Predicting Mental Health Service Utilization Among Working Professionals

Author : Aman Prajapati

Dataset : "Mental Health in Tech Survey" from Kaggle, created by OSMI (Open Sourcing Mental Illness)

Introduction

Mental health has become an important topic in today's workplaces, especially with rising levels of stress, burnout, and pressure. While awareness is improving, many professionals still hesitate to seek help sometimes because of stigma, sometimes because of a lack of support at work.

This project started with a simple but important question:
"Can we predict which working professionals are likely to seek mental health treatment?"

To answer this, I used real-world survey data from people in the tech industry. The dataset included questions about their age, gender, work environment, and mental health experiences. By analyzing these patterns and building a machine learning model, the goal was to understand the key factors that influence someone's decision to get help and how workplaces can create better support systems.

This isn't just about numbers and predictions. It's about using data to spark change, reduce stigma, and help companies take mental health more seriously.

Overview of the Data

The data used in this project comes from a public survey conducted by OSMI (Open Sourcing Mental Illness), which focuses on mental health in the tech industry. The survey collects anonymous responses from working professionals around the world and includes a wide range of questions not just about mental health, but also about their workplace environment, personal background, and whether they've ever sought treatment.
Here's what the dataset generally looks like:

- Demographic Information
  Includes age, gender, and country. This helps us understand how different groups perceive and handle mental health.

- Workplace-Related Questions
  Asks whether the respondent works remotely, works for a tech company, and whether their employer offers mental health benefits or support options.
- Mental Health History & Experience
  Covers whether the person has a family history of mental illness, whether their mental health affects their work, and if they've ever sought treatment.
- Target Variable
  The main column we're interested in is treatment, which tells us whether the individual has sought mental health treatment in the past - either "Yes" or "No". This is what we'll try to predict.

In total, the dataset includes hundreds of responses across several columns, and each row represents a different individual's experience. What makes this dataset powerful is that it blends both personal and professional aspects -giving us a more complete picture of what influences someone's decision to seek help.

Dataset link : https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey/data?select=survey.csv

Methodology

Data Preprocessing & Data Cleaning

When I first looked at the survey data, it was clear that some cleanup was needed before doing any analysis. There were missing values in several columns, inconsistent ways of writing things (like gender), and some age entries that didn't seem realistic. For example, the gender column had responses like "cis female", "F", "female (trans)", and even typos — so I grouped them into just three categories: Male, Female, and Unidentified. I also grouped ages into ranges (like under 25, 25–34, etc.) to make patterns easier to spot. For missing values, I used a smart strategy: if a column had just a few missing entries, I filled them in with common values; but if too much data was missing, I either handled it carefully or dropped the column altogether. This step helped ensure that the data going into the model was clean, consistent, and ready for meaningful analysis.

Exploratory Data Analysis

Once the data was cleaned, I spent time exploring it to understand the bigger picture and uncover patterns. I wanted to know — who's more likely to seek mental health treatment? So I looked at how treatment rates varied by age group, gender, family history, remote work, and region. Some things stood out right away: people with a family history of mental illness were much more likely to seek treatment. Younger professionals, especially those under 35, were also more open to getting help. Another interesting pattern was that remote workers were more likely to seek treatment, possibly because they feel more comfortable doing so in a flexible work environment. I also explored how company support (like benefits or HR policies) impacted decisions — and the data showed that lack of support often went hand-in-hand with not seeking help. Visualizing these trends helped build context before jumping into model-building, and confirmed that both personal and workplace factors play a big role in mental health behavior.

Feature Engineering

After exploring the data, I created new features to help the model better understand what might influence someone's decision to seek mental health treatment. These weren't just random columns — they were based on real-world logic. For example, I created a has_support feature that checks if a person has access to mental health resources at work, and a has_benefits feature to capture whether their employer offers mental health benefits. I also added a remote_worker flag to indicate if someone works remotely. To make geographic patterns more meaningful, I grouped individual countries into broader regions like North America, Europe, and Asia. One of the more interesting features I engineered was an interaction term that multiplies a person's age with whether they have a family history of mental illness — the idea being that someone younger with a family history might be more at risk or more aware. These new features gave the model more useful signals and allowed it to pick up on complex patterns that wouldn't be obvious from the raw data alone.

