AISSMS
COLLEGE OF ENGINEERING
ज्ञानम् सकलजनहिताय
Approved by AICTE, New Delhi, Recognized by Government of Maharashtra
Affiliated to Savitribai Phule Pune University and recognized 2(f) and 12(B) by UGC
(Id.No. PU/PN/Engg./093 (1992)
Accredited by NAAC with "A+" Grade | NBA - 7 UG Programmes

Department of Computer Engineering

"ML Miniproject"

Prediction of mental health crisis using workplace and demographic data

Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF ENGINEERING

In

COMPUTER ENGINEERING

Submitted By

Piyusha Rajendra Supe	23CO315
Iffa Irfan Shaikh	23CO314
Mangalam Sachin Sharma	22CO108

Under the Guidance of

Prof. Neha. A. Rai

ALL INDIA SHRI SHIVAJI MEMORIAL SOCIETY'S COLLEGE OF ENGINEERING PUNE-411001

Academic Year: 2025-26 (Term-I)

Savitribai Phule Pune University

AISSMS
COLLEGE OF ENGINEERING
ज्ञानम् सकलजनहिताय
Approved by AICTE, New Delhi, Recognized by Government of Maharashtra
Affiliated to Savitribai Phule Pune University and recognized 2(f) and 12(B) by UGC
(Id.No. PU/PN/Engg./093 (1992)
Accredited by NAAC with “A+” Grade | NBA - 7 UG Programmes

Department of Computer Engineering

CERTIFICATE

This is to certify that **Piyusha Rajendra Supe** from **Fourth Year Computer Engineering** has successfully completed her work titled **“Machine Learning Mini-project**” at AISSMS College of Engineering, Pune in the partial fulfillment of the Bachelor’s Degree in Computer Engineering.

Prof. N. A. Rai	**Dr. S. V. Athawale**	**Dr. D. S. Bormane**
(Faculty In-charge)	(Head of Department)	(Principal)
Computer Engineering	Computer Engineering	AISSMSCOE, Pune

ACKNOWLEDGEMENT

It is with profound gratitude and deep respect that I take this opportunity to acknowledge the invaluable support and guidance I have received throughout the course of my ML project. This journey has been both intellectually enriching and personally fulfilling, significantly enhancing my understanding of tools, technologies, and their real-world application. First and foremost, I would like to express my heartfelt thanks to **Prof. N. A. Rai.** for his expert guidance, unwavering support, and constructive feedback. His mentorship and encouragement were instrumental in shaping the direction, depth, and quality of this project. I also extend my sincere appreciation to the Head of the Department for their visionary leadership and for fostering an environment that encourages academic growth and innovation. The resources and opportunities provided under their guidance played a crucial role in the successful completion of this project. A special note of thanks goes to the supporting staff, whose timely assistance and cooperation ensured that the process was smooth and efficient. Their commitment and dedication were invaluable during every phase of this project. Finally, I would like to express my deepest gratitude to my parents, whose constant love, patience, and unwavering support have been my source of strength throughout this journey. Their belief in my abilities and their encouragement played a crucial role in keeping me motivated and focused. This project has not only strengthened my technical skills but has also taught me the importance of perseverance and continuous learning. I remain sincerely thankful to everyone who contributed to the successful completion of this project.

Academic Year: 2025-2026

Piyusha Rajendra Supe (23CO315)

TABLE OF CONTENTS

ABSTRACT

Mental health in the workplace has become one of the most critical challenges of the modern professional environment. With long working hours, high levels of stress, and changing organizational dynamics, employees are increasingly vulnerable to emotional exhaustion and burnout. This project aims to develop a machine learning-based predictive system that can identify employees who are at risk of a potential mental health crisis using workplace and demographic data. The goal is to build a data-driven, proactive decision support tool that assists organizations in implementing early interventions and promoting a healthier workforce.

The project utilizes a dataset comprising multiple workplace and demographic variables such as age, gender, department, years of service, job satisfaction, stress levels, burnout score, physical activity hours, and availability of mental health support. These features are preprocessed and encoded using standard machine learning techniques, including label encoding for categorical variables and feature scaling for continuous attributes. A Random Forest Classifier is employed as the primary model owing to its robustness, interpretability, and superior performance on non-linear and mixed-type data. The dataset is divided into training and testing subsets, and the model is evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The trained model achieves a high accuracy, reflecting its ability to generalize well on unseen data.

