



A Data-Driven Study of Mental Health Trends in the Tech Industry: Statistical and Machine Learning Perspectives

Muneeb Javed¹, Amna Zafar^{1,*}, Beenish Ayesha Akram², Muhammad Waseem¹ and Talha Waheed¹

¹Department of Computer Science, University of Engineering and Technology, Lahore, 54000, Pakistan

²Department of Computer Engineering, University of Engineering and Technology, Lahore, 54000, Pakistan

*Corresponding Author: Amna Zafar. Email: amnazafar@uet.edu.pk

Received: 02 May 2024; Revised: 07 June 2024; Accepted: 24 July 2024; Published: 1 August 2024

AID: 003-02-000036

Abstract: Mental wellbeing is critical for people to survive in the fast-paced workplace of today in order to succeed both personally and professionally. The computer sector poses particular difficulties for mental health because of its intensive work ethic, long hours, and high levels of stress. Research by Open Sourcing Mental Illness (OSMI) has shown that mental health disorders are especially prevalent in the technology sector. This study aims to identify important elements and facilitate early detection through a comprehensive analysis of mental health trends in the tech sector. This paper looks at the frequency of mental health problems among technical professionals in comparison to their non-technical counterparts using the Mental Health in Tech survey dataset, collected from people all around the world. Using rigorous statistical analysis and predictive modelling, the study explores variations in the frequency of mental disorders across different geographic locations and examines workplace attitudes towards mental health. The study combines a Python-based approach comprising feature engineering, exploratory data analysis, data preparation, and machine learning model building to forecast mental health diagnoses. The findings of this study highlight important mental illness and treatment-seeking attitude determinants as well as the prevalence of mental health problems in tech-related companies. This study aims to help create work environments that give mental health a top priority by elucidating effective strategies for promoting worker well-being and encouraging help-seeking behavior.

Keywords: Predictive Mental Health; Workplace Wellness; AI in Mental Health; Machine Learning; Mental Health Risk; Work-Life Balance;

1. Introduction

Our mental health is the emotional, psychological, and social well-being we have that shapes our relationships with others, our job performance, and our physical health maintenance. The IT industry, famous for its great stress levels, long work hours, and great drive to succeed, is coming under closer examination for how these elements influence employees' mental health. The World Health Organization (WHO) has highlighted how prevalent stress and strain at work are in technical sectors and how such factors can lead to significant loss of production and job churn [1, 2]. Because employees are often younger and have been subjected to lengthy workdays, the computer sector is more likely to experience burnout and a

bad work-life balance [3]. The tech sector's fast expansion highlights the need to handle workplace elements influencing mental health, with more than 53.2 million full-time workers in 2019 and a projection to reach 62 million by 2023 [4, 5].

Though mental health problems are common, many people wait a long time to get help—an average of ten years before consulting a general medical practitioner and much longer before seeing a psychiatrist [6]. Untreated mental health problems might get worse with time; hence, early diagnosis and treatments are quite vital [7]. A staggering 51% of tech professionals have been diagnosed with a mental health condition, and 57% report that their productivity is affected by mental health issues [8, 7].

Tech workers are five times more likely to suffer from mental health problems compared to other sectors, with stress, anxiety, and depression being prevalent [9]. The long-hours culture and the pressure to excel contribute to this trend, leading to severe physical symptoms like headaches, sleep deprivation, and anxiety attacks [10]. The stigma surrounding mental health in tech is a significant barrier to seeking help. Many employees fear job security and judgment from colleagues, with 38% concerned about speaking up and 17% worried about-facing prejudice [11]. Almost a quarter of tech workers lack official protocols for addressing mental health concerns within their teams, and 65% feel that physical health issues are given less prejudice than mental health issues [12].

Machine learning is being used to help diagnose mental health problems. This special tool can look at complicated patterns that have to do with how people are feeling and can guess better who might have mental health issues. Because machine learning uses a lot of data, doctors can find out sooner if someone has a mental health problem. It can also tell us trends about mental health in workplaces. For our project, we're using different kind of machine learning strategies. These let us identify the main causes of mental health problems in the world of technology. With this, we can get a better idea of mental health patterns, find out faster who might have problems, and help create plans to support the health of workers before problems come up.

The rest of the paper is organized as follows. In Section 2, previous work defining the impact of long working hours in mental well-being is presented. In section 3 methodology for mental health wellbeing analysis is presented. Results of the analyzed models is presented under section 4, while section 5 concludes our work.

2. Literature Review

A wide range of research and programs focusing on comprehending, avoiding, and managing mental health problems among employees are together referred to as mental health in the workplace. Memish et al. (2017) carried out a systematic study to assess the standard and scope of employer-developed recommendations for handling mental health issues at work. In order to translate scientific data into useful suggestions for averting mental health problems at work, this review conferred with specialists in psychology, public health, and mental health promotion [11].

Kahn et al. (2003) offered a thorough manual for identifying, comprehending, averting, and addressing mental health problems in individuals and organizations at work. In addition to emphasizing the availability of high-quality mental health care, their work focuses on developing corporate productivity and employee mental health through the development of systems and cultures [12].

In 2009 Chopra et al. examined how productivity at work affects people's physical and mental health, especially in developing nations. Their research demonstrated the connection between workplace stress and the rise in prevalent mental illnesses, highlighting the necessity of studies on mental health promotion and intervention to enhance worker productivity and well-being [13].

In 2020 Sasaki et al. looked into the relationship between Japanese employees' mental health and productivity and workplace policies put in place in response to COVID-19. The impact of organizational responses to the pandemic on employee well-being was illuminated by their study, which found strong relationships between workplace measures, psychological distress, fear and worry related to COVID-19, and work performance [14].

Eaton et al. (2018) claim that because mental health issues get insufficient worldwide focus, society must address stigma around mental illness. Their study objective was twofold: first, to raise workplace awareness of mental health practices; second, to develop workplace projects supporting a health workplace environment by means of techniques preserving and strengthening worker mental health. Latest research uses machine learning tools on mental health analysis to find disorders as well as assess susceptibility to such conditions. Examining COVID-19 epidemic impacts on workplace mental health, associations investigated stress elements as well as employee stress patterns.

Diverse literature review investigating several preventative and treatment strategies for worker mental disorders reveal how mental health interacts with workplace settings, according to the literature review. Studies indicate that companies who get the need for proactive actions supporting and improving employees' mental health also grasp the workplace relevance of mental health.

3. Methodology

We used the openly available dataset. The results led us by the provide important points for further action.

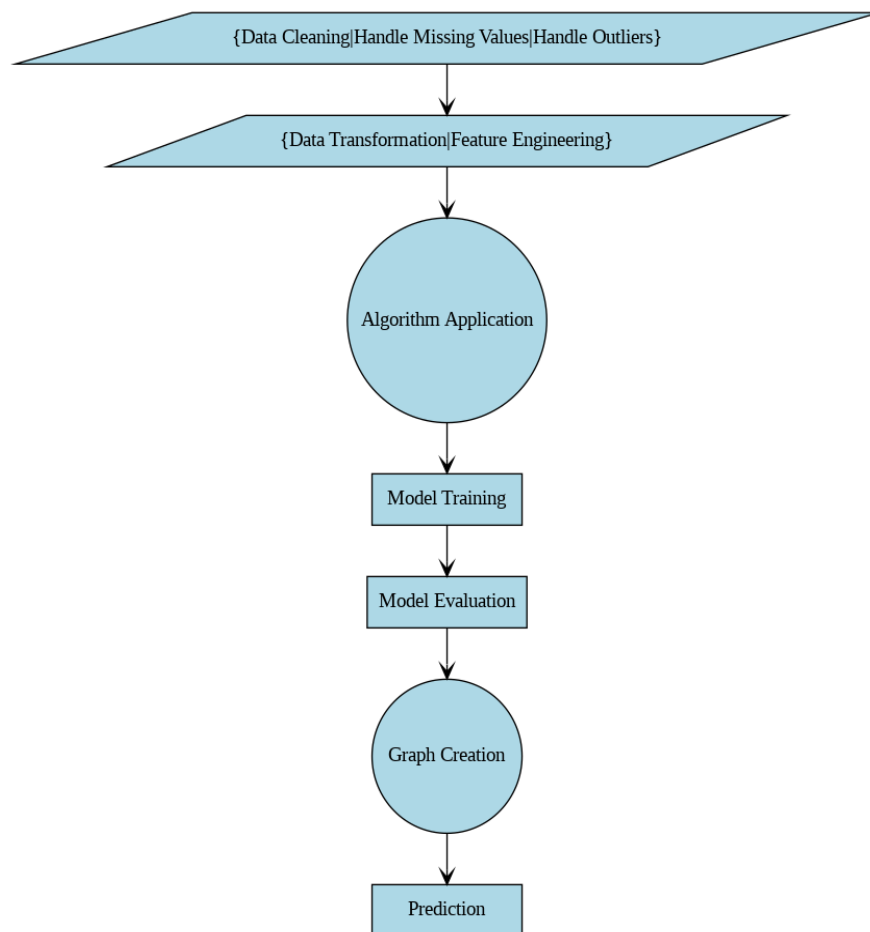


Figure 1: Methodology for Model Development and Prediction

3.1. Data Collection

The Open Sourcing Mental Illness (OSMI)—Mental Health in Tech Survey’s 2016–2021 data were used in this study [16]. Approximately 1400 responses were gathered for 63 questions concerning employee

mental health, attitudes toward mental health in the workplace, awareness of mental health, demographics, and other related topics. The purpose of the 2016 survey is to assess the attitudes toward mental health of tech employees and investigate the frequency of mental health issues among them. Comparably, the 2017–2021 OSMI Mental Health in Tech Survey data sets include 756 respondents, 417 in 2018, 352 in 2019, 180 in 2020, and 131 in 2021.

The data sets were utilized to analyze workplace culture trends, employee mental health situations, and the effects of COVID-19 on these areas.