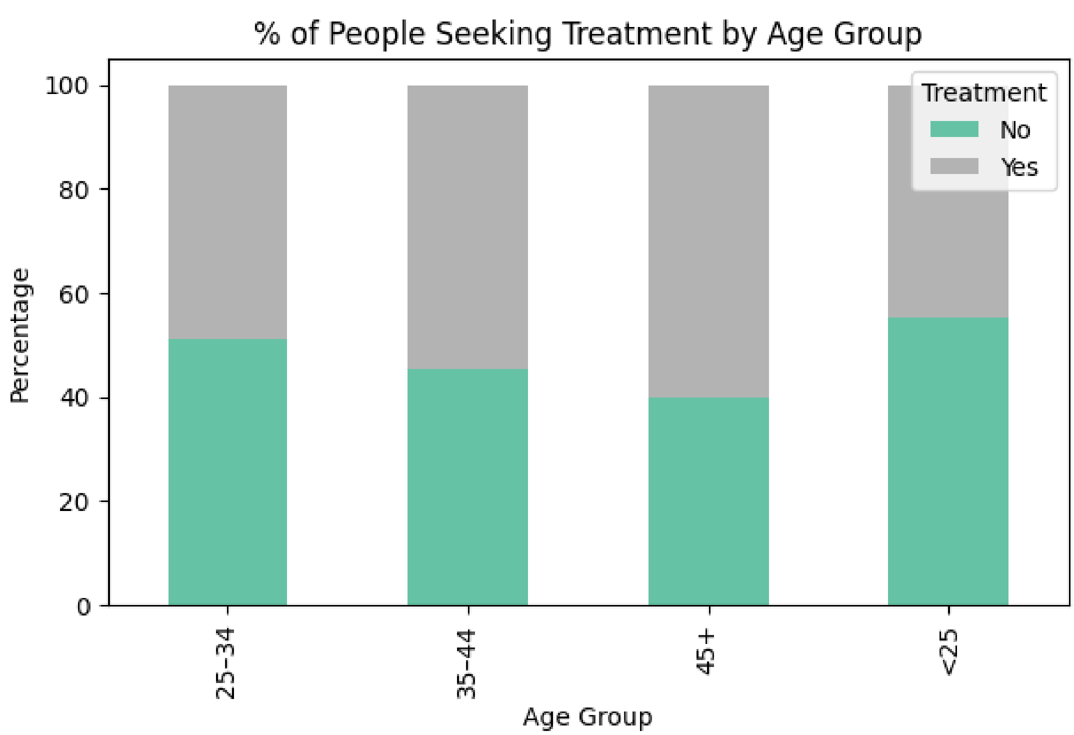
Predictive Modelling

With a cleaned and well-prepared dataset, I moved on to building machine learning models to predict whether someone would seek mental health treatment. Since our target variable (treatment) had only two possible outcomes — "Yes" or "No" — this was a binary classification problem. I trained and tested multiple models, including Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine (SVM). Each model was evaluated using key metrics like accuracy, F1-score, and ROC-AUC to understand both how often the model was correct and how well it balanced precision and recall. Among all models, Gradient Boosting performed the best overall, especially in identifying those who were likely to seek help. To make sure the models weren't just making predictions blindly, I used SHAP values to explain which features mattered most. This made the model's decision-making process more transparent and helped connect the predictions back to real human and workplace factors.

To make the data ready for machine learning, I had to convert some of the text-based columns into numbers — because models can't understand words like "Yes", "No", or "Sometimes" directly. This is where one-hot encoding came in. It's a technique that transforms each category into its own binary column — using 1s and 0s to indicate presence. For example, instead of keeping a single column called remote_work with values like "Yes" or "No", I created two separate columns like remote_work_Yes and remote_work_No. The same was done for features like gender, region, family history, and workplace support. This approach allowed the model to process all categories equally and helped avoid giving unfair importance to any one category just because of its label. Overall, one-hot encoding was a crucial step in preparing the data for modeling while preserving the meaning behind categorical responses.

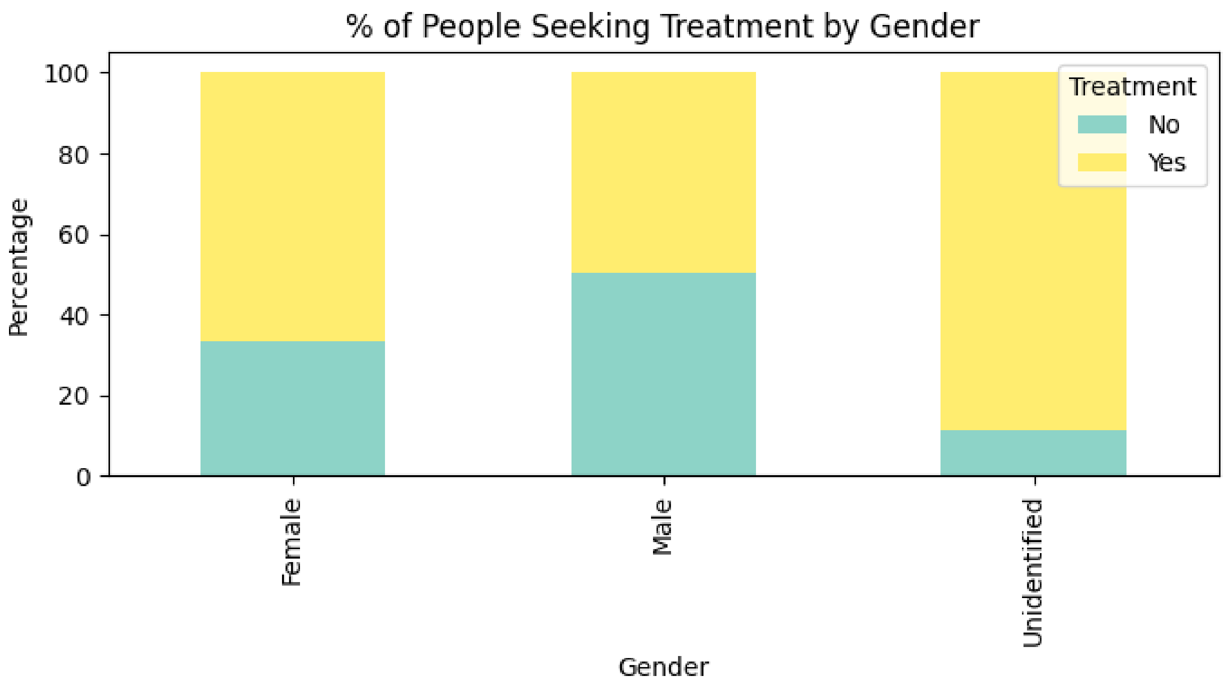
Analysis :

