply addressed stress, may be one explanation. Consequently, distinct predictors and risk variables may apply to a larger set of mental health illnesses.

A number of employer-related characteristics also demonstrate how important the models are to the employers. For instance, the value the employer places on both physical and mental health, as well as the simplicity of requesting a leave of absence.

The same conclusion was reached regarding the employer's perspective on physical and mental health in feature importance analyses conducted **by Reddy et al. (2018) and HENRY and ISA (2022)**.

The working environment and corporate regulations will reflect how important a firm believes its employees' physical and mental health is. For instance, having access to doctors or counsellors in the office, having physical and mental health benefits, or having a collaborative environment that promotes psychological safety may help employees cope with the stress of the job. In contrast, if employees' health is given little priority, they may be more susceptible to concerns relating to their mental health.

Many things could be contributing causes in someone having a mental health problem. Based on the findings of this thesis, it is believed that internal factors are more significant than external ones in predicting mental health disorders for those who work in the IT sector.

Openness to discussing mental health issues and active involvement in resolving mental health-related issues show greater value in the model than demographic questions.

People frequently have little choice over the demographic questions, such as location or race. Personal questions, on the other hand, generally concern one's attitude and behaviour, both of which are malleable. With this in mind, one may choose to handle mental health difficulties by concentrating more on personal, determinable factors than on immovable ones.

For instance, more proactive behaviour to ask for assistance when mental health-related problems interfere with employment. In order to manage a mental health illness, people may alter their strategy or behaviour. The desired outcome could not be possible, though, if the environment is unfavourable or there aren't any resources accessible.

The assistance from the employer is also a crucial component in the framework of this thesis.

**Discussion and critical analysis**

Problem Setting: An April 2018 paper in the Journal of Occupational and Environmental Medicine found that after receiving treatment for depression, over 86% of workers reported better work performance and lower absenteeism rates. For employers, this means significant increases in productivity and retention. The IT company can start cultivating a culture of understanding and compassion by granting employees access to mental health insurance. Additionally, it is good for business to have contented and caring personnel.

Employers can make more effective use of their resources by using this model to understand employee mental health issues and offer benefits to those who qualify. This technique can assist in reducing the additional costs associated with giving mental health benefits to those who do not request them and using that money for the employee's other benefits instead. As a result, employee retention will eventually rise, increasing employee satisfaction.

**6.1 questions & answers**

### **Q1: Have you sought treatment for a mental health condition?**

A blue and red squares

Description automatically generated

‘Treatment’ is the target variable.

There is no class imbalance, hence resampling is not necessary.

Approx 50% of respondents said they wanted to receive therapy. Workplaces that support employees with mental illnesses and encourage mental health are more likely to see declines in absenteeism and productivity as well as related economic gains.

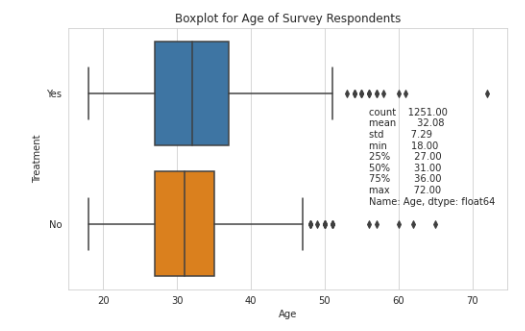
Employees who are in good mental health are better able to maximise their potential, deal with life's challenges, and contribute fully to their relationships, workplaces, and communities.

### **Q2: What is your age?**

A graph of a graph

Description automatically generated

Clearly both distributions are merging.



**The data are highly skewed and with Positive skewness** where the mode is smaller than mean or median. It's indicated that most of the employees that fill the survey around the end 20s to early 40s. Given that younger people tend to work in the technology sector, the age distribution is right-skewed, as would be predicted.

### **Q3: Are you self-employed?**

A graph of a bar graph

Description automatically generated

10% of people are self employed. Even though there is a significant difference, the percentage of people seeking treatment is the same in both categories.

Therefore, whether a person gets treatment or not doesn't actually depend on whether they are self-employed or not.

