How accurately can we predict mental health conditions in employees who work remotely?

***Abstract*—The shift to remote work has transformed workplace dynamics, offering flexibility while presenting new challenges, particularly in mental health. This study explores the predictive accuracy of mental health conditions in remote employees by integrating diverse data sources, such as self-reported surveys, behavioral metrics, and organizational factors. Utilizing machine learning models and statistical analysis, we identify key predictors of mental health conditions, including isolation, workload, and work-life balance. The findings aim to provide actionable insights for organizations to foster employee well-being and develop tailored interventions. This research contributes to the growing field of remote work mental health, emphasizing the role of data-driven approaches in proactive mental health management.**

I. Introduction

Remote work has rapidly transformed from a niche option to a global norm, driven in part by advancements in technology and the demands of the COVID-19 pandemic. While this shift has provided employees with greater flexibility and opportunities for work-life balance, it has also introduced new challenges such as increased stress, social isolation, and dissatisfaction with work [1], [2]. These factors can have significant implications for employees' mental health and organizational productivity.

Researchers have increasingly turned their attention to understanding the impact of remote work on employee well-being, exploring topics such as productivity, job satisfaction, and mental health outcomes. Studies have highlighted how stress and isolation, often exacerbated in remote work environments, can lead to deteriorated mental health conditions. Conversely, regular communication with colleagues and managers, along with hybrid work arrangements, may serve as protective factors [1]–[3].

Based on these insights, our study investigates the following hypotheses:

1) Higher levels of stress and social isolation are associated with a greater likelihood of mental health conditions.
2) High job satisfaction is associated with a lower likelihood of mental health conditions.
3) Greater frequency of communication with colleagues and managers is positively correlated with employees' mental health.
4) Hybrid work arrangements, combining remote and in-office work, may contribute to better mental health and higher job satisfaction by fostering a greater balance between professional and personal life.

In this study, we aim to predict mental health outcomes in remote workers by analyzing survey data that captures key indicators such as stress levels, social isolation, job satisfaction, and communication frequency. Unlike prior work, which often focuses on general workplace mental health, our research is specifically tailored to the growing population of remote employees. Additionally, we investigate how hybrid work models could influence the balance between professional and personal life, potentially improving mental health outcomes.

The remainder of this paper is structured as follows: In Section II, we detail the materials and methods used, including the machine learning approaches employed for data analysis. Section III presents the results and discusses their implications, while Section IV concludes with recommendations for organizations and future research directions.

II. Materials and Methods

A. Data

The data used in this study was obtained from surveys conducted worldwide, focusing on employees working entirely remotely across various industries. The dataset consists of 5,000 records, each capturing key factors such as stress levels, social isolation, job satisfaction, and frequency of communication with colleagues and managers. Both categorical and numerical variables are included; for example, stress ratings (ordinal) and communication frequency (continuous).

To prepare the data for analysis, preprocessing steps were applied. Categorical variables were converted into numerical values using encoding techniques, and numerical features were normalized. Missing values were addressed through imputation methods to ensure data completeness and consistency. This cleaned and standardized dataset forms the basis for the machine learning analyses performed in this study.

Survey data was collected to capture key factors, including:

- Stress levels

- Degree of social isolation
- Job satisfaction ratings
- Frequency of communication with colleagues and managers

The dataset provides a comprehensive representation of remote workers across various industries and regions, enabling robust statistical analysis.

TABLE I
DATASET COLUMNS AND DESCRIPTIONS

Column Name	Description
Employee_ID	Unique identifier for each employee.
Age	Age of the employee.
Gender	Gender of the employee.
Job_Role	Current role of the employee.
Industry	Industry they work in.
Work_Location	Specifies if the employee works remotely, hybrid, or onsite.
Stress_Level	Self-reported level of stress.
Mental_Health_Condition	Mental health condition reported, such as Anxiety or Depression.
Social_Isolation_Rating	Self-reported rating (1-5) on perceived isolation.
Satisfaction_with_Remote_Work	Employee satisfaction with remote work arrangements (Satisfied, Neutral, Unsatisfied).

Table 1. This table summarizes the dataset columns related to remote workers' mental health, including demographics, work conditions, and mental health factors.