Younger Professionals Are More Open to Seeking Mental Health Treatment



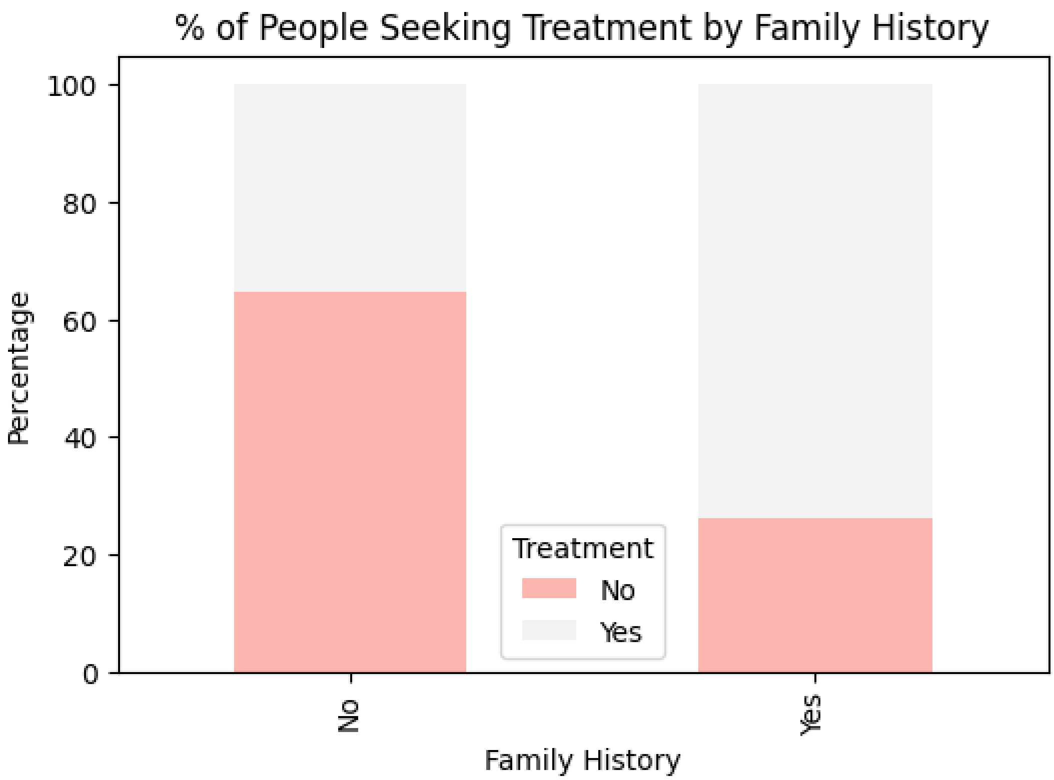
This chart shows the percentage of people in each age group who said “Yes” or “No” to seeking mental health treatment. A clear pattern stands out: younger age groups, especially those under 25 and between 25–34, are more likely to seek help. As age increases, the percentage of people who seek treatment gradually drops. This could be due to changing attitudes — younger professionals may feel more comfortable talking about mental health or have more awareness and access to support. On the other hand, older age groups might still carry stigma or hesitation around getting help. These insights suggest that efforts to normalize mental health conversations are working especially among the newer workforce generations.

Females and Unidentified Genders Show Higher Willingness to Seek Mental Health Help



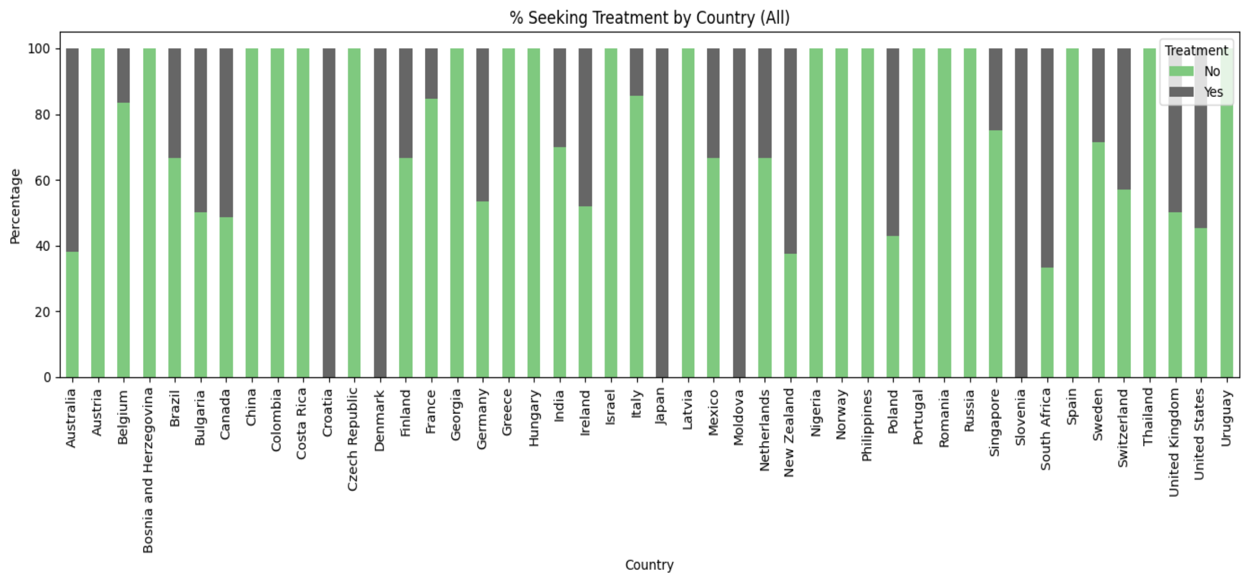
This chart shows how people from different gender groups responded when asked if they had ever sought mental health treatment. The results suggest that female respondents were more likely to seek help compared to male respondents, and those who identified outside the traditional binary (Unidentified) showed the highest treatment-seeking behavior overall. In contrast, a larger portion of male respondents said they had not sought treatment. These patterns could reflect deeper social factors — such as men feeling more stigma or pressure to "stay strong," while other gender groups may feel more open or aware of mental health resources. This highlights the importance of tailoring mental health support to different groups and creating safe, stigma-free spaces for everyone**.**