Although there is an imbalance in the categories, each category's class distribution is quite uniform. Thus, the model may not be affected.

### **Q4: If you have a mental health condition, do you feel that it interferes with your work?**

A graph of a bar graph

Description automatically generated

A number with black text

Description automatically generated

A graph of different colored squares

Description automatically generated

Interruption at work has been encountered by about 78% of respondents, with a ratio of rarely, occasionally, and frequently.

About 45% of the time, mental health issues interfere with one's ability to work. The graphs demonstrate that over 80% of patients desire therapy. It's interesting to learn that despite the fact that mental illness has never been an issue at work, a small percentage of people still desire to receive treatment before it causes difficulty on the job. When a worker's abilities, resources, or demands are not met by the job's criteria, it may result.

If you are in charge of a tech company, you should think about offering resources to those who need medical care. This will improve the work experience and undoubtedly raise productivity.

### **Q5: Do you have a family history of mental illness?**

A graph of a bar chart

Description automatically generated

People are more likely to seek therapy if they have a family history of mental illness. About 35% of those with no family history are also looking for assistance. insight for creating models.People are more likely to seek treatment if they have a family history than they are without one.

According to 40% of respondents, those who report a family history of mental illness are substantially more likely to wish to receive treatment than those who don't. This is acceptable since those who have a family history of mental illness are more aware of it. A substantial risk factor for many mental health issues is family history. When a family member has a history of issues similar to those experienced by someone with mental illness, it is not uncommon for the apple to fall far from the tree.

Family history will be an important feature.

### **Q6: Do you work remotely (outside of an office) at least 50% of the time?**

A graph of a bar chart

Description automatically generated

A graph of a bar

Description automatically generated

Whether they engage in remote work or not, about 50% of both kinds of persons seek treatment.

There are slightly more people who seek treatment who work remotely.

Since almost 70% of respondents do not work remotely, the workplace was cited as the main trigger for mental health disorders. On the other hand, there is a tiny difference between an employee who wants to receive treatment and one who does not. However, things start to become interesting when we find a respondent who works remotely for 50% of the workday. One or two percentage points more workers wish to receive treatment. The data doesn't provide that information, so I'm not sure why those people are working remotely to conduct additional analysis.

The lack of social connection in distant mode may be to blame.

### **Q7: Does your employer provide mental health benefits?**

A graph of a bar graph

Description automatically generated

A black text with numbers

Description automatically generated

A graph of different colored squares

Description automatically generated

We can see that about 38% of respondents claimed their workplace offered them mental health benefits, whereas a sizable portion (32%) didn't even know if they had this benefit.

In the second graph, we can observe that 63% of those who responded "YES" to the question about mental health advantages claimed they were seeking medical attention. Thus, it is clear that the employer is making better use of its resources.

Even if you consider the cost, you should still go for it because the staff use it effectively.

Surprisingly, nearly 45% of those who rejected the company's offer of mental health benefits nevertheless desire to receive therapy for their mental health.

### **Q8: Do you know the options for mental health care your employer provides?**

A graph with a blue rectangle

Description automatically generated

A close-up of numbers

Description automatically generated

A graph of different colored squares

Description automatically generated

25% of employees are unsure whether the employer offers care options, while 40% of employees receive no care options.

We can observe that 60% of employees whose employers don't offer healthcare options are seeking help. These organisations must deal with this problem. This can support our argument that people with care alternatives are genuinely seeking therapy.

### **Q9: How many employees does your company or organization have?**

A graph of a bar graph

Description automatically generated

A graph of different colored bars

Description automatically generated

It's an interesting finding that larger businesses (those with more employees) offer benefits more so than smaller ones.

A graph of different colored bars

Description automatically generatedNo matter how big or small the organisation, more than 50% of its members are looking for assistance.

This issue is especially significant in small organisations, possibly because there aren't as many health benefits offered.

### **Q10: Has your employer ever discussed mental health as part of an employee wellness program?**

A graph of a bar

Description automatically generatedA graph with red and blue squares

Description automatically generated

The majority of respondents' employers didn't mention mental health as part of their employee wellness initiatives.