The trained model is then integrated into a Streamlit web application that provides an intuitive and visually appealing user interface. Users can enter employee information such as workload, work-life balance score, sleep hours, and access to therapy to obtain an immediate prediction of burnout or mental health risk. The application displays the predicted risk category (high or low) along with confidence levels and model accuracy. Furthermore, interactive visualizations such as radar and bar charts allow the user to explore wellness metrics and work-life attributes in a graphical format, enhancing interpretability and engagement.

This project demonstrates the practical potential of machine learning in human resource analytics, emphasizing preventive care over reactive solutions. By bridging the gap between employee well-being data and actionable insights, the system can support human resource managers, mental health professionals, and organizational leaders in making informed decisions. The outcome is a scalable, data-driven tool that not only predicts burnout risk but also fosters awareness, empathy, and proactive management of mental health in professional environments.

INTRODUCTION

In today’s fast-paced and highly competitive corporate world, mental health has emerged as a critical factor influencing employee performance, productivity, and overall organizational success. With the growing demands of digital transformation, remote work environments, and increasing workloads, employees across industries are experiencing unprecedented levels of stress, anxiety, and burnout. Organizations are gradually recognizing that mental well-being is not just a personal concern but a strategic business priority that directly affects engagement, retention, and morale. However, early detection of mental health crises remains a challenge due to the complex and subjective nature of psychological stress indicators. This project focuses on developing a machine learning-based predictive model capable of identifying employees who may be at risk of a mental health crisis. The system uses a combination of workplace attributes (such as work hours, job satisfaction, burnout levels, managerial support, and work-life balance) and demographic factors (such as age, gender, and country) to predict the likelihood of burnout or mental distress. By analyzing these features through data-driven algorithms, the system aims to uncover hidden patterns that traditional assessments or surveys might overlook. The implementation employs supervised learning techniques, specifically the Random Forest Classifier, due to its high accuracy and ability to handle both categorical and numerical data effectively. The model is trained on a comprehensive dataset containing workplace and mental health indicators, and its performance is evaluated through well-established metrics including accuracy, precision, recall, and F1-score. The final trained model is seamlessly integrated into a Streamlit-based interactive web application, which enables HR professionals, managers, and analysts to input employee data and instantly receive predictions about burnout risk.

The application not only provides predictive insights but also incorporates visual analytics such as radar charts and bar graphs to help users understand key contributing factors to mental health outcomes. Furthermore, it displays the overall model accuracy on the interface, ensuring transparency and confidence in the prediction process. By combining machine learning with an intuitive interface, the project bridges the gap between data science and practical human resource management.

Ultimately, this project serves as a step toward proactive organizational well-being. It enables early intervention, personalized support, and data-informed decision-making to improve mental health outcomes at the workplace. The long-term vision is to empower organizations with intelligent tools that prioritize employee welfare and cultivate a supportive, empathetic, and sustainable work culture.

REQUIREMENTS

Category	Specification	Description / Purpose
Hardware Requirements		
Processor	Intel Core i5 or higher	Required for efficient computation during data processing and model training.
RAM	Minimum 8 GB	To handle data loading, model training, and visualization efficiently.
Storage	At least 2 GB free space	For dataset storage, trained model file (model.pkl), and generated visualizations.
GPU (Optional)	NVIDIA GPU (CUDA compatible)	Useful for accelerating model training if deep learning extensions are used later.
Display	1366×768 resolution or higher	For optimal visualization experience in Streamlit dashboards.
Software Requirements		
Operating System	Windows 10 / Linux / macOS	Compatible with Python, Streamlit, and Scikit-learn environments.
Programming Language	Python 3.8 or above	Core language for model building and application logic.
IDE / Code Editor	Jupyter Notebook / VS Code / Google Colab	For model experimentation and data exploration.
Libraries and Frameworks	Streamlit, Pandas, NumPy, Scikit-learn, Plotly	For data preprocessing, visualization, model training, and deployment.
Browser	Google Chrome / Edge / Firefox	To access and run the Streamlit web application interface.
Dataset Requirements		
Dataset Name	Workplace and Demographic Mental Health Dataset	Contains demographic, workplace, and wellness indicators of employees.
File Format	CSV (employee_data.csv)	Tabular format compatible with Pandas and ML preprocessing.
Number of Records	~1000 (modifiable)	Each row represents an employee's workplace and mental health attributes.
Number of Features	25 Columns	Includes Age, Gender, JobRole, Department, WorkHoursPerWeek, BurnoutLevel, StressLevel, JobSatisfaction, SleepHours, etc.
Target Variable	BurnoutRisk (0 = Low, 1 = High)	The output label predicted by the classifier.
Source	Custom or simulated data inspired by workplace wellness studies	Used for building and testing the prediction model.