Having a Family History of Mental Illness Increases the Likelihood of Seeking Treatment



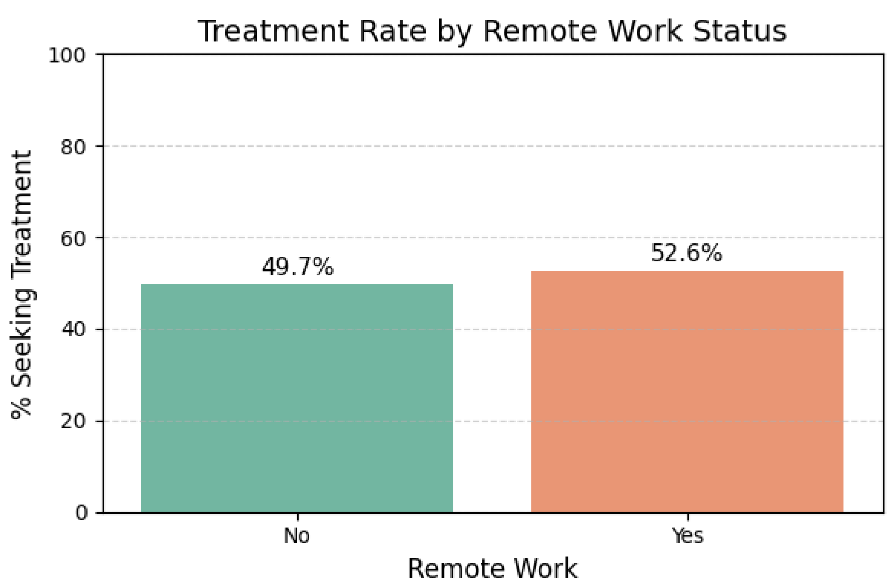
This chart shows a strong pattern: people who reported having a family history of mental illness were much more likely to seek treatment compared to those who didn't. Among respondents with no family history, a large portion said they had not sought help. But for those with a family history, the majority did seek treatment. This makes sense — having someone close go through a mental health issue can make a person more aware, less hesitant, and more comfortable reaching out for support themselves. It also shows how personal experience can shape not just awareness, but action.

Mental Health Treatment-Seeking Behavior Varies Greatly Across Countries



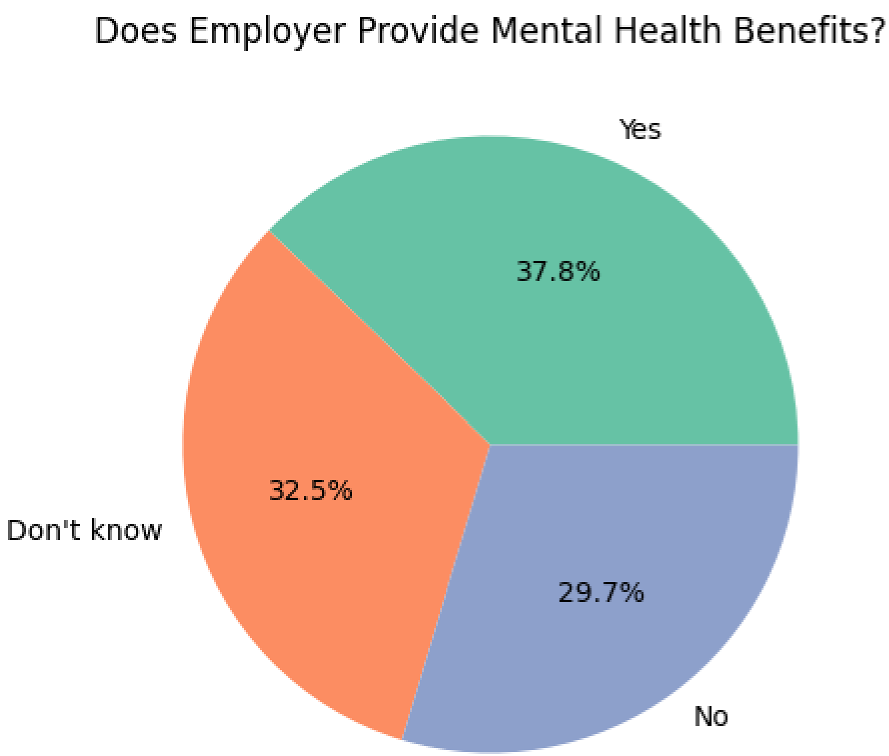
This chart highlights how treatment-seeking behavior differs from country to country. In some countries — like the United Kingdom, the United States, and Belgium — a large portion of respondents said they had sought mental health treatment. But in many other countries, especially where mental health awareness may still be growing, the majority of people said they had not sought help. These differences could be due to a variety of reasons — access to care, cultural beliefs, stigma, or simply how openly people talk about mental health. What's clear is that geography matters: where someone lives can strongly shape their experience with mental health support, and this reinforces the need for region-specific awareness and policies.