Records	Features	Dataset Size
1259	27	296 KB

Id	Features	Description
01	Timestamp	Time the survey was submitted.
02	Age	The age of the person.
03	Gender	The gender of the person.
04	Country	The country name where person belongs to.
05	state	The state name where person belongs to.
06	self_employed	Is the person self employed or not.
07	family_history	Does the person's family history had mental illness or not?
08	treatment	Have you sought treatment for a mental health condition?
09	work_intefere	If you have a mental health condition, do you feel that it interferes with your work?
10	no_employees	How many employees does your company or organization have?
11	remote_work	Do you work remotely (outside of an office) at least 50% of the time?
12	tech_company	Is your employer primarily a tech company/organization?
13	benifits	Does your employer provide mental health benefits?
14	care_options	Do you know the options for mental health care your employer provides?
15	wellness_program	Has your employer ever discussed mental health as part of an employee wellness program?
16	seek_help	Does your employer provide resources to learn more about mental health issues and how to seek help?
17	anonymity	Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
18	leave	How easy is it for you to take medical leave for a mental health condition?
19	mental_health_consequence	Do you think that discussing a mental health issue with your employer would have negative consequences?
20	phy_health_consequence	Do you think that discussing a physical health issue with your employer would have negative consequences?
21	coworkers	Would you be willing to discuss a mental health issue with your coworkers?
22	supervisor	Would you be willing to discuss a mental health issue with your direct supervisor(s)?
23	mental_health_interview	Would you bring up a mental health issue with a potential employer in an interview?
24	phts_health_interviw	Would you bring up a physical health issue with a potential employer in an interview?
25	mental_vs_physical	Do you feel that your employer takes mental health as seriously as physical health?
26	obs_consequence	Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?
27	comments	Any additional notes or comments.

Figure 2: Features Description

We have a situation where the 'state' field contains 40% missing values and 40% of the data comes from outside the US.

- It is noteworthy that the location of California, which is in the United States, is the mode for the 'state' field.
- Therefore, it wouldn't be appropriate to replace missing values in the 'state' column for different countries with 'California'.

3.2. Exploratory Data Analysis

A Series of questions to uncover patterns, trends, anomalies, relationships, and key insights without making any formal assumptions about the data.

- Q. What is the association between Gender & Treatment?

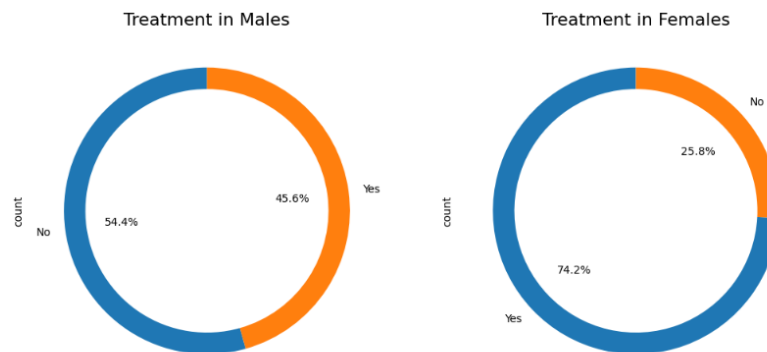


Figure 3: Association Between Gender and Treatment

Based on the data, it appears that individuals identifying as Females exhibit a higher tendency to seek treatment for mental health issues in comparison to Males. The statistics shows 48% of Males and 71% of Females have gone through treatment among the top 3 countries.

- Q. What is the association between treatment and work_interference?

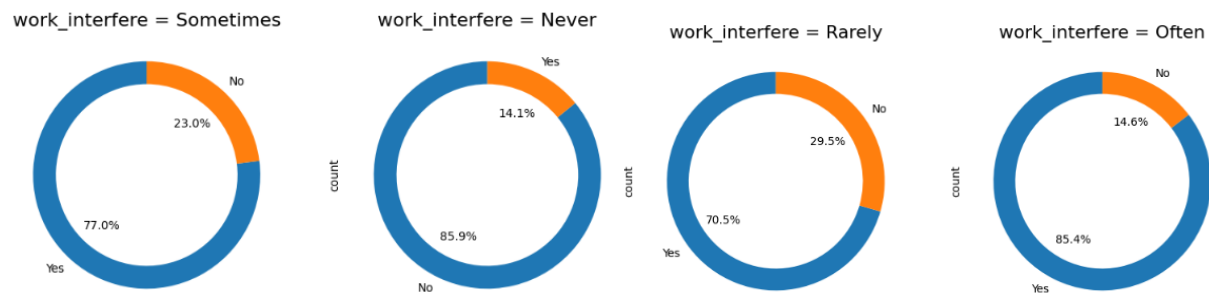


Figure 4: Association Between Treatment and work_interference

We can observe that employees who are more 'Often' & 'Rarely' interfered during work are likely to have Mental health issues and hence are seeking Treatment.

- Q. Do individuals show a greater willingness to seek treatment for mental health issues if there is a family history of such conditions?

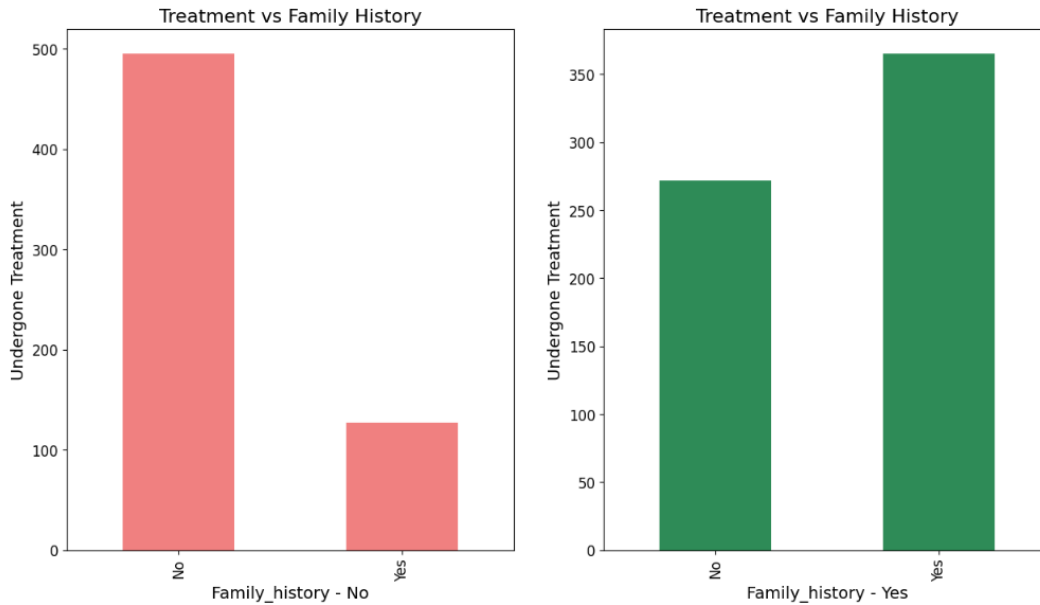


Figure 5: Association Between family_history and Treatment

We observe that employees with a family history of mental health issues are more inclined to choose treatment.

In contrast, employees without a family history of mental health issues may have lower awareness and, consequently, a reduced likelihood of seeking treatment.

- Q. What is the association between treatment and employee count in a company?

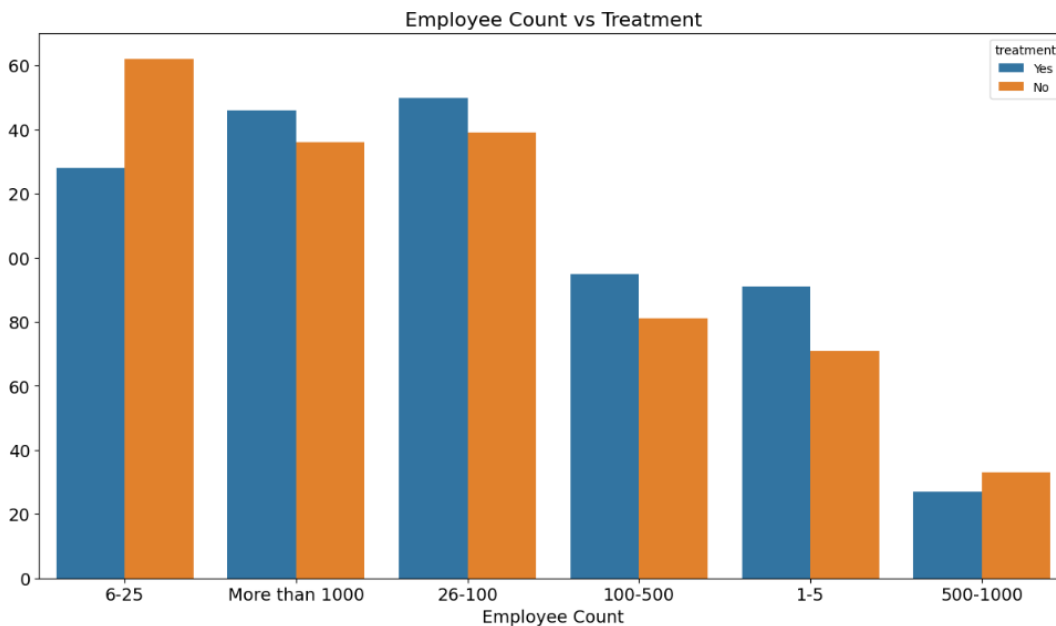


Figure 6: Association Between Employee Count in Company and Treatment

Based on the data, it can be inferred that the highest number of employees who sought mental health treatment belong to companies sized between 26-100 employees.

Conversely, the largest number of employees who did not seek treatment come from companies sized between 6-25 employees.

3.2.1. Data Set Analysis and Findings.

- Q. Top 10 Countries recorded for mental health treatment?

