Leveraging machine Learning to identify Workplace Mental health Risks

A Predictive Modeling Approach

Introduction

- **Problem - Mental health challenges in Tech are widespread but often go undetected.**

- **Solution – A predictive model using workplace and demographic data to identify at-risk individuals.**

Business Objectives

- Early Identification of At-risk Employees
- Identify Key risk Factors
- Enable Proactive Intervention

Stakeholders & Impact

- Who Benefits?: HR teams, employees, managers
- Business Value: Reduced turnover, better morale, personalized wellness initiatives.

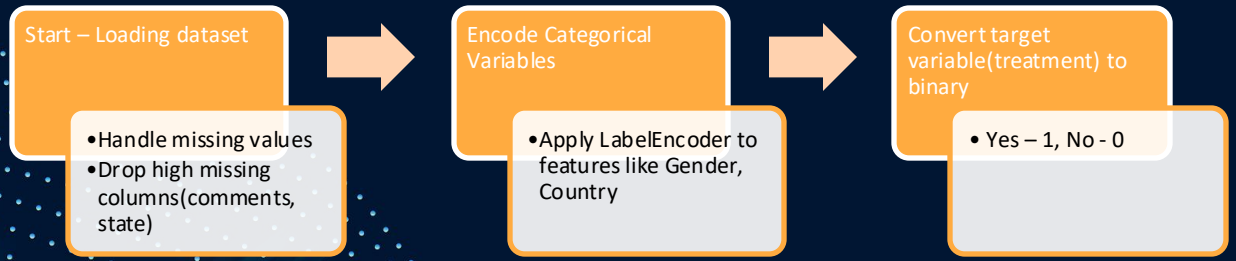
Data Understanding

- The dataset was 'survey_csv' which had a size of 1259 entries, It had key features which were Age, Gender, Family History, Remote work, Work stress.

- The target Variable was Treatment (Whether someone has sought mental health treatment)

Data Preparation

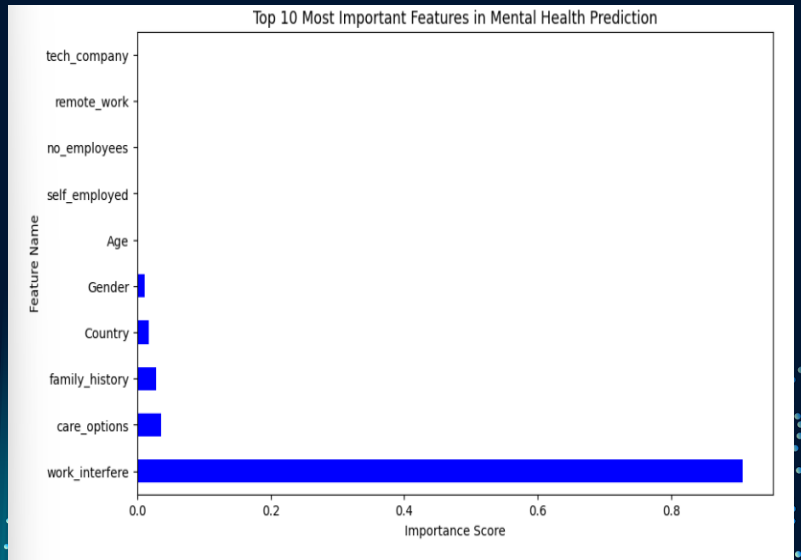
- Handling missing data(dropping/filling missing value)
- Encoding categorical variables using labelEncoder



Exploratory Data Analysis

- Choice of algorithms: Decision Trees and Logistic Regression.
- Model training with cross-validation and hyperparameter tuning.