Remote Workers Are Slightly More Likely to Seek Mental Health Support



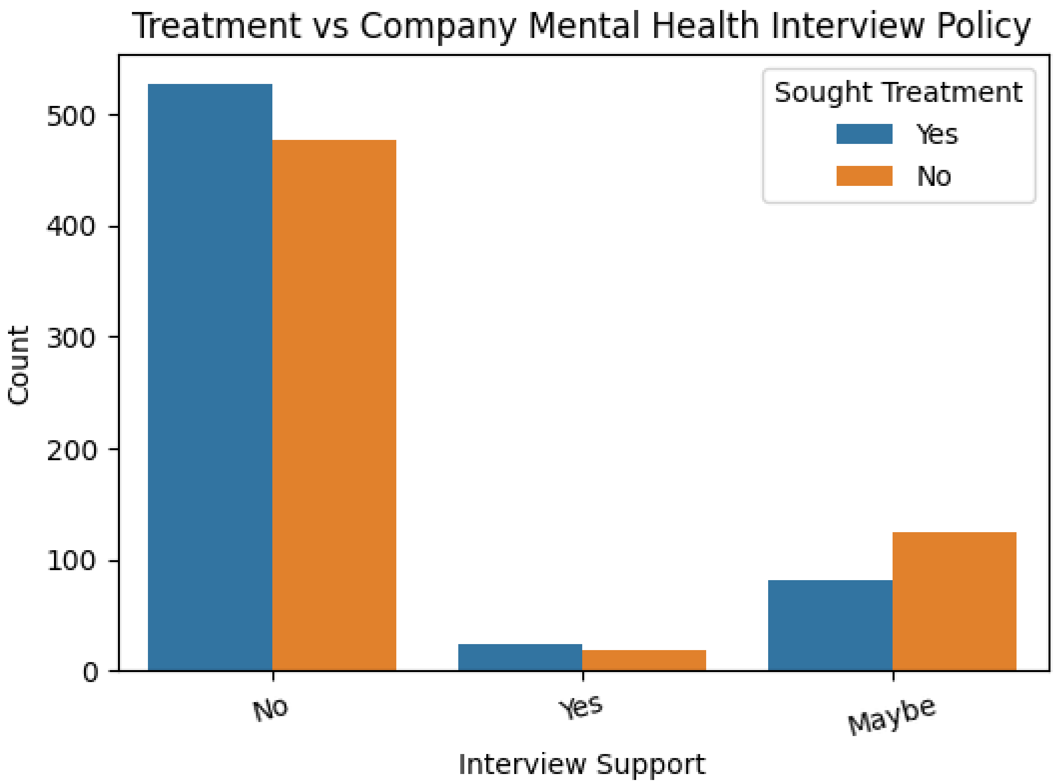
This chart compares how often people sought mental health treatment based on whether they worked remotely. Interestingly, those who work remotely showed a slightly higher treatment rate (around 52.6%) compared to those who don't (49.7%). While the difference isn't huge, it suggests that remote workers may feel a bit more comfortable or flexible when it comes to taking care of their mental health. Working from home could offer more privacy or a better work-life balance, making it easier for individuals to seek help without feeling judged. Even small differences like this can highlight how workplace structure can influence personal decisions.

Many Employees Are Unsure If Mental Health Benefits Exist at Work



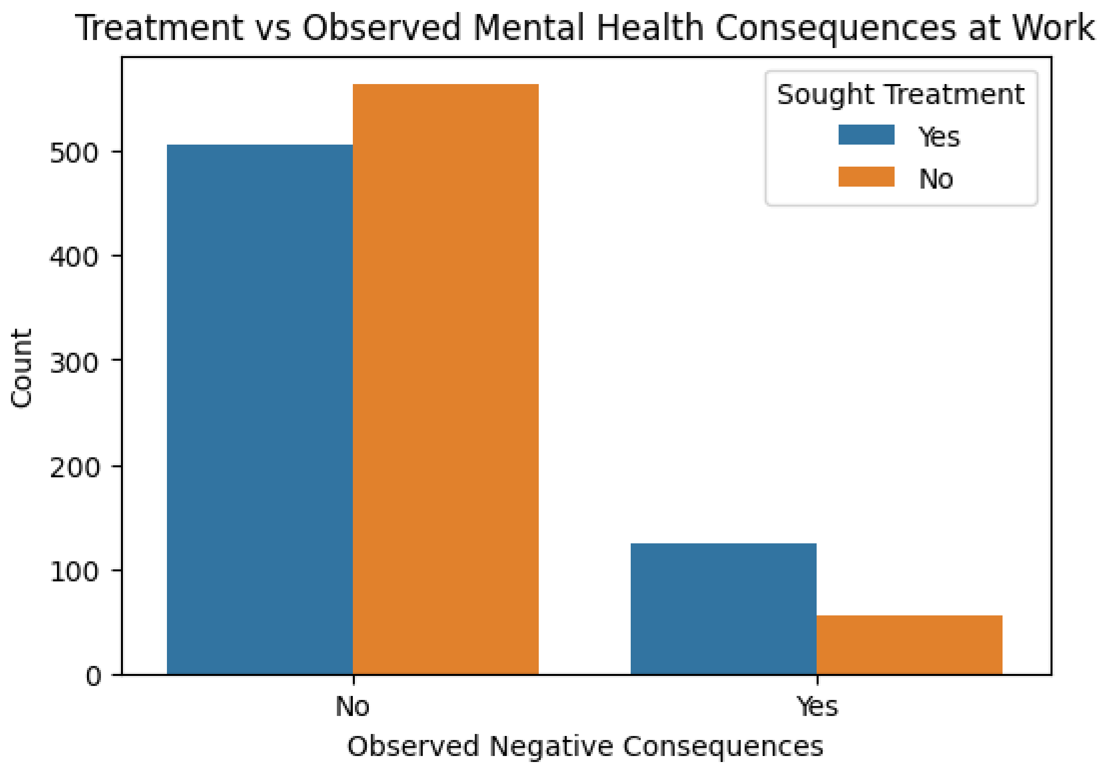
This chart shows how people respond when asked if their employer provides mental health benefits. Surprisingly, only about 38% said "Yes", while nearly 33% weren't even sure if such benefits existed. Around 30% said "No" outright. This means that more than 60% of employees either don't have access to mental health support or aren't aware of it. That's a big gap. It shows that even when companies offer support, they may not be doing enough to communicate it clearly. It also highlights the need for workplaces to talk more openly about mental health policies so that employees feel informed, supported, and empowered to ask for help when they need it.