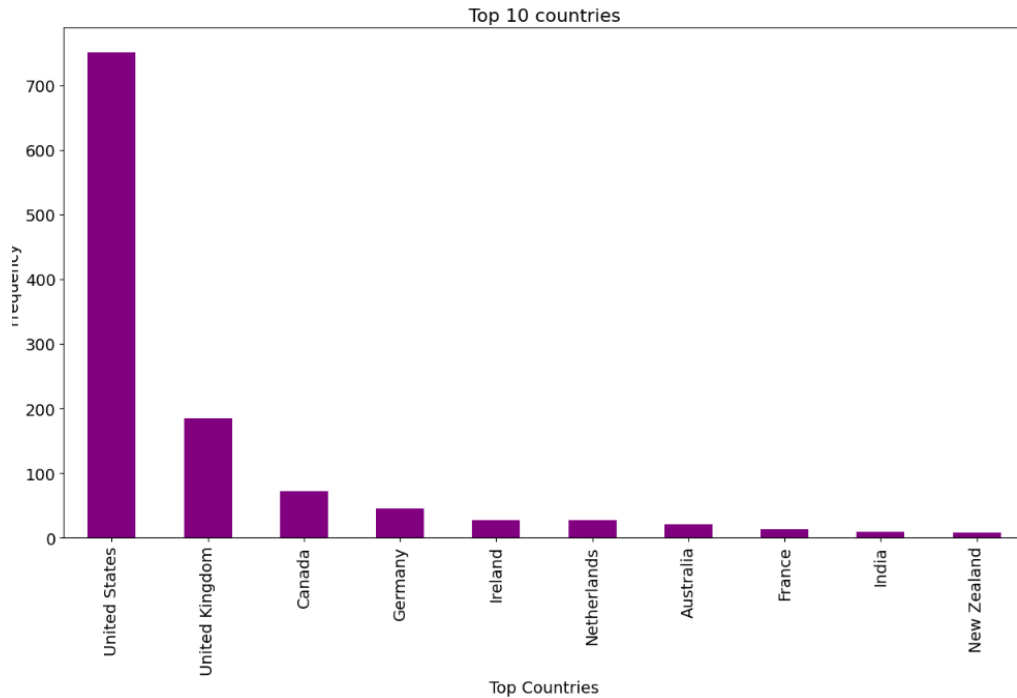


Figure 7: Country Wise Mental Wellness Treatment

The majority of the records are from the United States, followed by the United Kingdom and Canada.

- Q. Which countries are actually contributing more for mental health treatment?

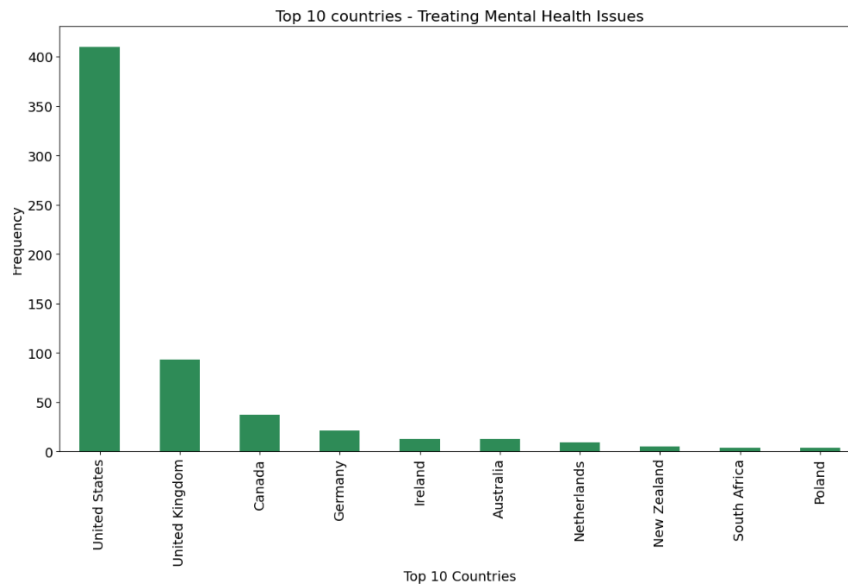


Figure 8: Country Wise Mental Issues Treatment

A change can be seen in the bar chart's lower portion, which shows the nations where more people are seeking mental health therapy.

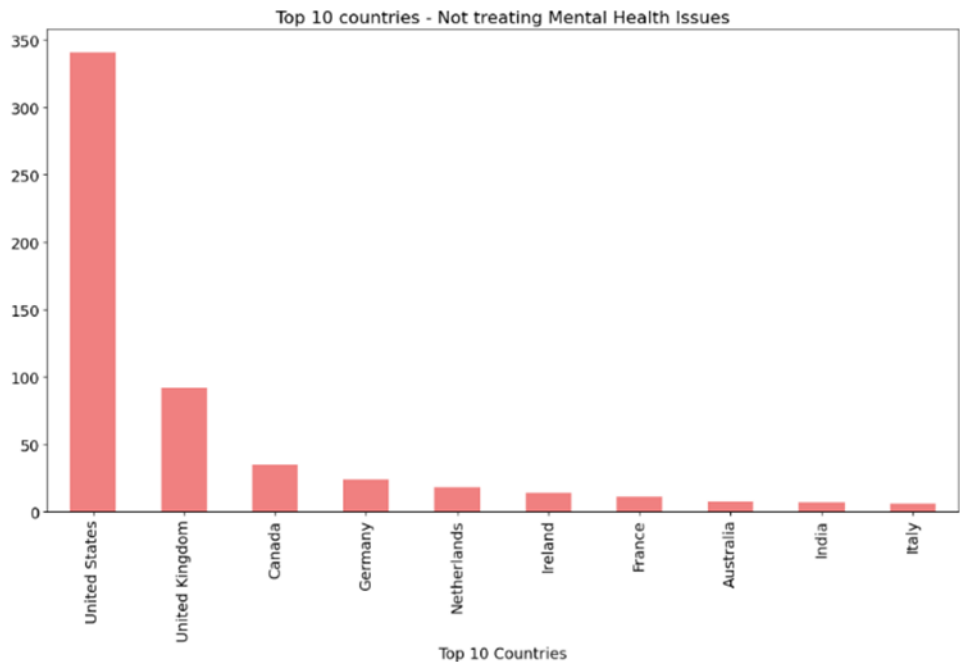


Figure 9: Country Wise Mental Issues Not Being Treated

Here are the Top 10 countries where individuals are least inclined to seek treatment for their mental health concerns.

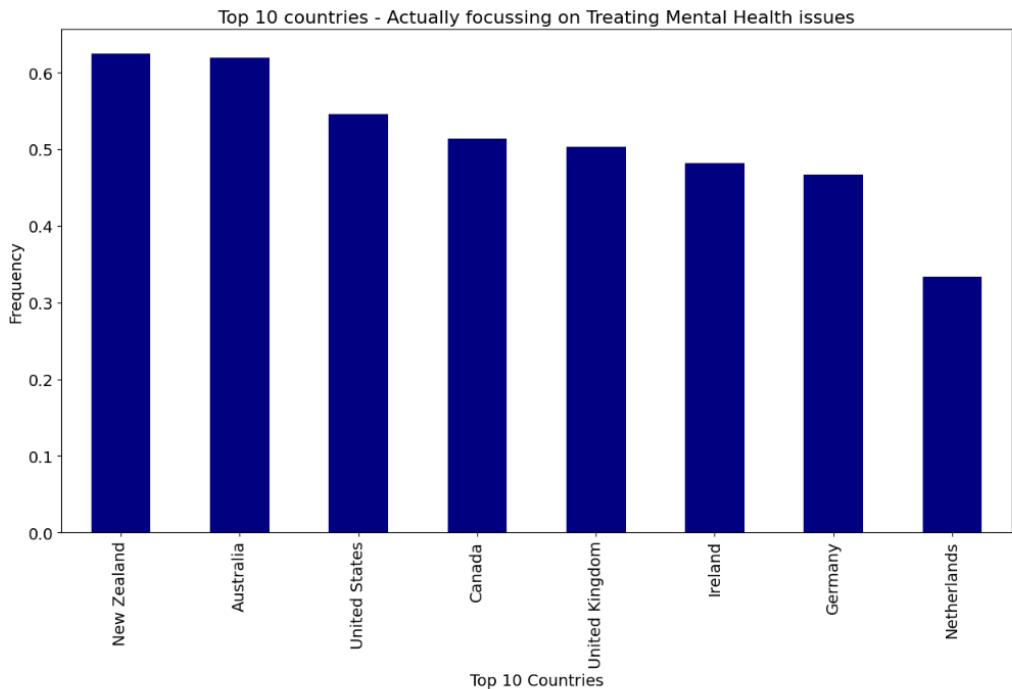


Figure 10: Top Countries Focusing Mental Health Wellness

According to the data, the top countries for both categories are the United States, the United Kingdom, Canada, and Germany, which treat a large number of people with mental health problems while also having the greatest incidence of untreated mental health cases. The inconsistency of our assertion is shown by this statistical contradiction.

To resolve the aforementioned paradox, let's conduct a comprehensive analysis focusing on the data distribution, specifically examining countries that meet the condition where Treatment equals 'Yes' out of the total values recorded.

3.2.2. Age vs Gender Distribution of Tech Company Employee

Let's calculate the ratio of observations from countries addressing mental health issues to the total number of countries included in the dataset shown in figure 10

These countries prioritize influencing a significant proportion of their population to address mental health issues, considering the total number of reported issues.

New Zealand & Australia top the list, followed by United States & Canada.

- Q. What is the frequency distribution of work interference among employees for the top 3 countries?

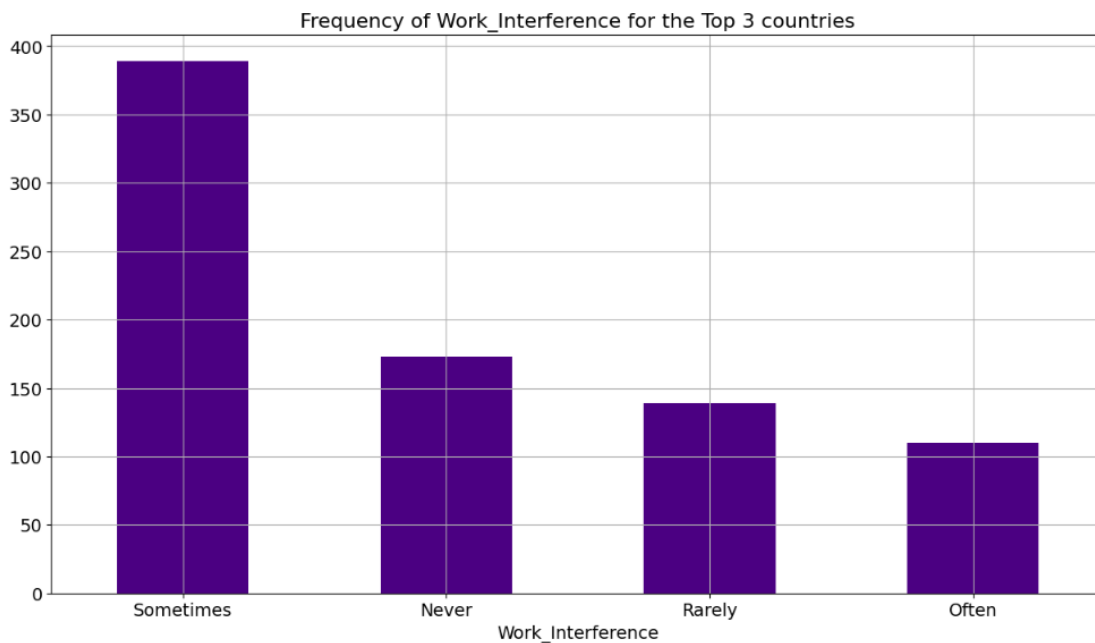


Figure 11: Work Interference Causing Mental Health Treatment

The majority of individuals seeking treatment for their mental health issues experienced interference with their work at times.