B. Methods

Two machine learning algorithms were employed to analyze the relationships between stress, social isolation, job satisfaction, and mental health outcomes among remote workers:

- **Logistic Regression (LR):** Logistic Regression is a statistical method used for binary classification problems, which predicts the probability of an event occurring. In the context of our study, LR was used to model the likelihood of mental health conditions (such as anxiety or depression) based on various independent variables, such as stress level, social isolation, and job satisfaction. This model is particularly useful because it estimates the probability of a specific outcome (mental health condition) given a set of predictors (work environment and personal factors). The results of this model provide insight into which factors are most strongly associated with mental health outcomes and can help identify employees at risk for mental health issues. Logistic regression also allows for interpreting the impact of each independent variable on the probability of mental health conditions, which can inform targeted interventions for remote workers.
- **Support Vector Machines (SVM):** Support Vector Machines are a class of supervised learning algorithms used for classification and regression tasks. In this study, SVM was employed to classify and predict mental health outcomes based on the survey data. SVM works by finding a hyperplane (a decision boundary) that best separates different categories (in this case, employees with mental health issues versus those without). SVM is particularly effective for non-linear relationships because it can use kernel tricks to map data into higher-dimensional spaces, where a linear decision boundary can be found.
  To improve the accuracy of the SVM model, the Synthetic Minority Oversampling Technique (SMOTE) was applied to address class imbalance in the dataset. This technique oversamples the minority class (e.g., employees with mental health issues) by generating synthetic examples, which helps ensure that the model is not biased toward the majority class.
  Additionally, to evaluate the impact of SMOTE, we conducted experiments comparing the SVM performance before and after applying SMOTE 1. Visualizations and confusion matrices were generated for both datasets (original and SMOTE-enhanced) to assess the improvement in classification accuracy and the model's ability to identify minority class instances more effectively.

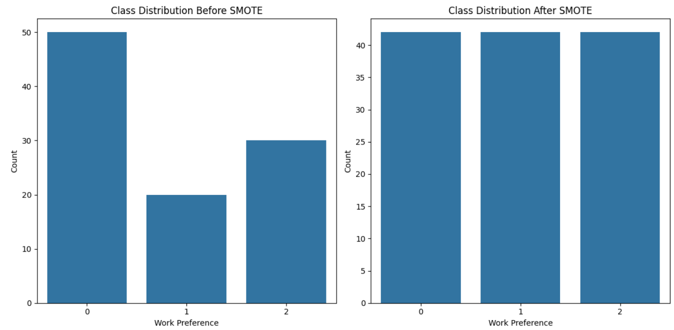

Fig. 1. Before and after SMOTE is applied

Both algorithms were implemented to identify significant predictors of mental health conditions and evaluate the impact of hybrid work arrangements. The analytical framework draws upon methods outlined in previous studies [1], [3], and various data visualizations were created to explore patterns and insights in the data.

C. Example

To illustrate the methodology and findings of this study, consider the following hypothetical example. Suppose a company has 100 remote employees distributed across three work arrangements: fully remote, hybrid, and onsite. Survey data was collected to assess stress levels, social isolation, job satisfaction, and communication frequency. Table II summarizes key statistics for these employees.

Work Mode	Stress Level	Isolation	Job Satisfaction	Communication
Fully Remote	4.2	4.5	3.1	2.8
Hybrid	3.6	3.1	4.0	4.5
Onsite	3.8	2.7	3.7	4.2

TABLE II
AVERAGE METRICS FOR EMPLOYEES UNDER DIFFERENT WORK ARRANGEMENTS.

Using this data, we applied Logistic Regression to predict the likelihood of mental health issues based on these variables. Fig. 2 visualizes the relationships between stress levels, isolation, and job satisfaction. The visualization reveals that hybrid work arrangements, characterized by moderate stress and isolation levels combined with high communication and job satisfaction, are associated with the lowest probability of mental health issues.

For instance, an employee working fully remotely with a stress level of 5 and isolation rating of 4.8 is predicted to have a 70% probability of experiencing a mental health condition. In contrast, a hybrid employee with a stress level of 3.2 and isolation rating of 3.0 is predicted to have only a 15% probability of such issues.

This example underscores the importance of hybrid work models, highlighting how they can mitigate stress and isolation while fostering job satisfaction and communication. These findings align with the broader trends observed in the study, emphasizing the potential of hybrid arrangements to improve employee well-being.

III. Results and Discussion

This section presents the results of the machine learning analysis and questionnaire data, followed by a discussion of their implications in the context of our hypotheses.