Most Employees Say Their Company Doesn’t Offer Mental Health Support During Interviews



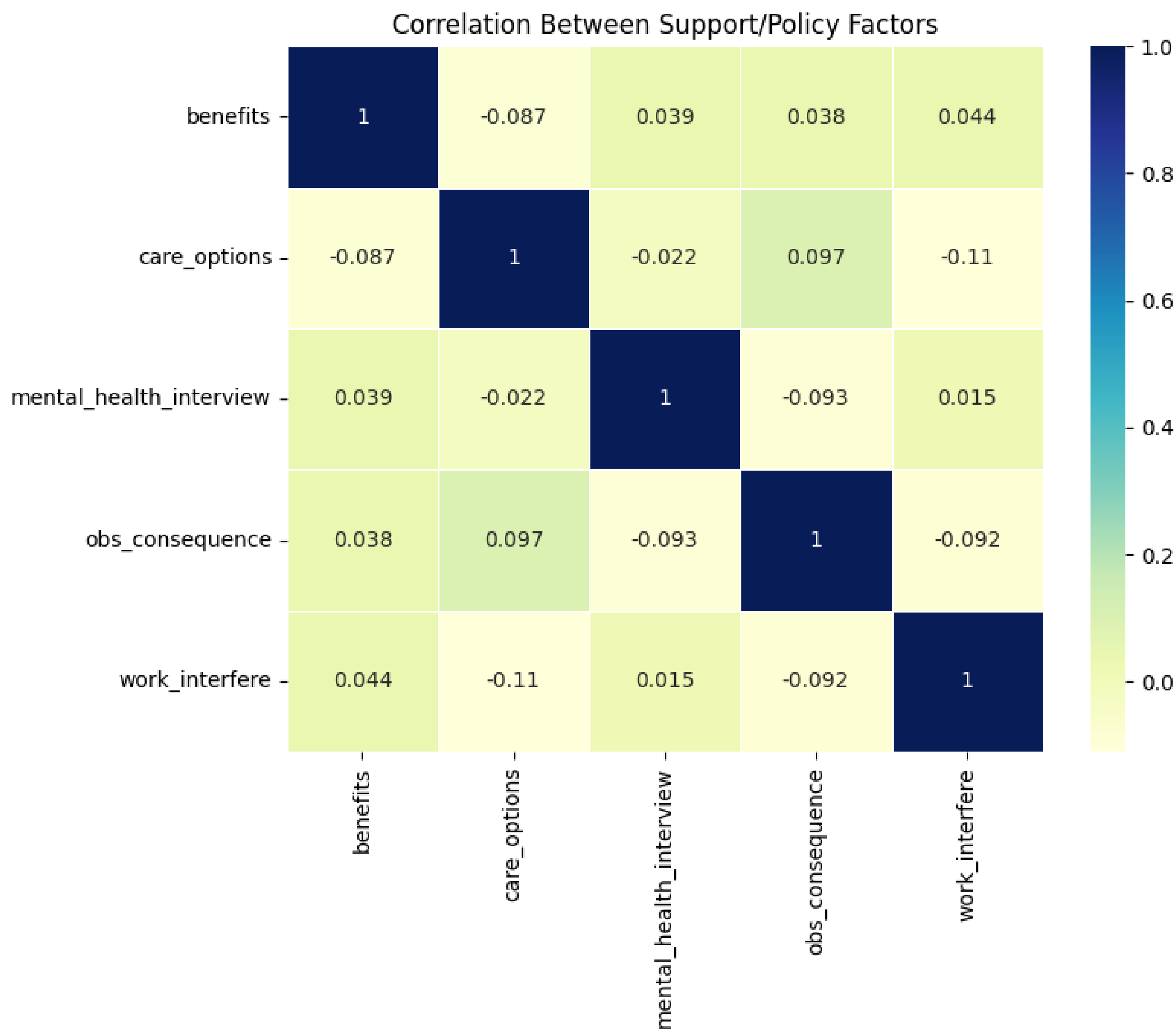
This chart explores whether companies support mental health discussions during the job interview process — and how that relates to whether people seek treatment. The majority of employees said their company does not offer any support or openness around mental health during interviews. In those cases, we still see a mix of people who do and don’t seek treatment. Very few companies clearly said “Yes” to offering support, and even fewer respondents knew for sure. Interestingly, when companies answered “Maybe,” the number of people not seeking treatment slightly increased. This suggests that uncertainty or silence from companies can make employees less likely to speak up or get help. Mental health support shouldn't just be internal — it should start from the hiring process to help build a safe and open work culture from day one.