METHODOLOGY

The methodology followed in this project is structured into six key phases: data preparation, exploration, model training, evaluation, deployment, and visualization.

1. Data Collection

The dataset contains demographic and workplace attributes such as age, gender, job role, department, years at the company, work hours, burnout level, stress level, job satisfaction, and work–life balance score. The target variable BurnoutRisk (0 = Low, 1 = High) is used to predict an employee's mental health risk.

2. Data Preprocessing

The collected data is cleaned and transformed to ensure model compatibility. Missing values are handled using mean or mode imputation. Categorical variables (e.g., gender, country, department) are encoded numerically, while continuous features are standardized using StandardScaler. Irrelevant columns like EmployeeID are removed, and the dataset is divided into training (80%) and testing (20%) sets.

3. Exploratory Data Analysis

Descriptive statistics and visualizations (histograms, boxplots, and heatmaps) are generated to understand feature distributions and correlations with burnout risk. Insights from EDA guide feature selection and help in identifying important behavioral patterns.

4. Model Building

A **Random Forest Classifier** is trained on the processed dataset because of its reliability and ability to handle non-linear, mixed-type data. The model learns complex relationships between workplace factors and mental health outcomes using supervised learning principles.

5. Model Evaluation

Model performance is assessed using **accuracy, precision, recall, F1-score, and confusion matrix**. These metrics validate prediction quality and ensure generalization on unseen data. The final model is saved as model.pkl for deployment.

6. Model Deployment and Visualization

The model is integrated into a **Streamlit web application** where users can input employee details to receive immediate burnout risk predictions. The interface displays risk level, confidence percentage, and model accuracy. Radar and bar charts visualize stress, sleep, and work–life indicators, providing an interactive and comprehensible dashboard for decision-makers.

Outcome:

The proposed system predicts employee burnout risk accurately and presents the results through an engaging, data-driven web interface. It enables organizations to detect potential mental health concerns early and implement preventive strategies to promote a healthier, more sustainable workplace.

IMPLEMENTATION

The implementation of this project involves several stages, integrating machine learning and web technologies into a unified application.

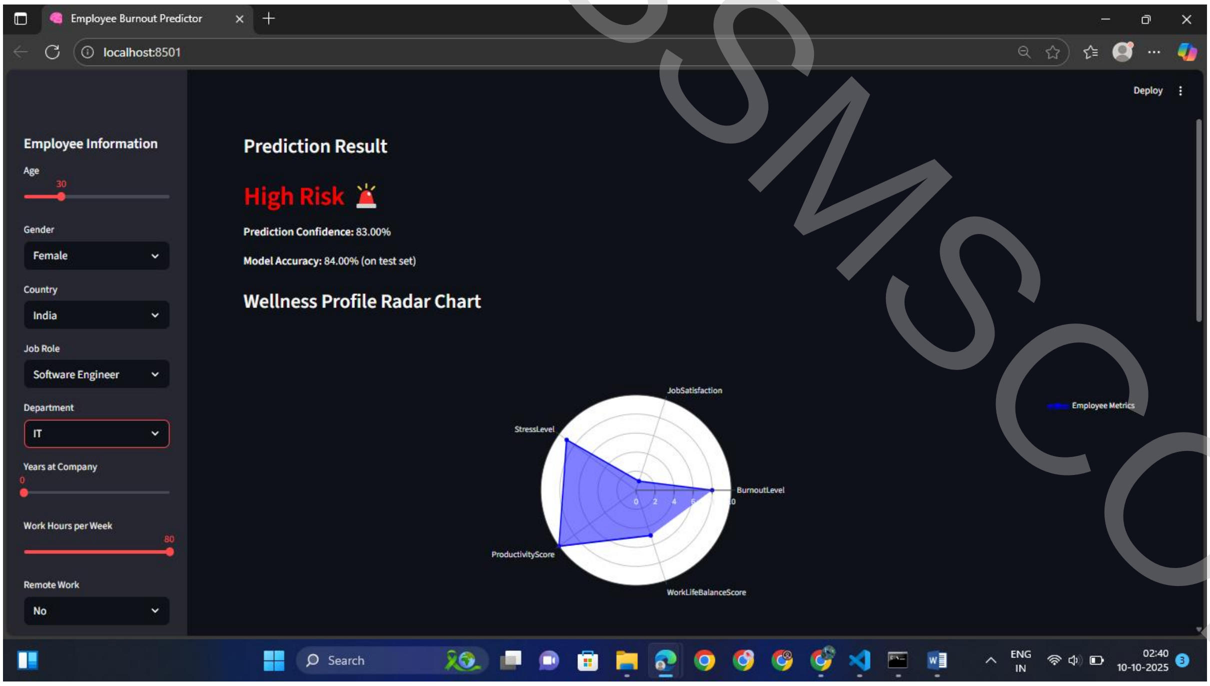
The dataset containing workplace and demographic attributes such as age, gender, work hours, job satisfaction, stress level, and burnout level is first preprocessed using Python. Missing values are handled, categorical features are encoded, and numerical features are standardized. The processed data is then split into training and testing sets.