- Relation between Treatment and Mental Health Consequence?
mental_health_consequence - Do you think that discussing a mental health issue with your employer would have negative consequences

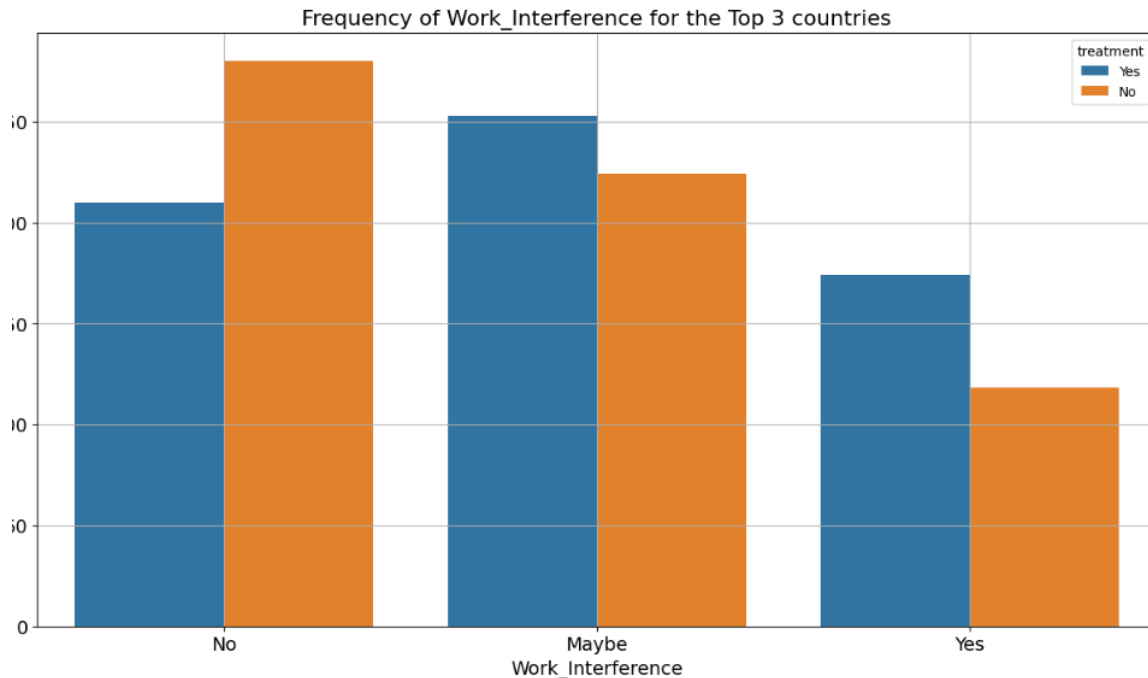


Figure 12: Top Countries of Work Interference

People who fear unfavorable outcomes from talking to their employers about mental health difficulties are more likely to seek help for their problems.

Comparably, people who feel at ease talking to their employers about mental health problems also typically have a decreased readiness to seek help for their worries.

- Q. What is the relationship between mental health consequences and the attitude?

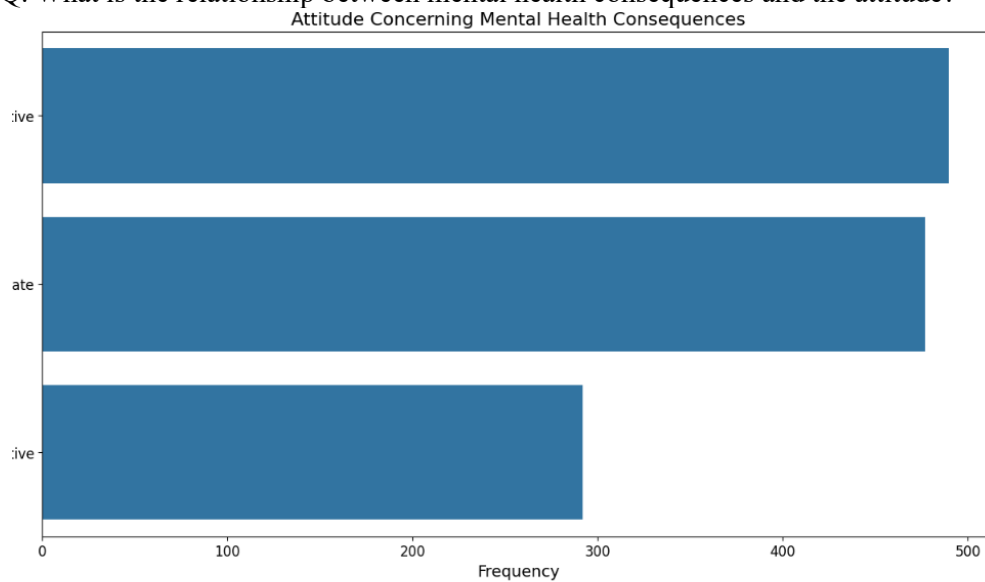


Figure 13: Attitude Affecting Mental Health

The majority of individuals perceive their employers' attitudes to be more positive or moderately supportive rather than negative when addressing their mental health concerns.

- Q. How does age relate to various behaviors and/or their awareness of their employer's attitude

toward mental health?

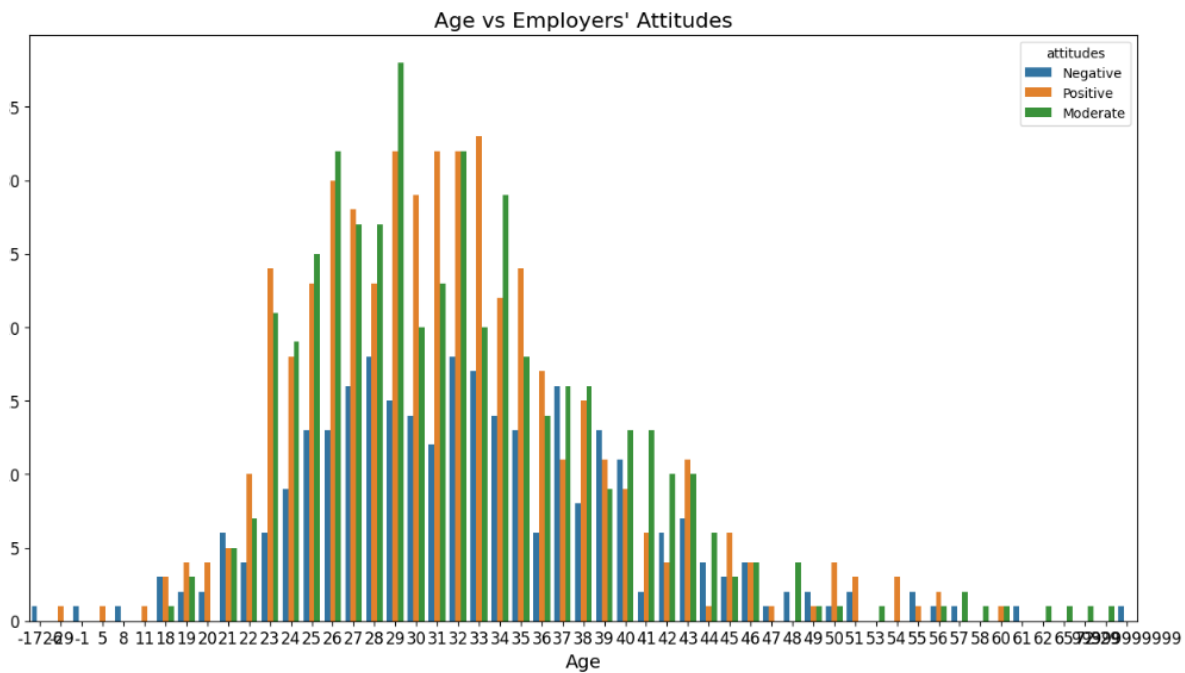


Figure 14: Age vs Employee's Attitude

This suggests that individuals in their mid-20s to mid-30s perceive their employers' attitude to be more positive or moderately supportive rather than negative when they discuss their mental health concerns.

3.2.3. Summarization

The mental health survey has helped us to understand the mental condition of employees working in tech firms across countries.

A total of 1259 entries were recorded during the survey out of which 1007 were recorded from the top 3 countries. The United States leads the chart in terms of participation in the survey followed by the United Kingdom and Canada. 45% of males, 69% of females, and 79% of trans were found to have sought treatment concerning the overall survey.

Likewise, data indicates that 48% of males, 71% of females, and 80% of trans individuals have received treatment within the top three countries in the recorded dataset. The following set of parameters are found to be affecting mental health the most and thus requires treatment:

- Age
- Family history,
- Work Interference,
- Number of employees working in a company,

New Zealand and Australia lead in prioritizing the resolution of employees' mental health issues, encouraging a higher number of individuals to seek treatment, followed by the United States and Canada.

The data shows a prominent peak occurring between the mid-20s to about mid-30s, indicating that the majority of individuals fall within this age range.

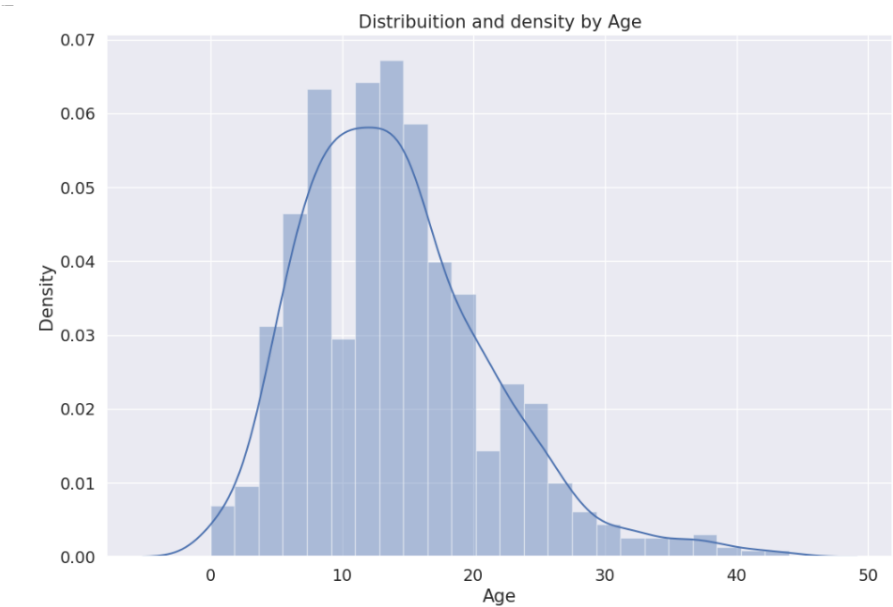


Figure 15: Data Distribution and Density by Age

Age groups demonstrate heightened awareness of their mental health. The dataset parameters and their graphical representations are displayed here.