Seeing Mental Health Discrimination at Work Encourages Some to Speak Up-and Others to Stay Silent



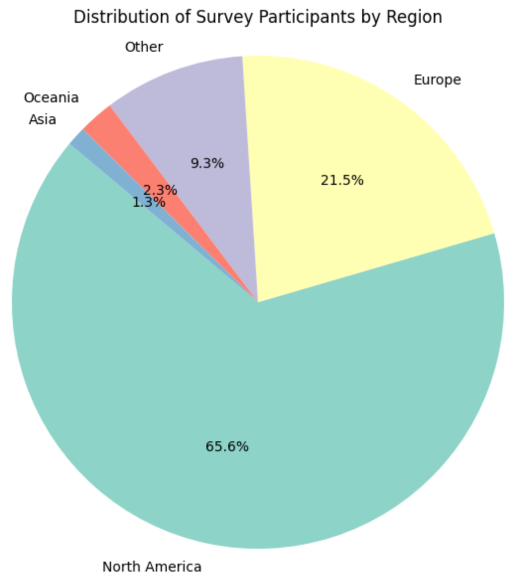
This chart looks at whether people have witnessed negative consequences at work for someone dealing with mental health issues like- being judged, treated unfairly, or denied opportunities and how that affected their decision to seek help. Interestingly, among those who did observe such consequences, more people actually chose to seek treatment than not. This could mean that seeing others struggle might push someone to prioritize their own mental health sooner. On the other hand, among those who did not observe any consequences, we still see a fairly even mix of treatment and non-treatment. This tells us that while a supportive environment matters, people also respond differently to workplace culture some feel encouraged to speak up, while others may stay quiet if they fear judgment.

Support Systems at Work Don’t Always Go Hand-in-Hand — but Every Bit Helps



This heatmap shows how different workplace support and policy factors relate to each other. Interestingly, most of these factors like having mental health benefits, care options, or interview support don’t have strong correlations with each other. In simple terms, just because a company offers mental health benefits doesn't always mean it also has HR interview support or open care options. Each policy seems to operate somewhat independently. While these correlations are small, they still remind us that building a mentally healthy workplace isn’t about one big fix it’s about combining many small efforts. Even if no single support system dominates, every layer of support adds up to create a culture where people feel safer asking for help.

Most Survey Respondents Came from North America — Global Views Were Limited



This chart shows where the people who took the survey are from. The majority of participants — about 66% — were from North America, followed by Europe at around 21%. Other regions like Asia, Oceania, and “Other” made up only a small portion of the responses. This matters because it shows that the insights from this project are mainly shaped by North American and European perspectives. While the findings are still valuable, we should be careful when applying them to other parts of the world where cultural views, access to care, and workplace support for mental health can be very different. It also highlights the need for more globally inclusive mental health data in the future.

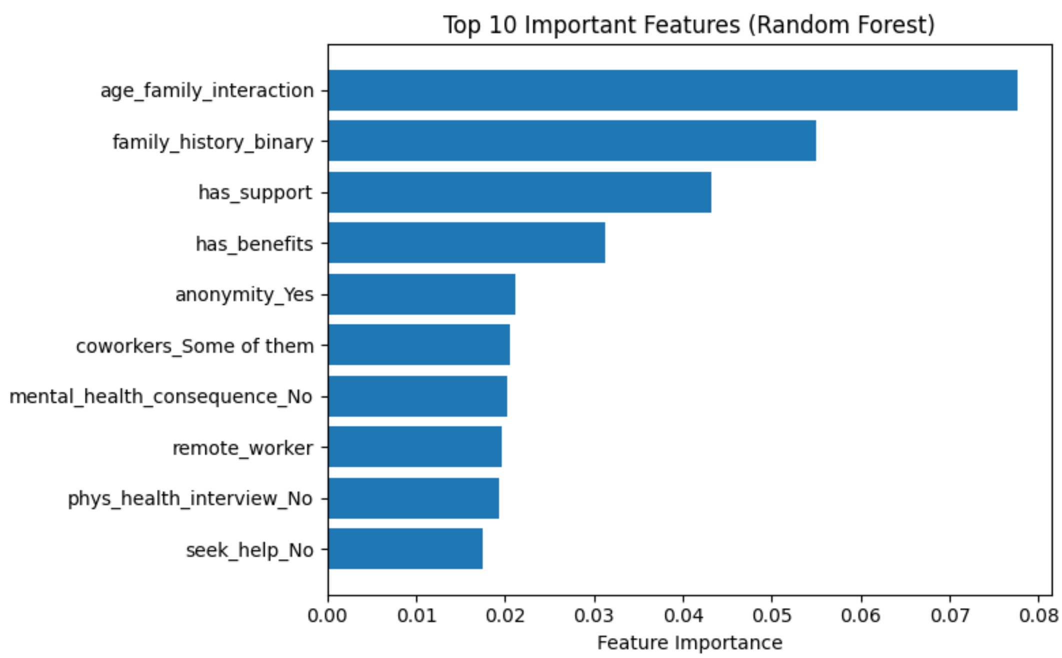
Model Performance Summary

	Model	Accuracy	F1-Score	ROC-AUC
2	Gradient Boosting	0.677291	0.684825	0.713351
3	SVM	0.669323	0.655602	0.710850
1	Random Forest	0.653386	0.656126	0.704149
0	Logistic Regression	0.673307	0.674603	0.699756

To predict whether someone is likely to seek mental health treatment, I tested four different machine learning models: Gradient Boosting, SVM (Support Vector Machine), Random Forest, and Logistic Regression. Among them, Gradient Boosting performed the best across all metrics

— with the highest accuracy (67.7%), F1-score (68.5%), and ROC-AUC (71.3%). This means it was the most reliable model in identifying patterns and correctly classifying treatment-seeking behavior. The other models, especially SVM and Random Forest, also performed reasonably well, but slightly behind. Logistic Regression was the simplest model and still gave competitive results. Overall, the Gradient Boosting model was chosen as the final model for its better balance of precision, recall, and ability to distinguish between those who sought treatment and those who didn't.

