Predictive Modeling for Early Detection of Mental Health Crises Among Employees

Introduction

Mental health concern is a growing issue in workplaces, highlighting the need for early detection and intervention to ensure employee well-being and productivity. Machine learning offers a promising field for predicting mental health crises based on various data sources. This study aims to develop a predictive model using machine learning algorithms to identify employees at risk. We will evaluate the model's accuracy in detecting such individuals, aiming for a system that can significantly improve early intervention efforts. The expected outcome is a model capable of accurately pinpointing employees at risk, along with an evaluation of its effectiveness using sensitivity, specificity, and overall accuracy.

Methodology

Data Collection and Preprocessing Methods

The proposed model is simulated using a single mental health in Tech survey dataset. The dataset consists of 1259 cases with 27 features. The data for this project was obtained from https://www.kaggle.com/osmi/mental-health-in-tech-survey. Originally, the data was collected through an online survey conducted in 2014 from respondents who self-reported having a mental health condition. It measures attitudes towards mental health and frequency of mental health disorders in the tech workplace. Cases of missing values, inconsistencies and invalid values were managed using various data cleaning and preprocessing techniques.

The data underwent a preprocessing pipeline involving data cleaning, feature engineering and normalization to ensure the quality and suitability of the input dataset. First, necessary libraries like pandas and numpy were loaded. The data was then loaded from its CSV format. To understand the data structure, we explored it by examining the initial rows and column names. This initial exploration helped identify data cleaning needs. We removed columns with high missing values (over 20%), such as comments, state, and timestamp. Inconsistencies in the gender column were addressed by combining entries with similar meanings. This involved cleaning the gender data and creating distinct gender groups.

Finally, the age data was categorized into ranges and encoded for better modeling.
Feature extraction was then done to identify relevant features from the dataset. This was done using Covariance Matrix and Variability comparison between categories of variables and some visualizations to see data relationship. The process resulted in the following most predictive features: 'Age', 'Gender', 'family_history', 'benefits', 'care_options', 'anonymity', 'leave', 'work_interfere', 'wellness_program', 'seek_help'.

ML/AI Model Development

After data preprocessing, different ML/AI models were employed including logistic regression, k-nearest neighbors (KNN) classifier, decision tree classifier, random forests, bagging, boosting and stacking. Each model was trained on the preprocessed data to learn the underlying patterns and relationships between the features and the target variable. Logistic regression was used to model the probability of a binary outcome. KNN classifier, made predictions based on the majority class among its nearest neighbors. Decision tree classifiers partitioned the feature space into regions to enable hierarchical decision making. Random forests which is an ensemble method combining multiple decision trees to improve prediction accuracy and robustness. Bagging and boosting further enhanced the performance of the prediction by aggregating multiple models and improving weak learners iteratively. Stacking combines the predictions of multiple base classifiers using a meta classifier for better results.

Evaluation of the Proposed System

To assess the effectiveness of each model in predicting treatment needs, we employed a range of evaluation metrics. Accuracy measured the overall percentage of correctly classified cases. The AUC score (Area Under the ROC Curve) assessed the model's ability to differentiate between patients requiring treatment and those who don't. To ensure a more reliable evaluation, we used cross-validated AUC, which involves multiple training and testing splits. Additionally, we considered precision (the proportion of true positives among predicted positives) and recall (the proportion of correctly identified positive cases). Finally, the F1-score provided a balanced view by combining precision and recall into a single metric. This comprehensive evaluation approach allowed us to identify the model that performed best in predicting depression treatment needs.