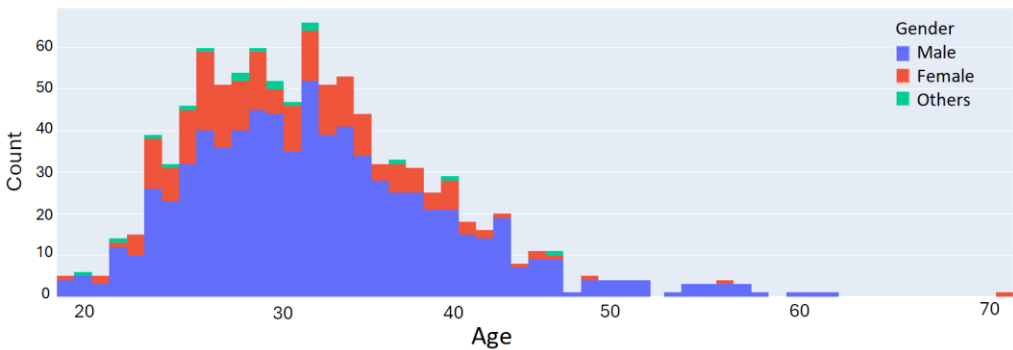


Figure 16: Mental Health Treatment by Age Group

3.2.4. Treatment by Age

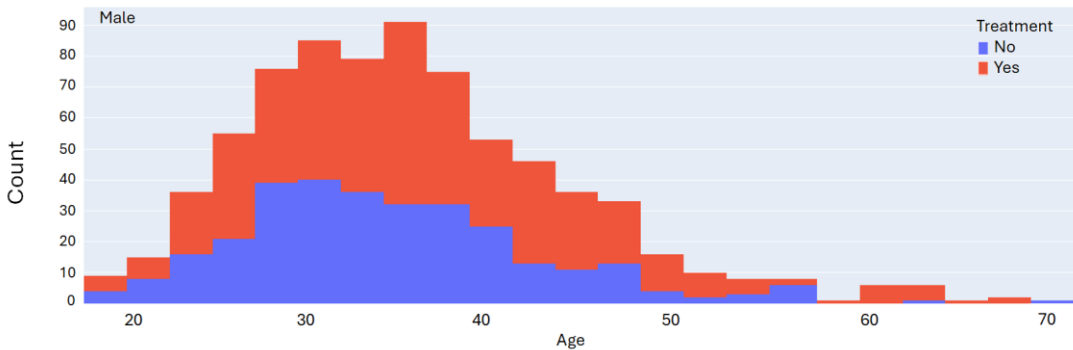


Figure 17(a): Frequency Distribution of Males Getting Treatment vs not-getting Treatment

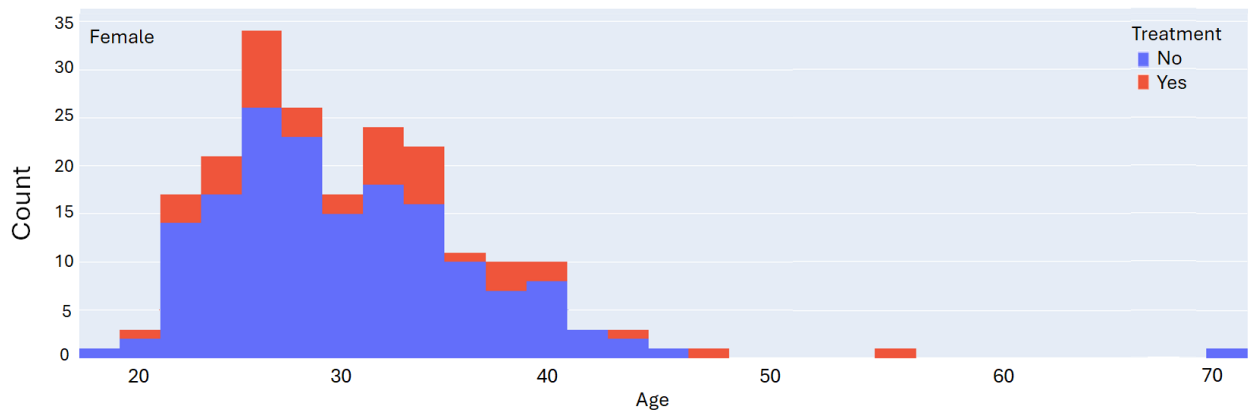


Figure 17(b): Frequency Distribution of Females Getting Treatment vs not-getting Treatment

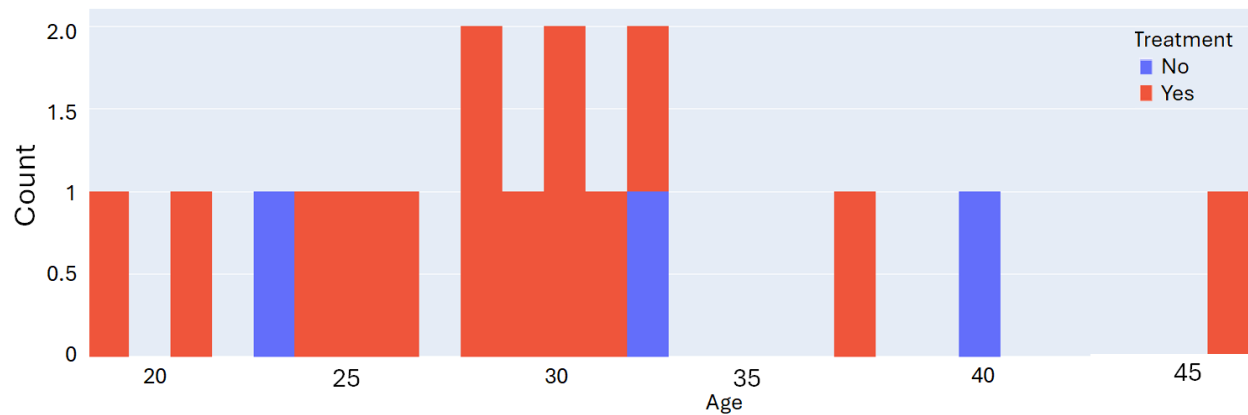


Figure 18: Age-wise Frequency Distribution of Other Gender Getting vs not-getting Treatment

Analyzing the proportions, it suggests that individuals over the age of 30 are addressing their mental health concerns.