- Distibution of key features like work_interfere, family history



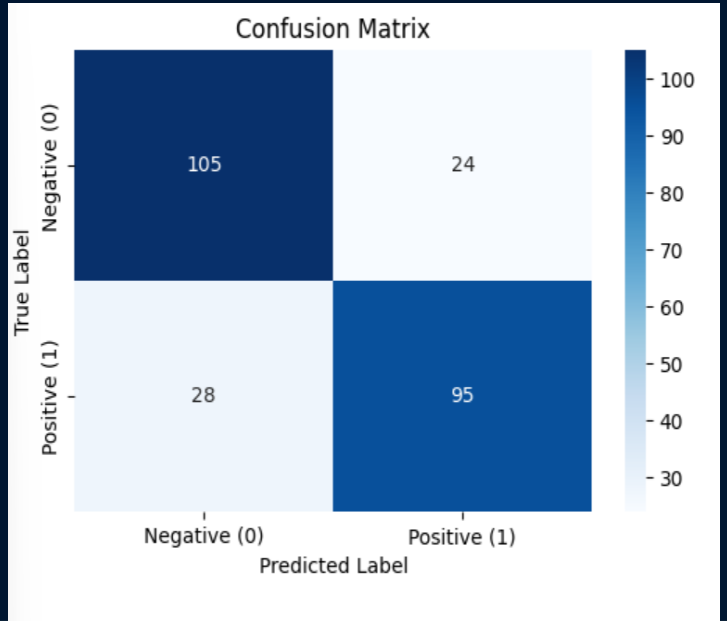
Modeling Approach

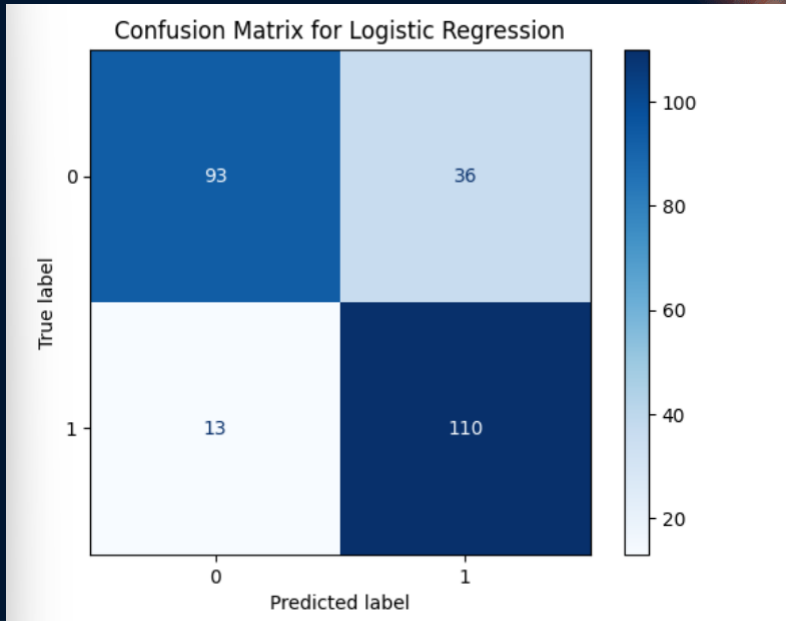
- Decision Tree
  - Tuned with GridSearchCV
  - Accurcy: 0.79 Recall: 0.77-0.81.

- Logistic regression
  - Balanced Class weights
  - SMOTE applied
  - Accuracy: 0.81 Recall for "At Risk": 0.89

Model Evaluation

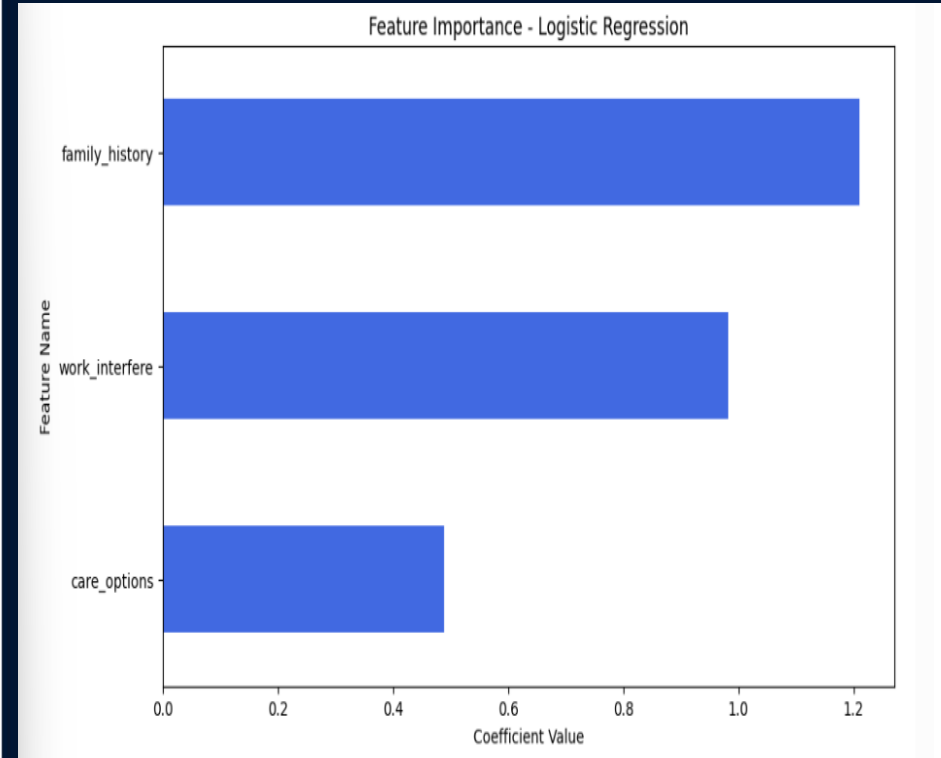
- Confusion matrix:

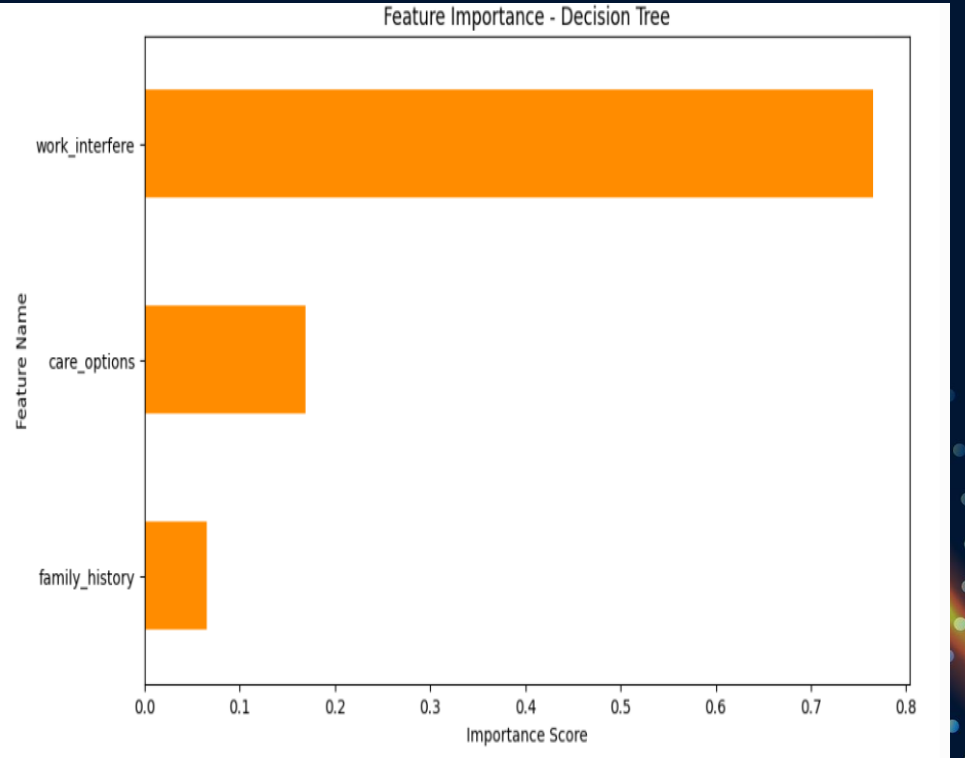




Feature Importance

Work_interfere, Family_history, care_options
Were the key features for both the Decisiontree and logistic models





Conclusion

Our predictive model effectively identifies employees at risk of mental health challenges by analyzing key predictors such as work stress (work_interfere) and family history. Feature selection enhances accuracy, ensuring high-risk individuals are not overlooked. The model enables proactive intervention through decision tree insights for wellness strategies and logistic regression for precise risk assessment. By balancing false positives and false negatives, it optimizes support allocation without overwhelming resources."

Thank you!

Cornelius Ngatia
cornelius.ngatia@student.moringaschool.com