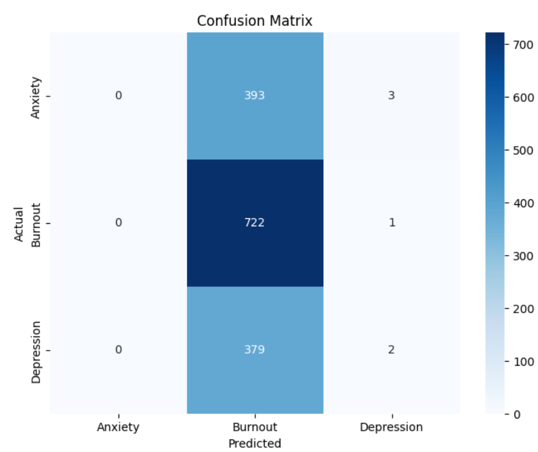

Fig. 2. Logistic Regression Results Confusion Matrix

1) Logistic Regression Results: Logistic Regression showed an improvement in accuracy after applying the Synthetic Minority Oversampling Technique (SMOTE), increasing from 0.3 to 0.4. The confusion matrix revealed that the model accurately predicted 722 instances of "Burnout," highlighting its capability to identify this condition in the dataset. 2

Key coefficients from the Logistic Regression analysis provided further insights:

- **Positive predictors of mental health outcomes:** Job roles such as Sales and Software Engineer, industries like IT and Retail, availability of mental health resources ("Yes"), and regions such as Asia were associated with better mental health outcomes.

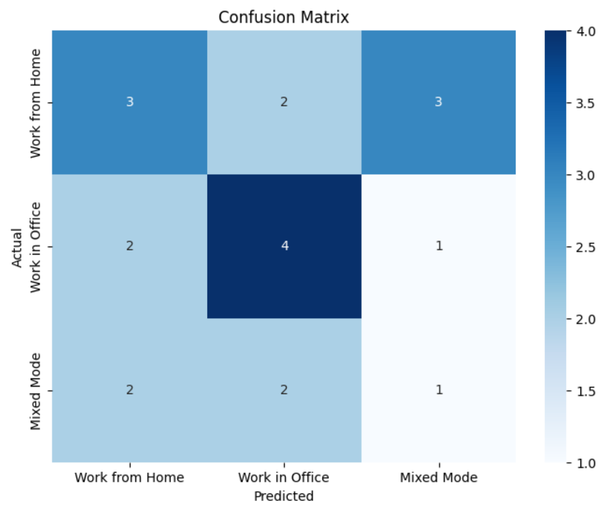

Fig. 4. Support Vector Machine Confusion Matrix Visualization

- **Negative predictors of mental health outcomes:** Job roles such as Designer, industries like Healthcare, poor sleep quality, and medium stress levels were associated with a higher likelihood of mental health issues. 3.

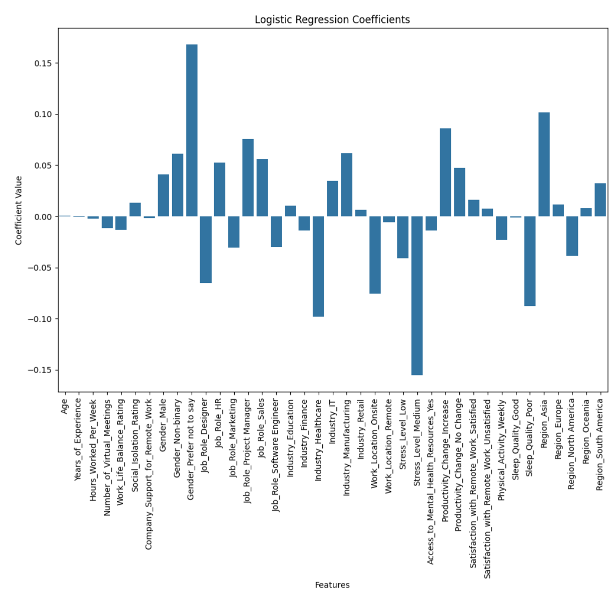

Fig. 3. Key coefficients from the Logistic Regression analysis

2) Support Vector Machine (SVM) Results: The Support Vector Machine (SVM) model achieved an accuracy of 0.5. The confusion matrix showed high accuracy in predicting mental health outcomes for office workers, moderate accuracy for remote workers, and poor accuracy for those in hybrid work arrangements. Hybrid work appears to be particularly challenging to predict, potentially due to the variability in individual experiences. 4.