About 50% of those who are unaware of the programme are looking for assistance. This means that businesses should describe the benefits they offer for mental health.

The employee wellness programme should cover mental health for businesses. This must not be disregarded.

More than 65% of respondents claim that their firm doesn't offer any wellness programmes. However, considering that half of the respondents wanted therapy, the business needed to start offering it fast. I believe it makes sense if it's about company budgeting based on my earlier inquiry about the benefit of the firm. I'm aware that it will cost a lot of money, and the business also has many employees that need to be taken care of. My second perspective is that budgeting is still important, but it will be very difficult for a tiny business.

### **Q11: Does your employer provide resources to learn more about mental health issues and how to seek help?**

A graph with a blue rectangle

Description automatically generated

A graph with red and blue bars

Description automatically generated

The majority of businesses either don't offer any resources or, possibly, don't provide information about the resources that are already available.

### **Q12: Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?**

A graph with blue squares

Description automatically generatedA graph of different colored squares

Description automatically generatedThe majority of people are unaware of their right to privacy protection when seeking medical attention.

For consumers to feel comfortable sharing their difficulties, businesses should offer a secure environment.

We can prove that people are more likely to seek therapy if they believe their identity is preserved.

Nearly 65% of respondents desire to receive treatment, and about 30% of respondents say yes if their anonymity is protected while using services for mental health or drug misuse therapy. The employee believes that their privacy was safeguarded by the company, and it was a wise decision on the part of the business to earn their trust. The employee wishes to receive treatment as a result in order to get better.

### **Q13: How easy is it for you to take medical leave for a mental health condition?**

A graph of a bar graph

Description automatically generated

A graph of different colored squares

Description automatically generated

Despite the fact that fewer people believe it is difficult to obtain leave for mental health concerns, more people in this category seek assistance than in any other.