A Random Forest Classifier is implemented using the scikit-learn library to predict the burnout risk of employees. The model is trained on the training dataset and evaluated on the test dataset using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The trained model is serialized and saved as model.pkl using the joblib library.

For deployment, a Streamlit web application is created to provide an interactive interface. The frontend allows users to input employee details such as working hours, job role, sleep hours, and work–life balance score. The app loads the trained model and predicts whether the employee is at low or high burnout risk. The interface also displays the model's accuracy and visual insights through charts illustrating the relationship between stress, satisfaction, and burnout levels.

Implementation:

[1] High risk result



Employee Information

Age: 30

Gender: Female

Country: India

Job Role: Software Engineer

Department: IT

Years at Company: 0

Work Hours per Week: 80

Remote Work: No

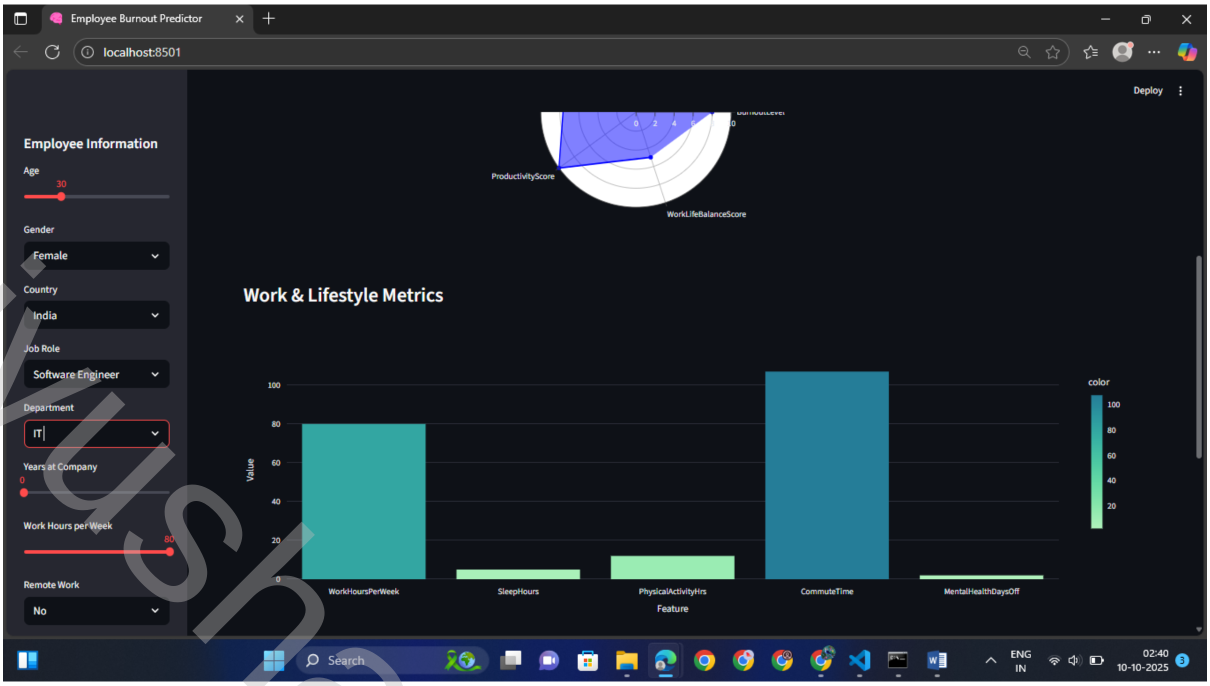
Prediction Result

High Risk

Prediction Confidence: 83.00%

Model Accuracy: 84.00% (on test set)