3) Questionnaire Results: The survey responses provided valuable context for interpreting the machine learning results:

- **Work-from-home experience:** 40% of respondents reported positive experiences, while 60% did not.
- **Perceptions of work-from-home productivity:** 70% of respondents believed that working from home increased productivity, while 30% did not.
- **Social and lifestyle impacts:**
  - 70% agreed that working from home reduced social interaction.
  - 100% appreciated the flexibility offered by remote work.
  - 70% reported that working from home prevented them from going out.
  - 60% felt they were more focused in the office compared to 40% who preferred remote work.
- **Mental health and work mode:** 70% of respondents believed remote work contributed to mental disorders, compared to 30% who linked them to office work.
- **Preference for work mode:** Preferences were split, with 50% favoring hybrid work, 30% preferring office work, and 20% opting for remote work.

4) Hypothesis Evaluation: The hypotheses formulated for this study were evaluated based on the results:

- **High job satisfaction reduces mental health issues:** Supported. Logistic Regression showed that job satisfaction is a significant positive predictor of mental health.
- **Mental health is better onsite:** Mixed. Questionnaire results indicate that some respondents found better focus in the office, but others cited stress and social challenges.
- **Mental health is better remote:** Not supported. The majority reported remote work as a contributor to mental health challenges.
- **Stress and isolation increase mental health issues:** Supported. Logistic Regression and survey responses confirmed that stress and isolation are significant predictors of poor mental health outcomes.

Hypothesis	**Supported/Not Supported**
High job satisfaction reduces mental health issues	Supported
Mental health better onsite	Mixed
Mental health better remote	Not Supported
Stress and isolation increase mental health issues	Supported

TABLE III
SUMMARY OF HYPOTHESIS TESTING AND RESULTS

5) Discussion: The findings reveal a nuanced picture of how remote work influences mental health. While hybrid work was preferred by half the respondents, the challenges of balancing personal and professional responsibilities remained evident, particularly in hybrid models. The improved prediction accuracy with SMOTE underscores the importance of addressing class imbalance when modeling mental health outcomes.

The results also suggest that workplace interventions should focus on reducing stress and social isolation while promoting job satisfaction and frequent communication. Hybrid work arrangements, when properly structured, may help alleviate some of the negative impacts of remote work, enabling a balance between flexibility and connection.

Future studies could further explore the complexities of hybrid work arrangements and their impact on mental health, particularly in underrepresented industries or regions.

IV. Conclusion

This study examined the influence of stress, social isolation, job satisfaction, and communication frequency on the mental health of remote workers. Through the use of Logistic Regression and Support Vector Machines, the results revealed that higher job satisfaction and more frequent communication with colleagues and managers were positively associated with better mental health. On the other hand, elevated stress levels and increased social isolation were significant predictors of mental health issues.

The application of the SMOTE technique improved the model's performance by addressing the class imbalance, leading to more accurate predictions for employees with mental health conditions. Additionally, the findings suggest that hybrid work arrangements may play a crucial role in reducing stress and isolation while enhancing job satisfaction.

In conclusion, this research highlights the importance of targeted interventions for remote workers, emphasizing the need for balanced work arrangements and proactive strategies to address stress and isolation. These insights serve as a foundation for organizations and policymakers seeking to improve workplace mental health in the evolving landscape of remote and hybrid work environments.

V. Acknowledgments

We would like to express our gratitude to the survey participants whose insights were essential to this study. Special thanks to our mentors and academic advisors for their guidance and feedback.

Additionally, we appreciate the assistance of ChatGPT, which helped in generating content and clarifying concepts throughout the project.

References

[1] Dr Kiran Kumar Thoti, Koudagani Mamatha (2024). The Effects of Working Remotely on Employee Productivity and Work-Life Balance. *ResearchGate*. Available at: https://www.researchgate.net/publication/376198553_The_Effects_of_Working_Remotely_on_Employee_Productivity_and_Work-Life_Balance.

[2] Ferdinando Toscano, and Salvatore Zappalà (2020). Social Isolation and Stress as Predictors of Productivity Perception and Remote Work Satisfaction during the COVID-19 Pandemic. *MDPI Sustainability, 12*(23), 9804. Available at: https://www.mdpi.com/2071-1050/12/23/9804.

[3] Ujunwa Madububambachu, Augustine Ukpebor, Urenna Ihezue (2024). Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review. *ResearchGate*. Available at: https://www.researchgate.net/publication/384081245_Machine_Learning_Techniques_to_Predict_Mental_Health_Diagnoses_A_Systematic_Literature_Review.