Family History and Workplace Support Play a Big Role in Treatment-Seeking Behavior



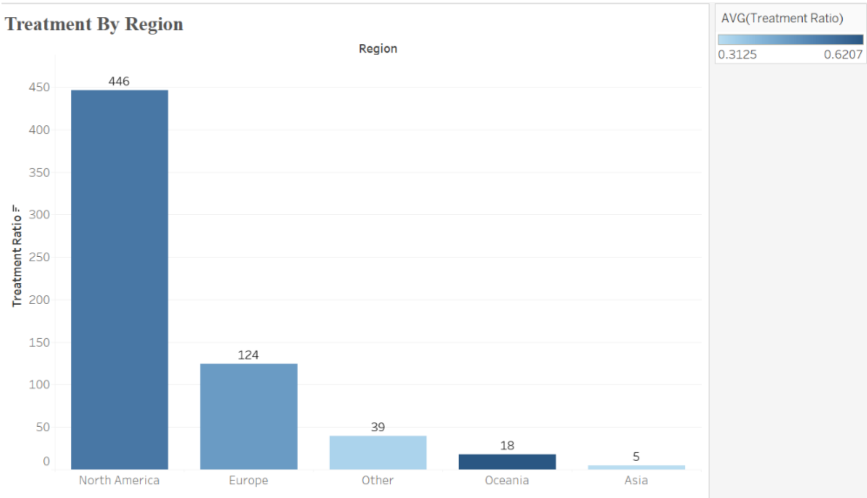
This chart shows the top 10 features that had the most impact on the model's predictions. The most influential factor was age_family_interaction — meaning age combined with family history played a strong role in whether someone sought mental health treatment. Unsurprisingly, having a family history of mental illness and access to support at work were also highly important. Features like mental health benefits, anonymity at work, and coworker attitudes also contributed to the decision-making process. These results show that a mix of personal background and workplace environment strongly influence whether someone chooses to get help. The takeaway? Employers can make a meaningful difference by creating supportive, stigma-free spaces — and personal history often shapes how people respond to those environments.

Access to Workplace Support Makes a Clear Impact on Mental Health Decisions

These SHAP plots show how having mental health support available at work influences the model's prediction — and the difference is striking. On the right side of the graph (positive SHAP values), you can see a dense cluster of individuals who had access to support and whose predicted likelihood of seeking treatment increased. On the left, people without support tend to have negative SHAP values, meaning their presence lowered the model's confidence in predicting treatment-seeking behavior.

In simpler words: when people know support is available, the model strongly believes they are more likely to seek help. This reinforces the idea that workplace resources aren't just a checkbox they directly affect how people act. SHAP helps us visually confirm that access to support systems is one of the most meaningful and positive predictors in this entire project.

North America Leads in Mental Health Treatment-Seeking — Other Regions Lag Behind



This chart shows how many people in each region reported seeking mental health treatment. North America stands out with the highest number of treatment-seekers by far, followed by Europe. Other regions like Oceania, Asia, and the "Other" category have much lower counts. While this could partly reflect sample size differences (more respondents from North America), it also suggests that people in North America may have better access to resources, less stigma, or more openness around mental health. The color shading — based on treatment ratios — also shows how varied the treatment behavior is across the globe. It's a reminder that mental health support needs to be both global and culturally tailored, because the willingness and ability to seek help still differ greatly depending on where someone lives.

Work Policy can likely effect and Employees will seek help of treatment

This chart shows a clear connection between workplace mental health benefits and the likelihood of seeking treatment. People who said "Yes" to having mental health benefits had the highest average treatment rate, while those who weren't sure had the lowest. This tells us that just having benefits isn't enough employees also need to know about them. Meanwhile, people who reported having no benefits still had a moderate treatment rate, which might reflect personal drive or outside resources. Overall, this highlights how clear communication and active promotion of mental health benefits can encourage more employees to get the help they need.

Challenges and Solutions

Working with real-world survey data came with its own set of hurdles. First, the data had many inconsistent entries (especially in the gender field) and missing values scattered across key columns. To tackle this, I standardized responses into clear categories (e.g., "Male," "Female," "Unidentified") and adopted a tiered imputation strategy—filling small gaps with common values while carefully dropping or flagging columns with excessive missingness. Another challenge was balancing model performance with interpretability: complex models like Gradient Boosting gave the best scores but can feel like a "black box." I addressed this by engineering intuitive features (like has_support and age_family_interaction) and using SHAP values with a model-agnostic explainer, ensuring each prediction could be traced back to real workplace or personal factors. Finally, visualizing these insights in Tableau required exporting and reshaping data so non-technical stakeholders could explore treatment trends interactively.

Conclusion

This project demonstrates how thoughtful data work and machine learning can spotlight the human side of mental health in the workplace. By cleaning messy survey responses, engineering meaningful features, and comparing several models, we created a reliable predictor of who is likely to seek treatment. More importantly, explainability tools like SHAP and clear Tableau dashboards translate these technical findings into actionable insights: for example, making sure employees know about available benefits, reducing stigma in interviews, and tailoring support by region. Ultimately, data isn't just numbers—it's a way to drive real change. I hope these results help HR teams and mental health advocates build more supportive, transparent workplaces where everyone feels safe to ask for help.