### **Q14: Do you think that discussing a mental health issue with your employer would have negative consequences?**

A graph of a graph

Description automatically generated

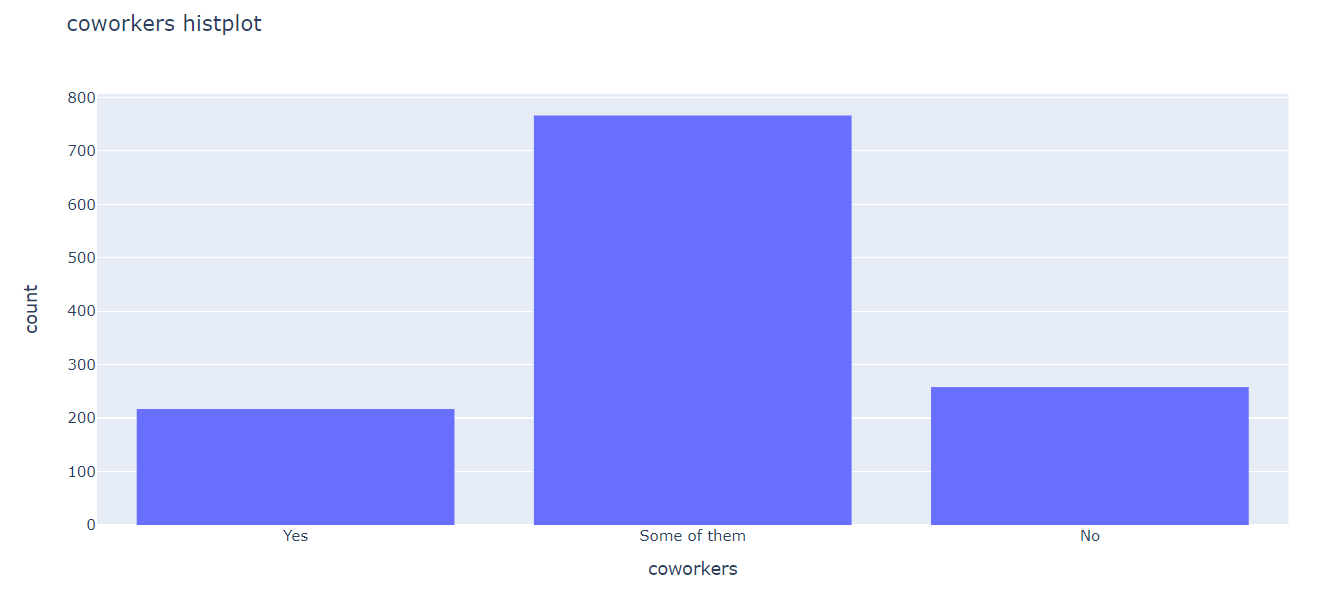
A graph of different colored squares

Description automatically generated

About 70% of people either believe that talking about mental health issues would have negative consequences or are unsure of whether they will.

And the majority of those who believe it will have a negative effect seek treatment.

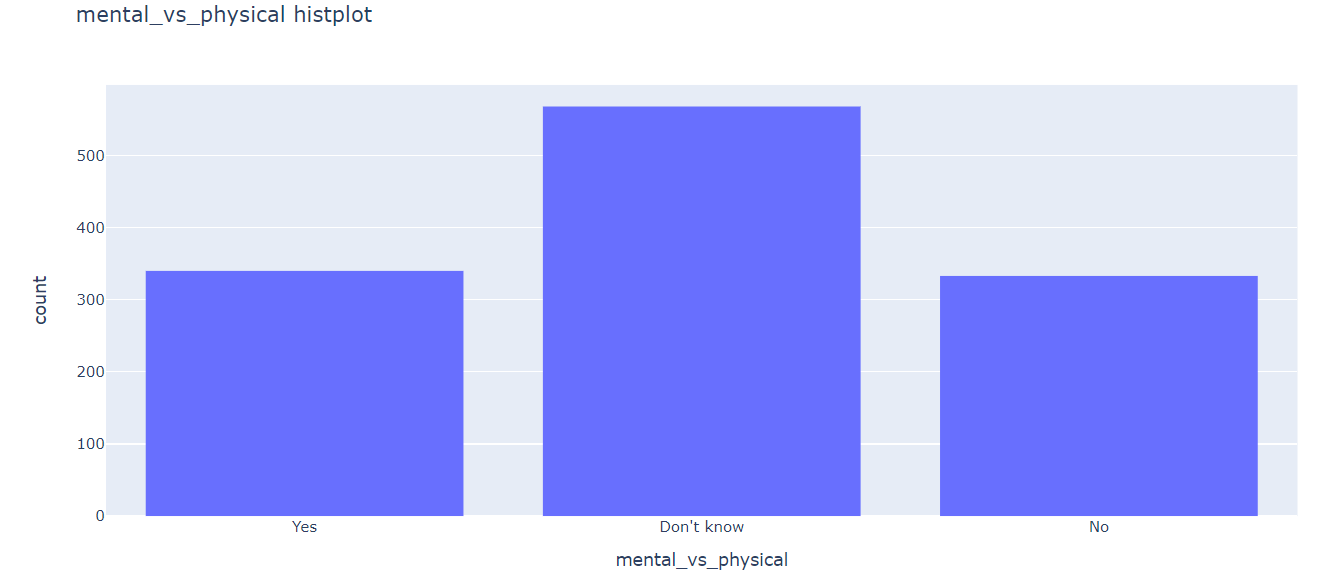
### **Q15: Would you be willing to discuss a mental health issue with your coworkers?**