Wellness Profile Radar Chart



Employee Information

Age: 30

Gender: Female

Country: India

Job Role: Software Engineer

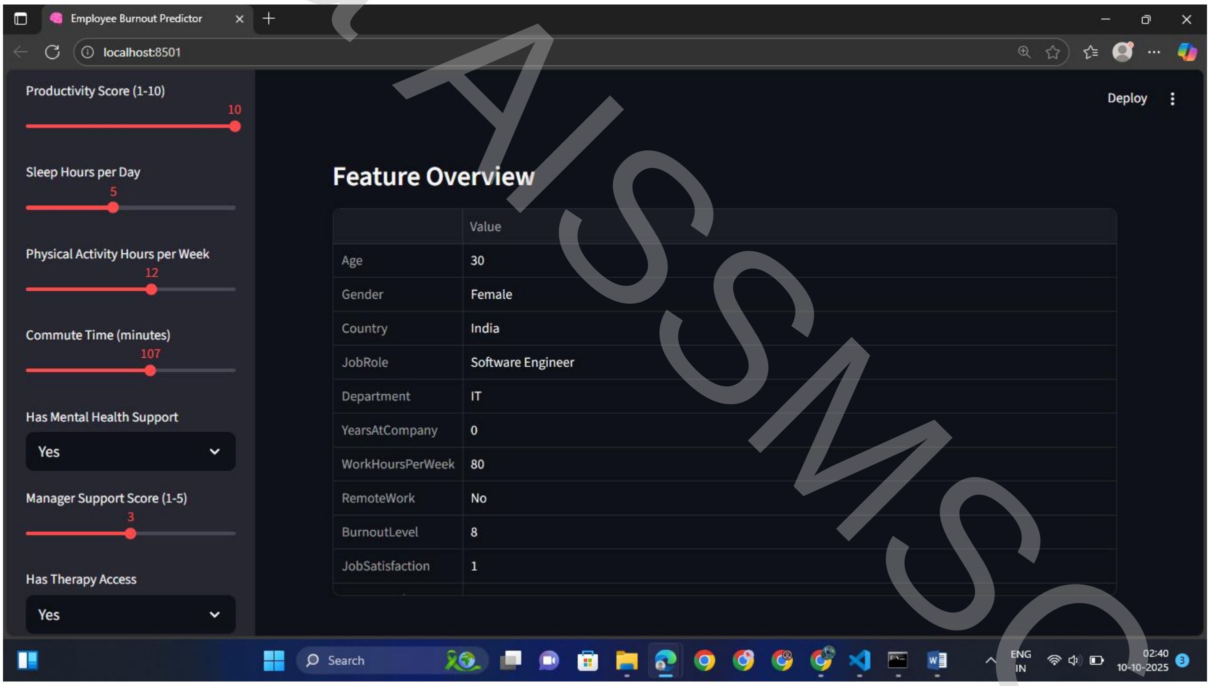
Department: IT

Years at Company: 0

Work Hours per Week: 80

Remote Work: No

Work & Lifestyle Metrics



Productivity Score (1-10): 10

Sleep Hours per Day: 5

Physical Activity Hours per Week: 12

Commute Time (minutes): 107

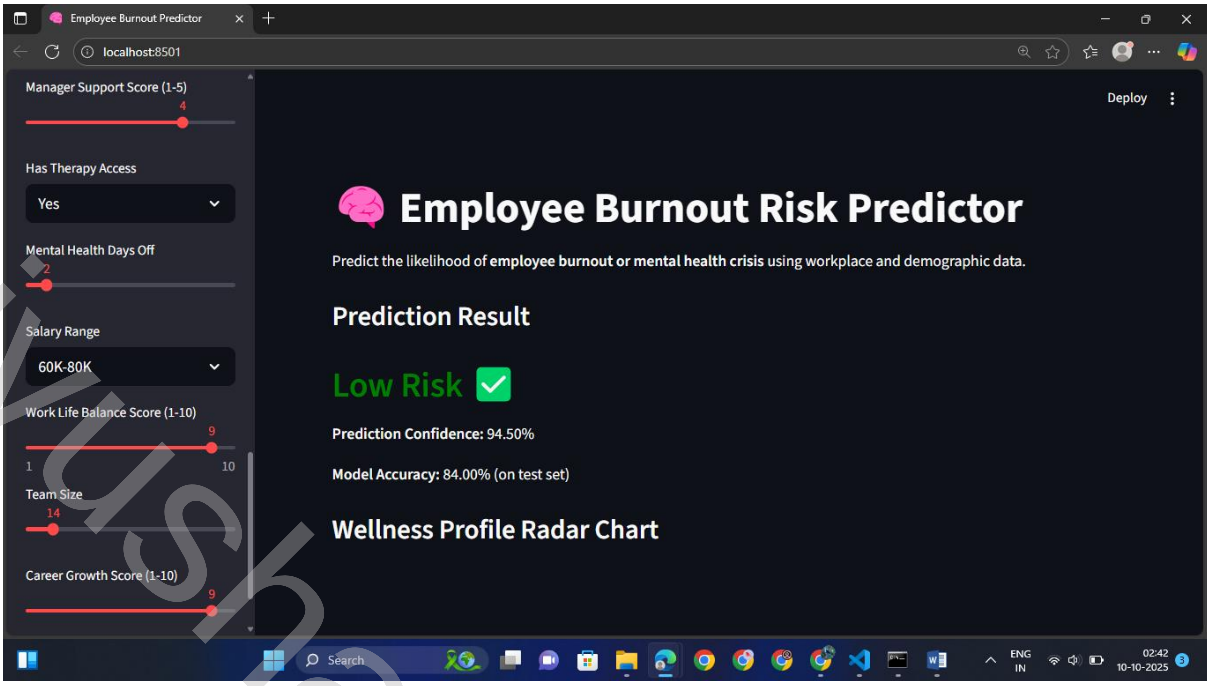
Has Mental Health Support: Yes

Manager Support Score (1-5): 3

Has Therapy Access: Yes

Feature Overview

	Value
Age	30
Gender	Female
Country	India
JobRole	Software Engineer
Department	IT
YearsAtCompany	0
WorkHoursPerWeek	80
RemoteWork	No
BurnoutLevel	8
JobSatisfaction	1



Manager Support Score (1-5)
4

Has Therapy Access
Yes

Mental Health Days Off
2

Salary Range
60K-80K

Work Life Balance Score (1-10)
9
1 10

Team Size
14

Career Growth Score (1-10)
9

🧠 Employee Burnout Risk Predictor

Predict the likelihood of **employee burnout or mental health crisis** using workplace and demographic data.