3.2.5. Wellness Program by Age

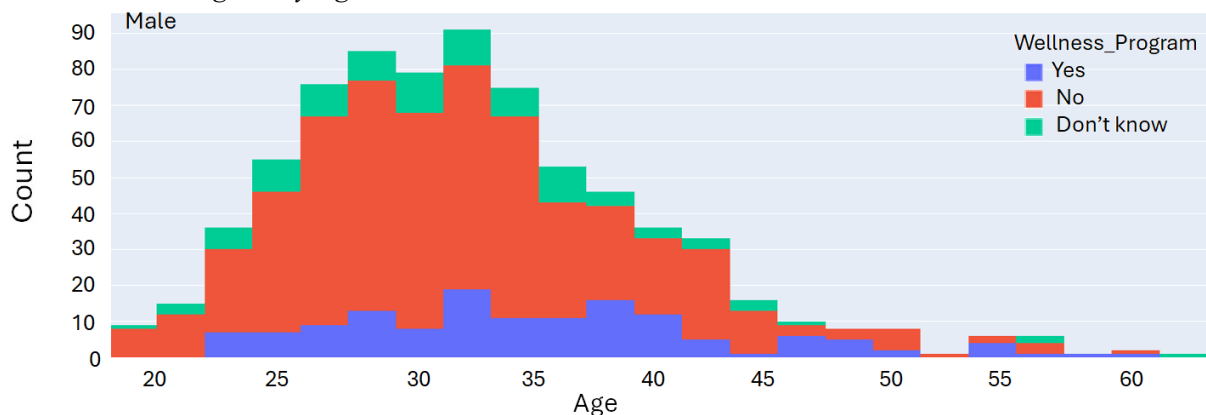


Figure 19: Age-wise Frequency Distribution of Wellness Program of Male

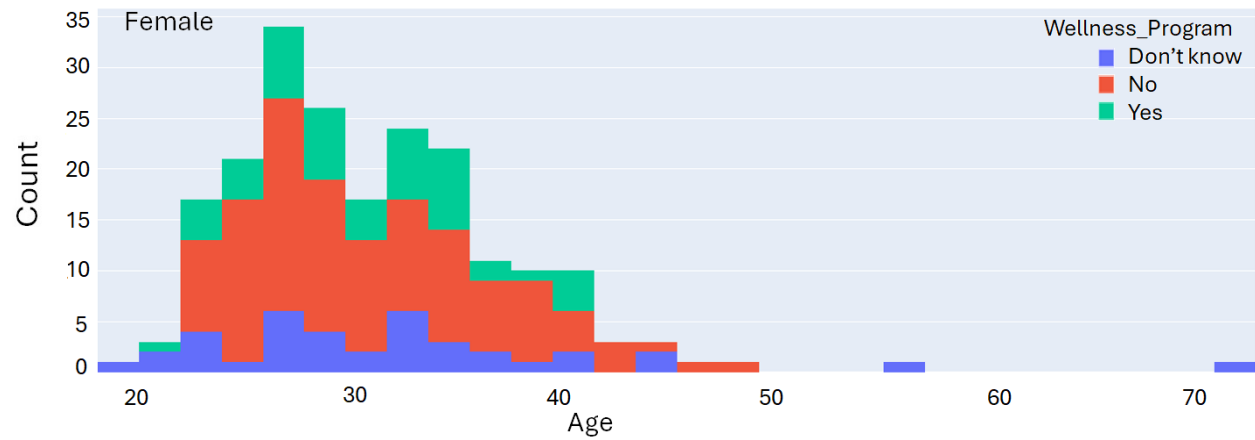


Figure 20: Age-wise Frequency Distribution of Wellness Program of Female

3.3. Feature Selection

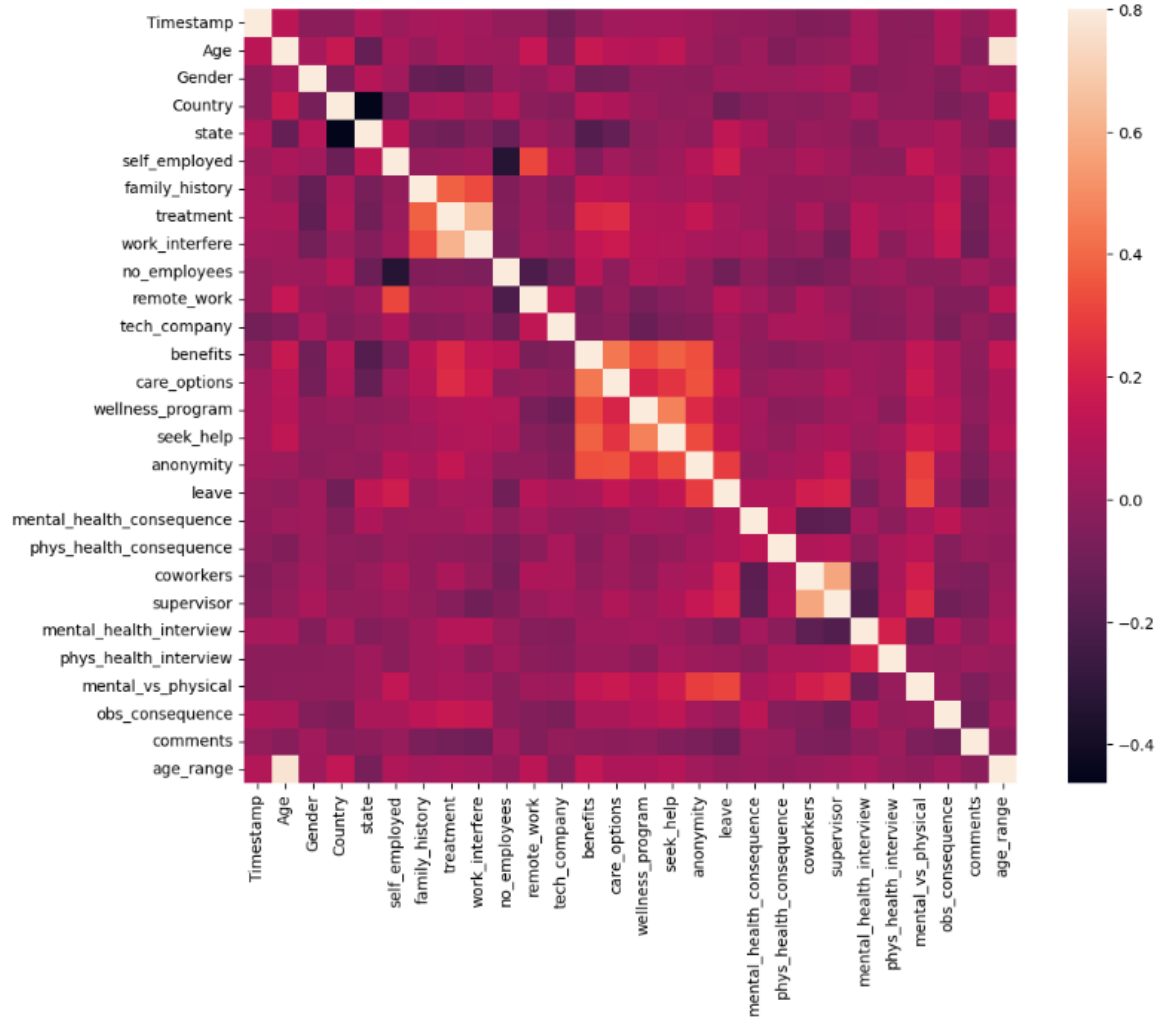


Figure 21 (a): Correlation Matrix of all Features in the Dataset

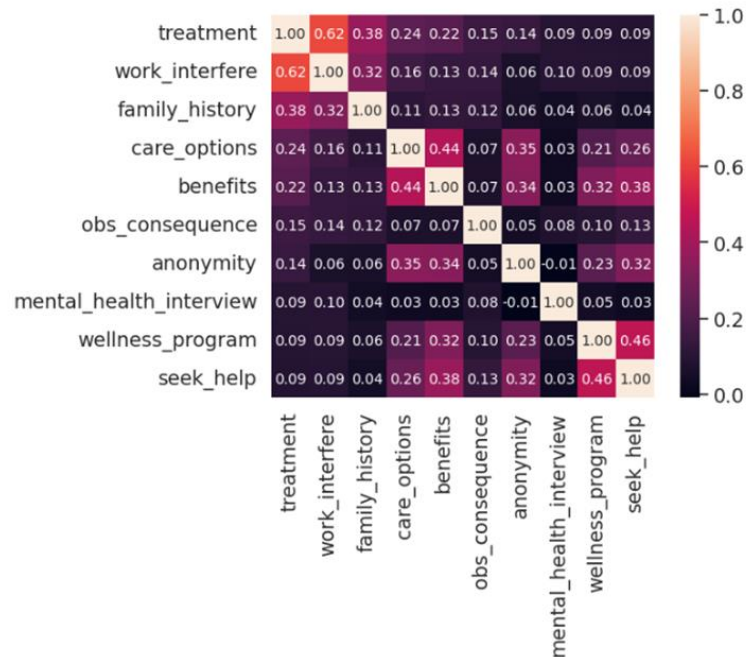


Figure 21 (b): Correlation Matrix of Key Features Related to Mental Health in the Tech Industry

The feature correlation matrix provides is shown in figure 21. it provides insights into the relationships between key features that influence mental health trends in the tech industry.

Based on correlation matrix, and PCA graph, shown in figure 22, features highly correlated with the target variable (mental health treatment) were prioritized while ensuring minimal multicollinearity. The selected features graph is shown in figure 23.

3.3.1. PCA

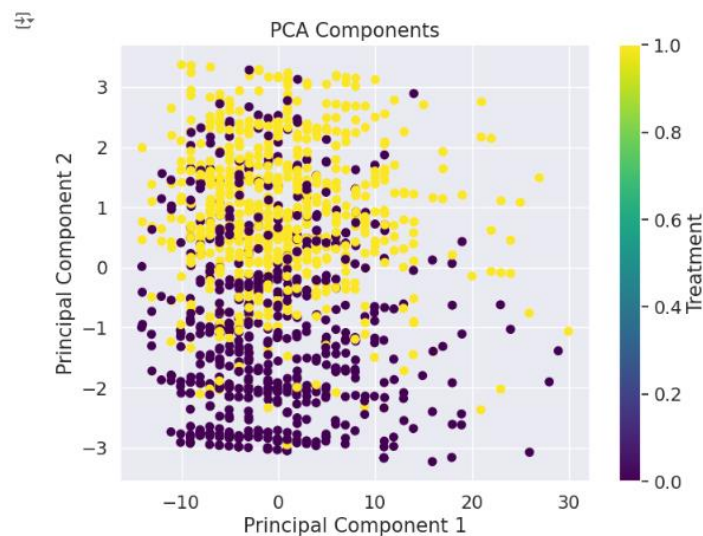


Figure 22: PCA

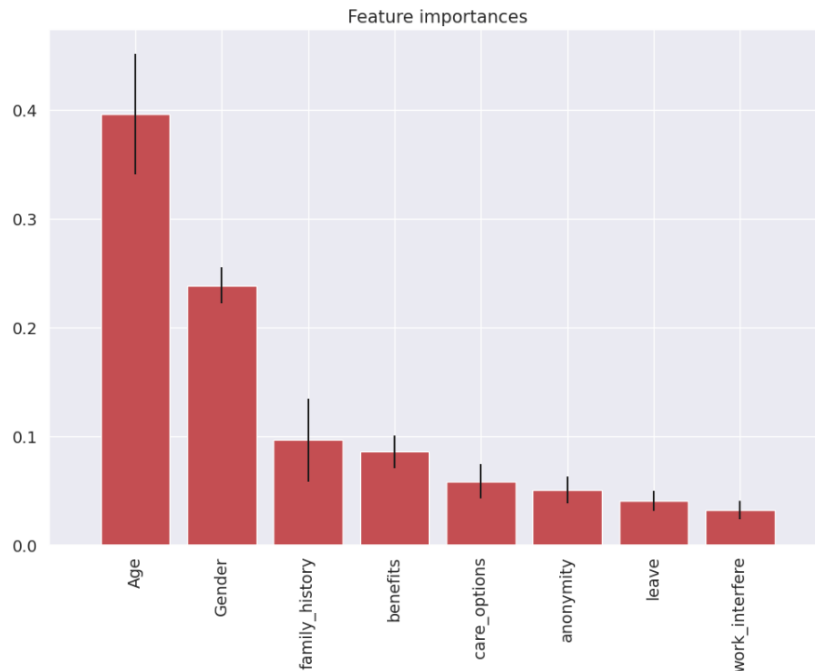


Figure 23: Features Importance on Target Variable, i.e., Mental Health in the Tech Industry

3.4. Model Selection and Training

To determine if a person has sought treatment for a mental health issue, four categorization models were trained and assessed. Logistic regression, K-Nearest Neighbors (KNN), Random Forest Classifier (RFC), and Naive Bayes were the models that were employed. AUC Score, Cross-validated AUC, False Positive Rate, Precision, Classification Accuracy, and Classification Error were among the evaluation criteria.

3.4.1. Logistic Regression

Achieved a classification accuracy of 79.37%, a classification error of 20.63%, and a false positive rate of 25.65%. The precision was 76.33%, with an AUC score of 79.42% and a cross-validated AUC of 87.52%.