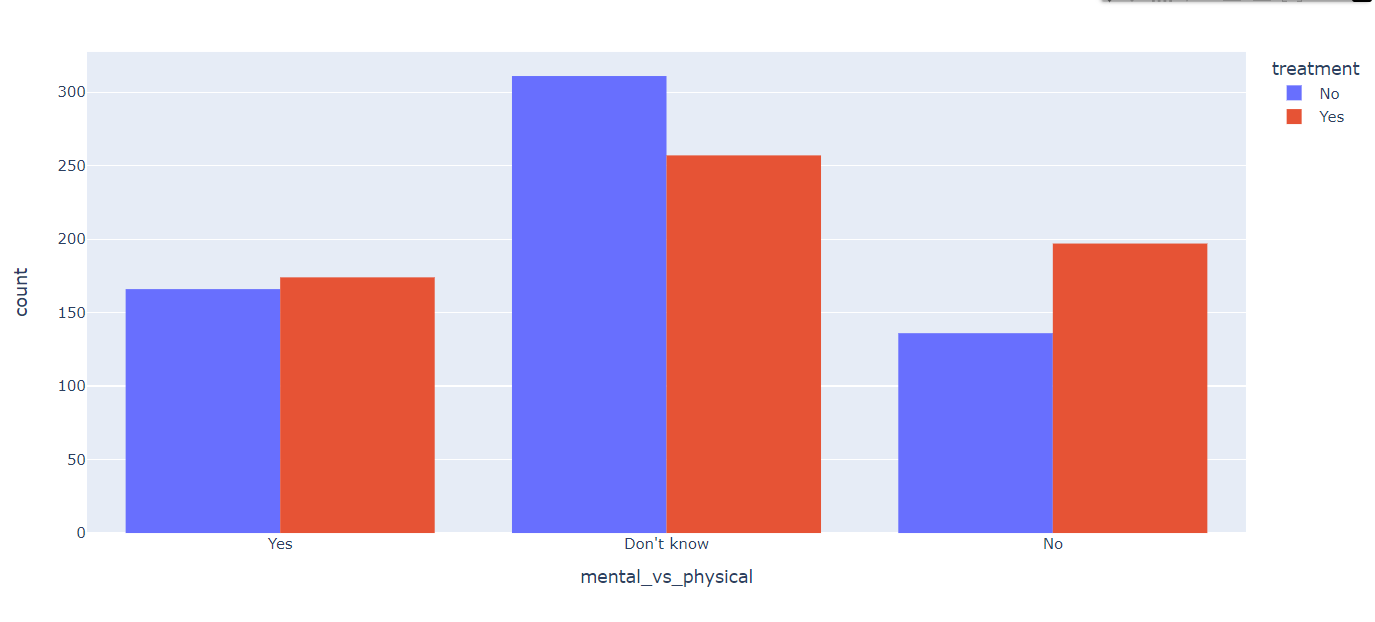
A graph with different colored rectangles

Description automatically generated

It is encouraging that the majority of people have at least one or two coworkers with whom they can discuss mental health difficulties.

### **Q16: Do you feel that your employer takes mental health as seriously as physical health?**





Employees are more likely to seek treatment than the other two groups if they believe their organisation doesn't take mental health seriously or if they are unsure.

**Q17: Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?** A blue rectangles and white squares

Description automatically generated

A graph of a bar graph

Description automatically generated

Nearly 85% of people claim to have never heard of or seen coworkers who were penalised for having mental health problems.

10% of the remaining individuals who saw unfavourable effects on coworkers are seeking assistance.

**Q18: Does Gender matter**

**A graph with a bar and a number of squares

Description automatically generated**

**A graph with different colored squares

Description automatically generated**

Unsurprisingly, the majority of respondents—nearly 79%—are male, particularly in the tech industry. Women are under more competitive pressure than men are due to the very wide divide between the sexes. According to the story, a substantial percentage of women—around 70%—want to receive treatment. Perhaps some of them experience racism or sexual harassment at work because there aren't many women working in the technology sector.

Less than 2% of entries are Queer. Despite the extremely limited LGBT population, some fresh perspectives should still be uncovered. For instance, such a little percentage can reveal a sizable difference in the number of people who desire the therapies, showing that mental health issues are also critical for the queer community. They might have encountered prejudice or hate speech at work, in my opinion.

**Q19: Would you be willing to discuss a mental health issue with your direct supervisor(s)?**

A blue and white graph

Description automatically generatedA blue and red squares

Description automatically generated

Only 55% of responders, out of the 40% who indicate they will discuss with their supervisor, actually desire treatment. I believe that getting some relief from talking to a higher authority figure would assist. When employees talk with coworkers, it's the opposite.

**Q20: Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?**

A blue and white rectangles

Description automatically generated

A graph with different colored squares

Description automatically generated

Nearly 70% of respondents (15%) who acknowledge recognising the harmful effects of employees with mental health conditions wanting treatment. When an employee is aware of the negative effects, it serves as a good motivator to seek treatment for mental health disorders.