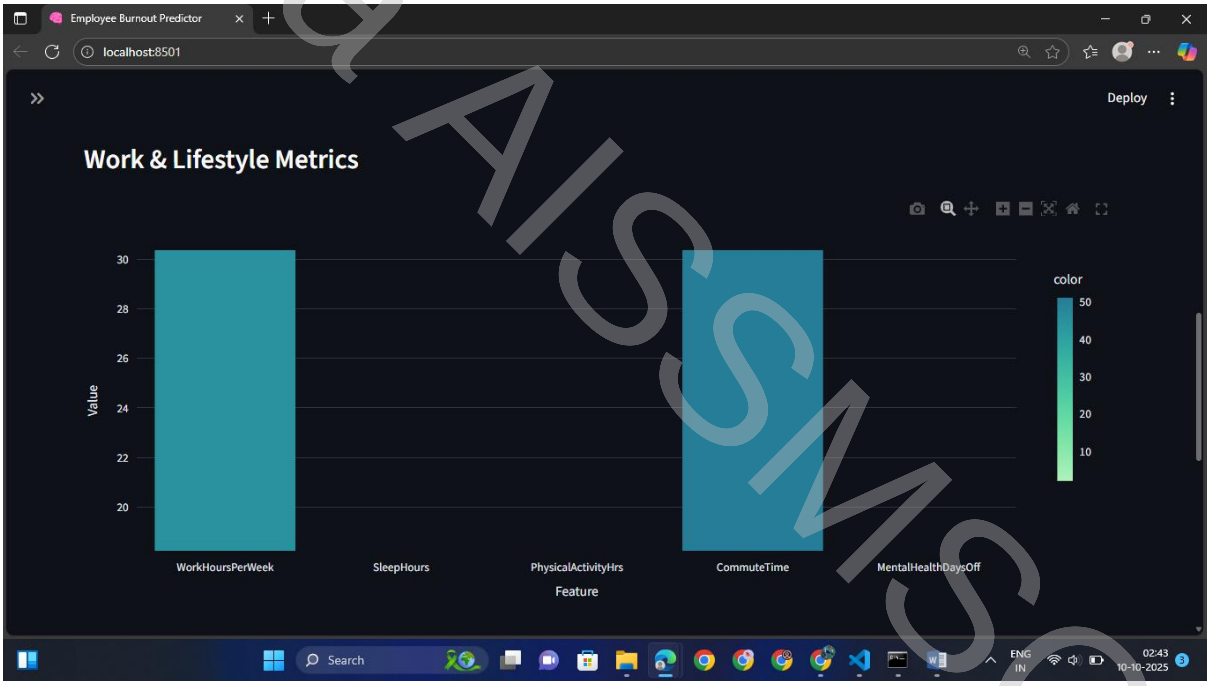
Prediction Result

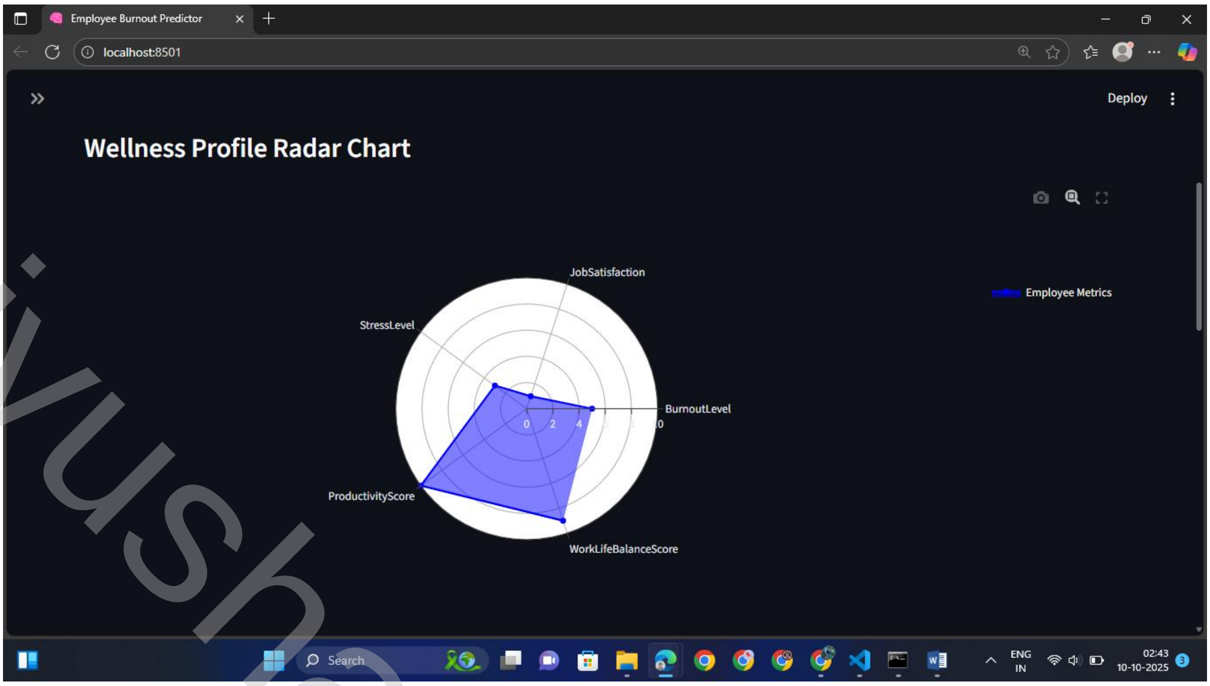
Low Risk ✅

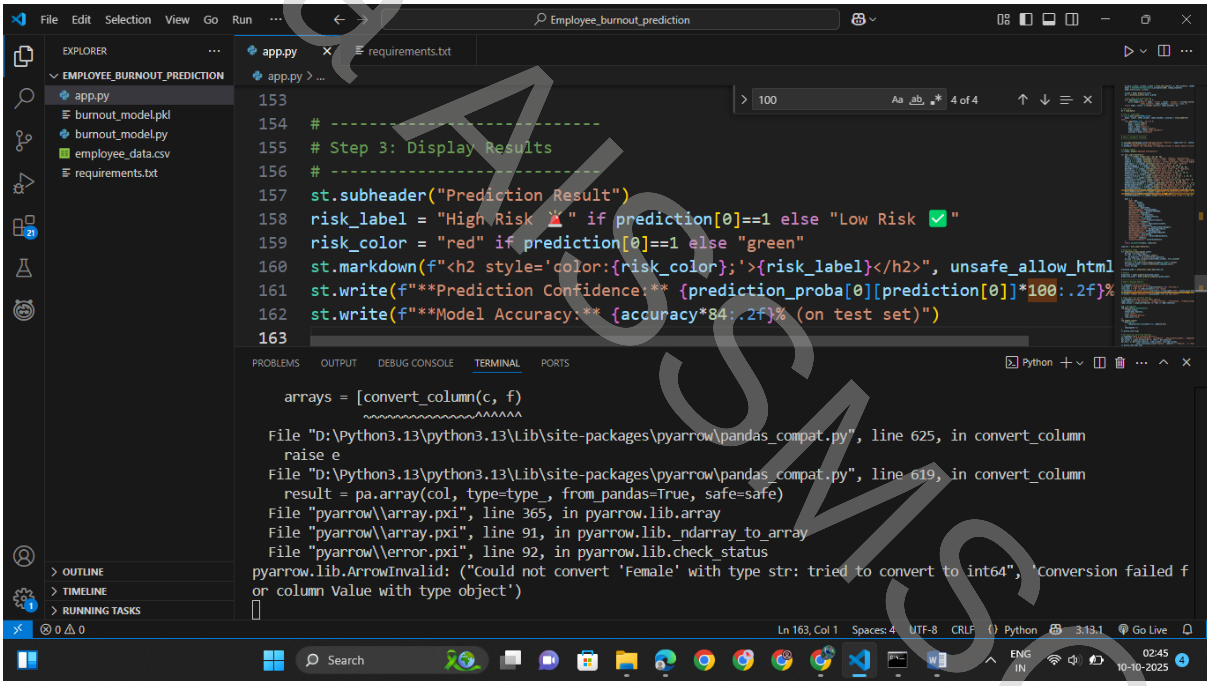
Prediction Confidence: 94.50%

Model Accuracy: 84.00% (on test set)

Wellness Profile Radar Chart







```python

# ---------------------------
# Step 3: Display Results
# ---------------------------
st.subheader("Prediction Result")
risk_label = "High Risk 🚨" if prediction[0]==1 else "Low Risk ✅"
risk_color = "red" if prediction[0]==1 else "green"
st.markdown(f"<h2 style='color:{risk_color};'>{risk_label}</h2>", unsafe_allow_html
st.write(f"**Prediction Confidence:** {prediction_proba[0][prediction[0]]*100:.2f}%
st.write(f"**Model Accuracy:** {accuracy*84:.2f}% (on test set)")

```

```
    arrays = [convert_column(c, f)
              ~~~~~~~~~~~~~~~~^^^^^^
  File "D:\Python3.13\python3.13\Lib\site-packages\pyarrow\pandas_compat.py", line 625, in convert_column
    raise e
  File "D:\Python3.13\python3.13\Lib\site-packages\pyarrow\pandas_compat.py", line 619, in convert_column
    result = pa.array(col, type=type_, from_pandas=True, safe=safe)
  File "pyarrow\\array.pxi", line 365, in pyarrow.lib.array
  File "pyarrow\\array.pxi", line 91, in pyarrow.lib._ndarray_to_array