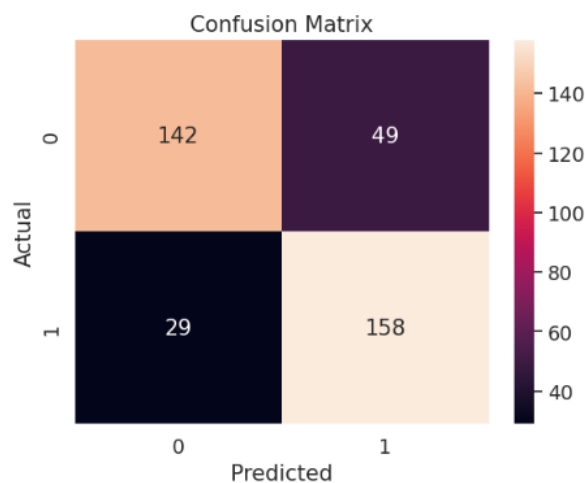


Figure 24: Confusion Matrix of Logistic Regression Model

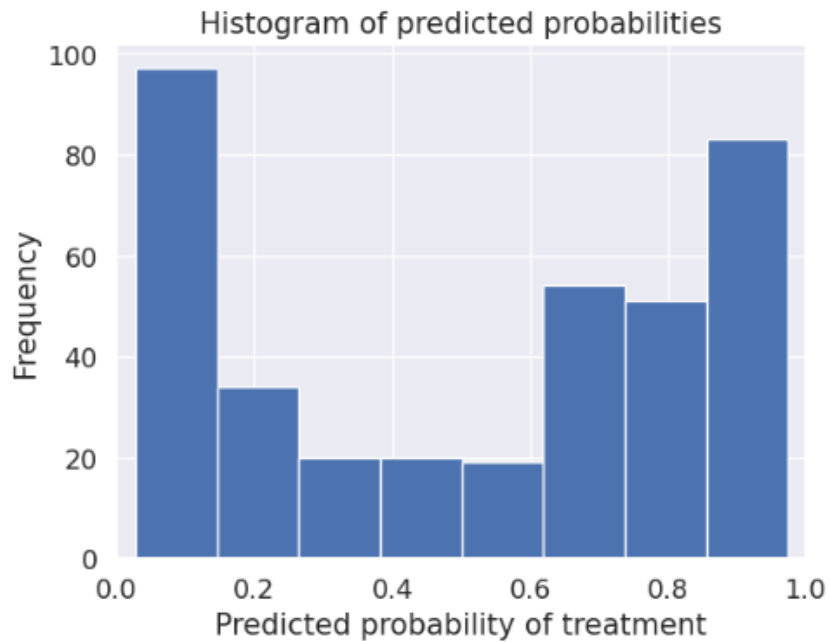


Figure 25: Histogram of Probability of Treatment

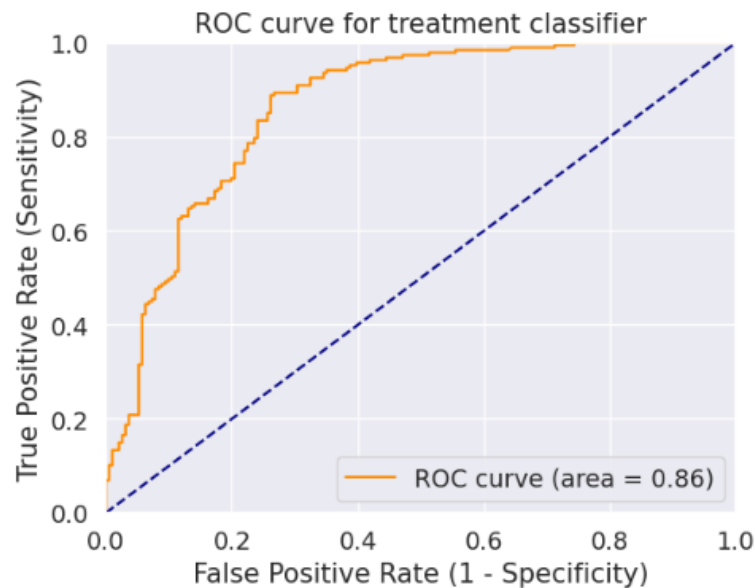


Figure 26: Logistic Regression ROC curve of treatment

3.4.2. Random Forest Classifier

Figure 27 demonstrated a classification accuracy of 81.22%, a classification error of 18.78%, and a false positive rate of 30.37% of RFC. The precision was 75.00%, with an AUC score of 81.34% and a cross-validated AUC of 89.33%.

Classification Accuracy: 0.8121693121693122
 Classification Error: 0.1878306878306878
 False Positive Rate: 0.3036649214659686
 Precision: 0.75
 AUC Score: 0.8134081809782457
 Cross-validated AUC: 0.893325664906046

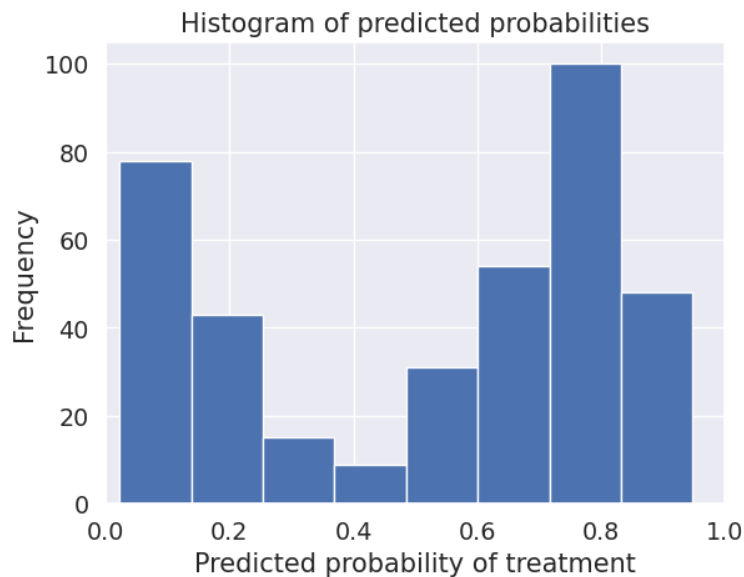


Figure 27: Classification Report and Histogram of RFC

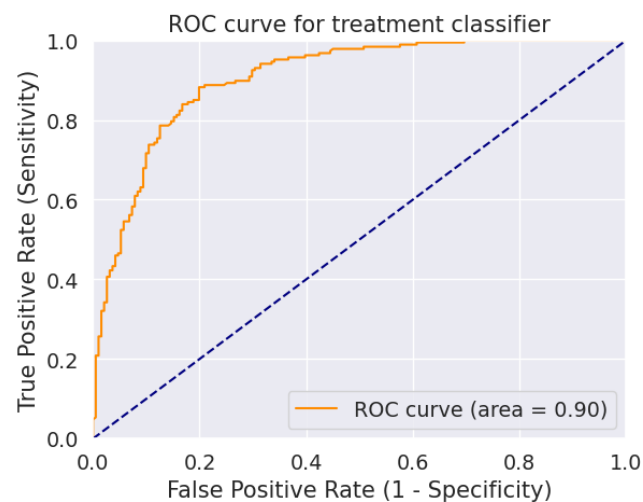


Figure 28: RFC ROC curve of treatment

3.4.3. KNN

Figure 29 shows classification report and confusion matrix of KNN mode. The accuracy of the model is 74.60%, a classification error of 25.40%, and a false positive rate of 30.89%. The precision was 71.77%, with an AUC score of 74.66% and a cross-validated AUC of 83.77%.

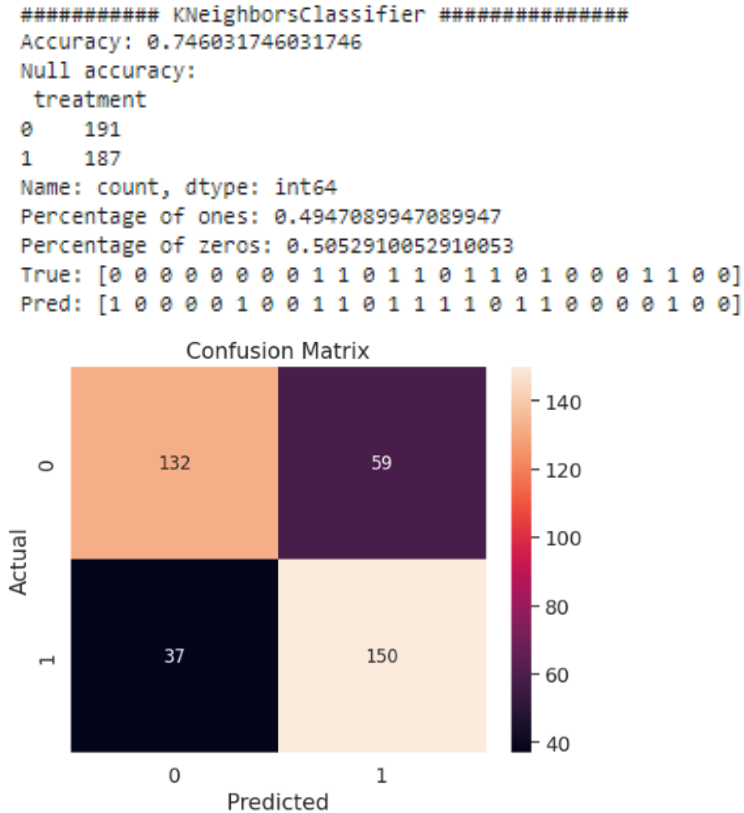


Figure 29: Classification Report and Confusion Matrix of KNN

```
Classification Accuracy: 0.746031746031746
Classification Error: 0.25396825396825395
False Positive Rate: 0.3089005235602094
Precision: 0.7177033492822966
AUC Score: 0.7466192569364729
Cross-validated AUC: 0.8377240683086258
```

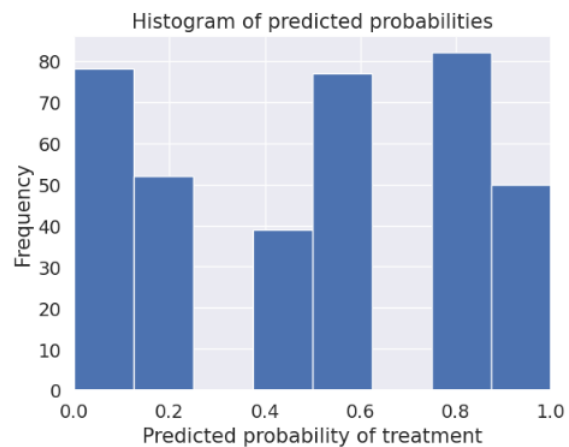


Figure 30: Histogram of Predicted Probabilities of KNN Model

3.4.4. Naive Bayes

Figure 31 showed a classification accuracy of 82.54%, a classification error of 17.46%, and a false positive rate of 19.90%. The precision was 80.71%, with an AUC score of 82.57% and a cross-validated AUC of 88.57%.