**6.2 future work**

The results of this study's model performance analysis are in line with earlier studies that found Boosting to be more effective than other algorithms. The efficacy of Boosting and various Deep Learning techniques should be compared in future research to better analyse the ideal model for forecasting mental health issues.

Future studies should investigate further the best method for predicting classes that are in between two labels, such as "Maybe," as this model only performs well in predicting the "Yes" and "No" classes. The performance of the models would enhance if the findings were consistently good across courses.There is no automated feature selection process used in this thesis.

The best indicators for future investigations could be selected automatically using data science techniques.

**Conclusion and references**

* 1. **Conclusion**

The goal of this thesis is to discover machine learning algorithms that perform well in forecasting the current mental health issues of tech professionals.

The concept of a mental health disorder in this thesis is whether someone

possesses a mental illness now or not. The top three models, based on f1 scores as the primary assessment metric, are XGBoost, Random Forest, and AdaBoost. These models produce encouraging and largely consistent results, with XGBoost being the model that performs the best.

To determine the variables that have the most impact on mental health prediction, feature importance analysis is also carried out. The interruption brought on by mental health problems and prior diagnoses of mental health disorders are the two components that contribute most significantly to the model. In the model, internal influences are more significant than external ones.

The first step for an employer to address issues linked to mental health issues is to evaluate the employee's mental health using a machine learning model. A person's mental health issues are influenced by a variety of circumstances. Therefore, in order to adequately address mental health issues, the entire ecosystem must be involved.

The usage of machine learning in the field of mental health could be expanded in the future by using other datasets and various work sectors.

* 1. **REFERENCES**

Awad, M., & Khanna, R. (2015). Efficient learning machines: theories, concepts, and applications for engineers and system designers. Springer nature.

Awal, G. K., & Rao, K. (2021). Can machine learning predict an employee’s mental health? In International conference on information, communication and computing technology (pp. 235–247).

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785–794).

Dharma, A., Pasaribu, H., Aritonang, L. A., Telaumbanua, A., Hulu, V. T., Samosir, F. J., et al. (2021). Method analysis on support vector machine and fully connected neural networks on mental health among tech workers. In 2021 2nd international conference on smart cities, automation & intelligent computing systems (icon-sonics) (pp. 65–69). Dimoff, J., & Kelloway, E. K. (2019). Mental health problems are management problems: Exploring the critical role of managers in supporting employee mental health.

Elias, J. (2020, November). Google is tackling mental health challenges among employees through ‘resilience training’ videos.

Retrieved from https://www.cnbc.com/2020/11/27/google-tackling-mental -health-among-staff-with-resilience-training.html HENRY, M., & ISA, S. M. (2022).

Mental health treatment prediction for tech employee with the implementation of ensemble methods. Journal of Theoretical and Applied Information Technology, 100(8). Jiang, H. (2021).

Machine learning fundamentals: A concise introduction. Cambridge University Press. Reddy, U. S., Thota, A. V., & Dharun, A. (2018). Machine learning techniques for stress prediction in working employees. In 2018 ieee international conference on computational intelligence and computing research (iccic) (pp. 1–4).

Saini, A. (2021, September). Adaboost algorithm – a complete guide for beginners. Retrieved from https://www.analyticsvidhya.com/blog/ 2021/09/adaboost-algorithm-a-complete-guide-for-beginners/ WHO. (2008, March).

Mental health: strengthening our response. Retrieved from https://www.who.int/news-room/fact -sheets/detail/mental-health-strengthening-our-response Zhang, T., Lin, W., Vogelmann, A. M., Zhang, M., Xie, S., Qin, Y., & Golaz, J.-C. (2021).

Improving convection trigger functions in deep convective parameterization schemes using machine learning. Journal of Advances in Modeling Earth Systems, 13(5), e2020MS002365.***.***

[http://arno.uvt.nl/show.cgi?fid=160766- Research](http://arno.uvt.nl/show.cgi?fid=160766-%20Research) paper