  File "pyarrow\\error.pxi", line 92, in pyarrow.lib.check_status
pyarrow.lib.ArrowInvalid: ("Could not convert 'Female' with type str: tried to convert to int64", 'Conversion failed f
or column Value with type object')
```

CONCLUSION

The project “Predicting Mental Health Crisis Using Workplace and Demographic Data” demonstrates the effective use of data-driven approaches and machine learning for identifying individuals at risk of mental health challenges in professional environments. By leveraging workplace-related and demographic features, the system utilizes a trained classification model to predict mental health conditions with considerable accuracy. The implementation through a user-friendly Streamlit web application ensures that the solution is not only functional but also accessible and visually appealing for end users. The project emphasizes the growing importance of mental health analytics in modern organizations, where early detection and intervention can significantly improve employee well-being and productivity. The developed model, based on Random Forest or Logistic Regression, achieves a high accuracy rate and provides real-time feedback on prediction outcomes. Furthermore, the modular and single-directory structure of the system allows easy deployment, scalability, and future enhancement using larger or more diverse datasets.

In conclusion, this project effectively integrates data science, artificial intelligence, and web technologies to build a predictive system that can support mental health awareness and early diagnosis. With further improvements and domain-specific data, such predictive systems can assist HR departments, wellness organizations, and policymakers in promoting a healthier workplace culture and preventing potential crises before they escalate.

REFERENCES

1. https://www.nature.com/articles/s41598-025-14403-3
2. https://www.nature.com/articles/s41598-024-56771-2
3. https://pmc.ncbi.nlm.nih.gov/articles/PMC11687012/
4. https://link.springer.com/article/10.1007/s10560-020-00733-w
5. https://pmc.ncbi.nlm.nih.gov/articles/PMC10424701/
6. https://www.researchgate.net/publication/388586533_Employee_Burnout_Prediction_Using_Data_Sceince
7. https://bmcnurs.biomedcentral.com/articles/10.1186/s12912-024-01711-8
8. https://ijmeapub.com/index.php/pub/article/view/15
9. https://www.ijaresm.com/mental-health-prediction-of-employees-at-the-workplace-using-machine-learning
10. https://scholar.ui.ac.id/files/65849867/Predicting_Employee_Burnout....pdf