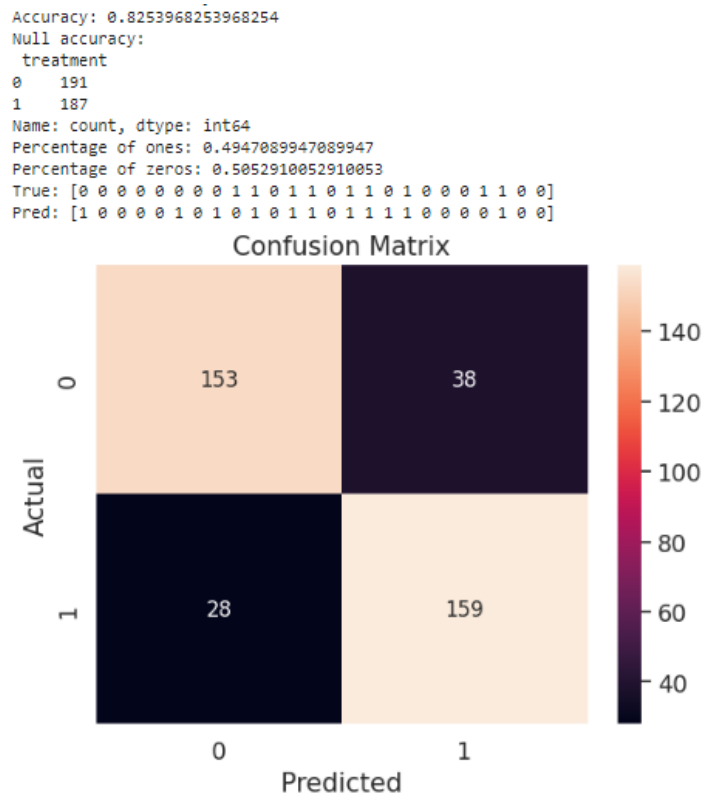


Figure 31: Classification Report and Confusion Matrix of Naïve Bayes

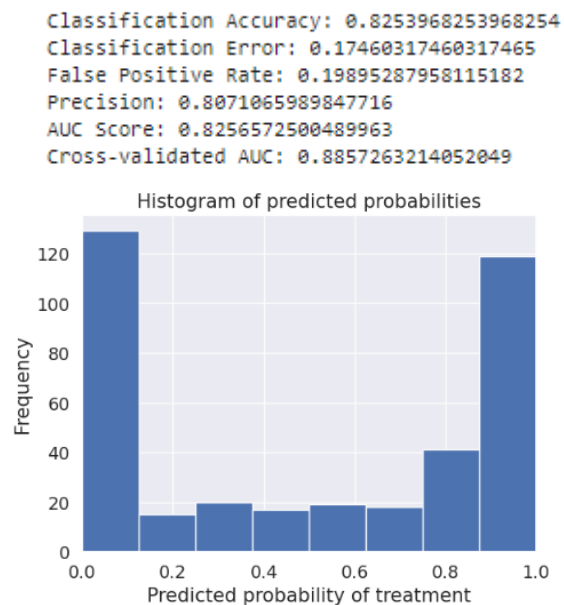


Figure 32: Histogram of Predicted Probabilities of Naïve Bayes Model

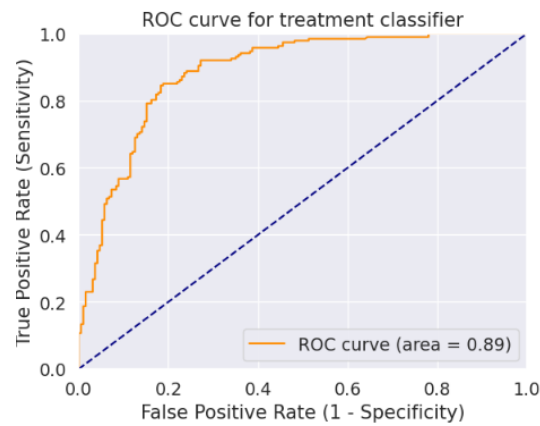


Figure 33: Naïve Bayes ROC Curve of Treatment

Using the classification report and confusion matrix supplied by scikit-learn, we assessed how well each model performed. The measures show how well the models separated "normal" from "anomalous" events. The results show that some models could be quite efficient in terms of computation as well as detection rates.

3.5. Performance Metrics

The model evaluation utilizes a confusion matrix together with a classification report to determine their performance levels. The evaluation measures give crucial information about model accuracy and precision as well as recall performance which helps determine the models' ability to detect intrusions.

4. Results and Discussion

Model performance is compared by us using several evaluation criteria including accuracy, recall, and precision. The results imply that some models could outperform others in terms of detection rates and efficient use of computational resources.

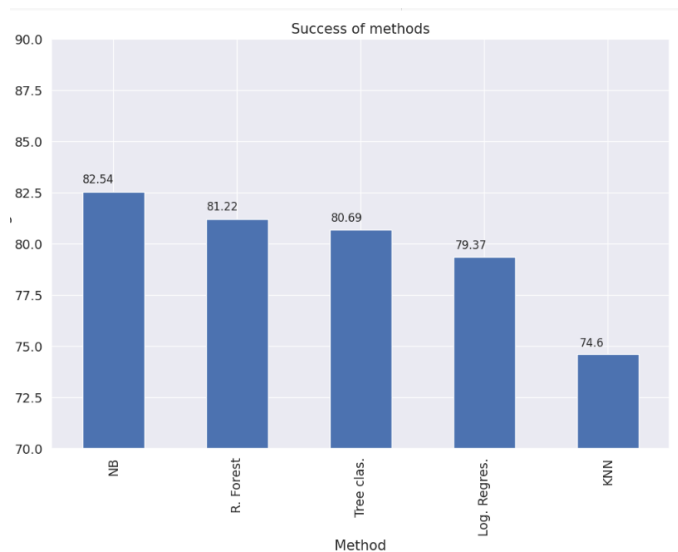


Figure 34: Classification Accuracy Results of ML Classifiers

5. Conclusion

Our study found that the Naive Bayes classifier was the best performer with an outstanding classification accuracy of 82.54% and the lowest classification error of 17.46%. The findings revealed an AUC score of 82.57%, a precision of 80.71%, and a strong cross-validated AUC of 88.57%. Conversely, the Random Forest Classifier (RFC) performed well with a classification accuracy of 81.22%, a precision of 75.00%, and the highest cross-validated AUC of 89.33%. Though the Random Forest Classifier showed a good balance across several measures, Naive Bayes had the most accuracy and proved to be a consistent predictor of whether tech sector personnel would seek treatment for mental health concerns. Its remarkable accuracy, precision, and AUC ratings help the Random Forest Classifier to significantly excel in this prediction challenge. Though both models show good promise, the Random Forest Classifier distinguishes itself with its steady and reliable performance.

References

- [1] WHO Team Mental Health, Brain Health and Substance Use (MSD), "Depression and Other Common Mental Disorders," *World Health Organization*, 2017. Available: <https://www.who.int/publications/i/item/depression-global-health-estimates>.
- [2] "Making the Investment Case for Mental Health : A WHO/UNDP Methodological Guidance Notes" *World Health Organization*, 2019, Available: <https://iris.who.int/bitstream/handle/10665/325116/WHO-UHC-CD-NCD-19.97-eng.pdf>
- [3] Goetzel, Ron Z., Enid Chung Roemer, Calliope Holingue, M. Daniele Fallin, Katherine McCleary, William Eaton, Jacqueline Agnew et al. "Mental health in the workplace: a call to action proceedings from the Mental Health in the Workplace—Public Health Summit." *Journal of occupational and environmental medicine* 60, no. 4 (2018): 322-330.
- [4] Chang, Josh, Felix Peysakhovich, Weimin Wang, and Jin Zhu. "The UK health care system." *United Kingdom* 30 (2011): 2019.
- [5] Heffernan, Margaret, and Tony Dundon. "Cross-level effects of high-performance work systems (HPWS) and employee well-being: the mediating effect of organisational justice." *Human Resource Management Journal* 26, no. 2 (2016): 211-231.
- [6] Dewanto, Wahyu, and Sofia Retnowati. "Intervensi kebersyukuran dan kesejahteraan penyandang disabilitas fisik." *Gadjah Mada Journal of Professional Psychology (GamaJPP)* 1, no. 1 (2015): 33-47.
- [7] Tanenbaum, A., Wetherall, D. *Computer Networks (5th Edition)*. Pearson, 2010.
- [8] Shatte, Adrian BR, Delyse M. Hutchinson, and Samantha J. Teague. "Machine learning in mental health: a scoping review of methods and applications." *Psychological medicine* 49, no. 9 (2019): 1426-1448.
- [9] Tran, Truyen, Dinh Phung, Wei Luo, Richard Harvey, Michael Berk, and Svetha Venkatesh. "An integrated framework for suicide risk prediction." In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1410-1418. 2013.
- [10] Buczak, Anna L., and Erhan Guven. "A survey of data mining and machine learning methods for cyber security intrusion detection." *IEEE Communications surveys & tutorials* 18, no. 2 (2015): 1153-1176.
- [11] Mittravinda, K. M., Devika S. Nair, and Gowri Srinivasa. "Mental health in tech: Analysis of workplace risk factors and impact of covid-19." *SN computer science* 4, no. 2 (2023): 197.
- [12] Wood, A. M., Maltby, J., Gillett, R., Linley, P. A., & Joseph, S.. The Role of Gratitude in the Development of Social Support, Stress, and Depression: Two Longitudinal Studies. *Journal of Research in Personality*, 42, (2019) 854-871.
- [13] Farrer, Louise, Amelia Gulliver, Jade KY Chan, Philip J. Batterham, Julia Reynolds, Alison Calcar, Robert Tait, Kylie Bennett, and Kathleen M. Griffiths. "Technology-based interventions for mental health in tertiary students: systematic review." *Journal of medical Internet research* 15, no. 5 (2013): e2639.
- [14] Torous, John, Keris Jān Myrick, Natali Rausero-Ricupero, and Joseph Firth. "Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow." *JMIR mental health* 7, no. 3 (2020): e18848.
- [15] Kumar, Koushal, and Jaspreet Singh Batth. "Network intrusion detection with feature selection techniques using machine-learning algorithms." *International Journal of Computer Applications* 150, no. 12 (2016).