Analysis and Findings

Model Performance

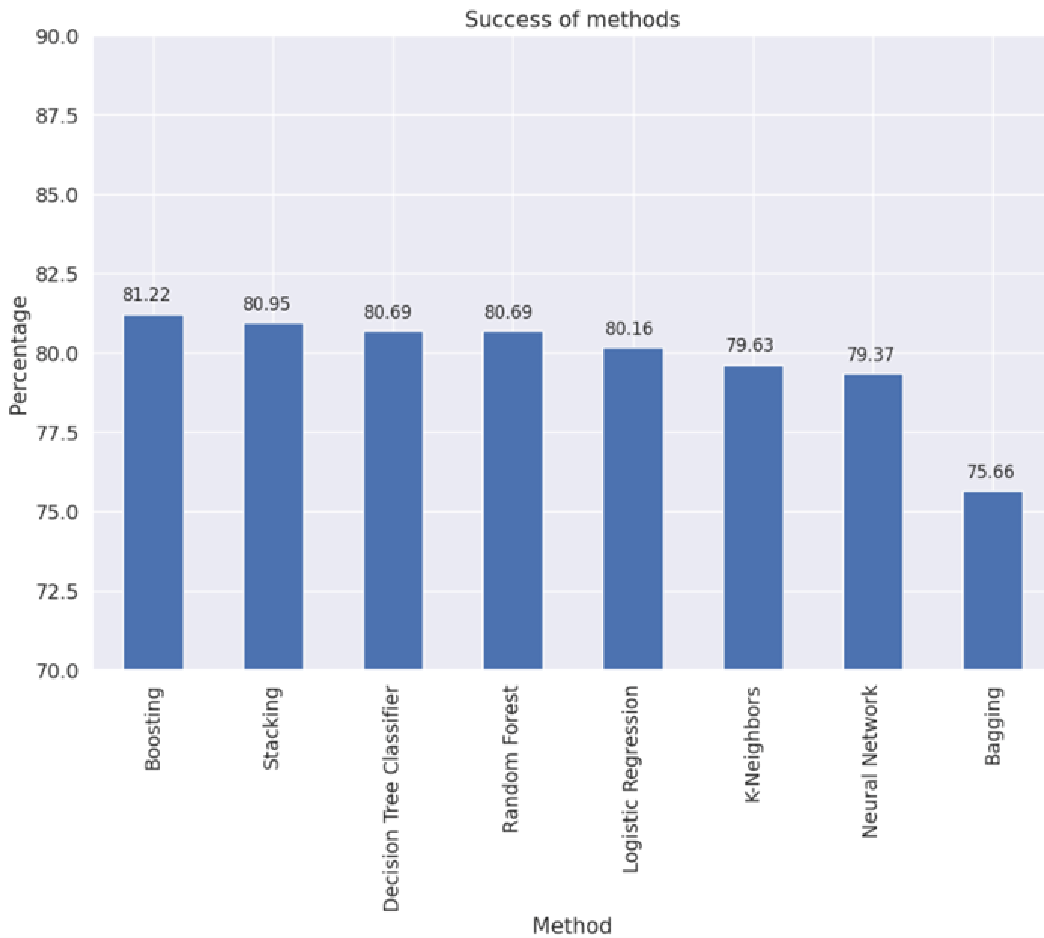
Our findings highlight the effectiveness of AI-based models in predicting the need for mental health treatment based on demographic and work-related factors. By comparing the performance of these models with baseline methods, such as traditional machine learning techniques, we demonstrated the superiority of AI-based approaches in terms of predictive accuracy and robustness.

The dataset containing patient features and treatment labels was split into training and testing sets. The training set (70%) was used to train the model and the testing set (30%) was used to evaluate the model. Different models were trained on the training data including logistic regression, K-Nearest neighbors (KNN), decision tree, random forest, bagging classifier, boosting, stacking classifier and deep neural network (DNN). The performance of each trained model was evaluated on the testing set. Here are the key metrics used:

Model	**Metric**			
	Accuracy	Precision	AUC Score	Cross-validated AUC
Logistic Regression	0.802	0.766	0.802	0.875
K-Nearest Neighbors (KNN)	0.796	0.748	0.797	0.878
Decision Tree	0.807	0.742	0.808	0.876
Random Forest	0.807	0.742	0.808	0.863
Bagging Classifier	0.757	0.732	0.757	0.857
AdaBoost Classifier (Boosting)	0.812	0.759	0.813	0.875

Stacking Classifier	0.810	0.773	0.810	0.842
Deep Neural Network (DNN)	0.79			

The performance metrics for each model were compared to identify the best performing model. The performance was also compared to baseline methods. Boosting had the highest accuracy (81.22%)

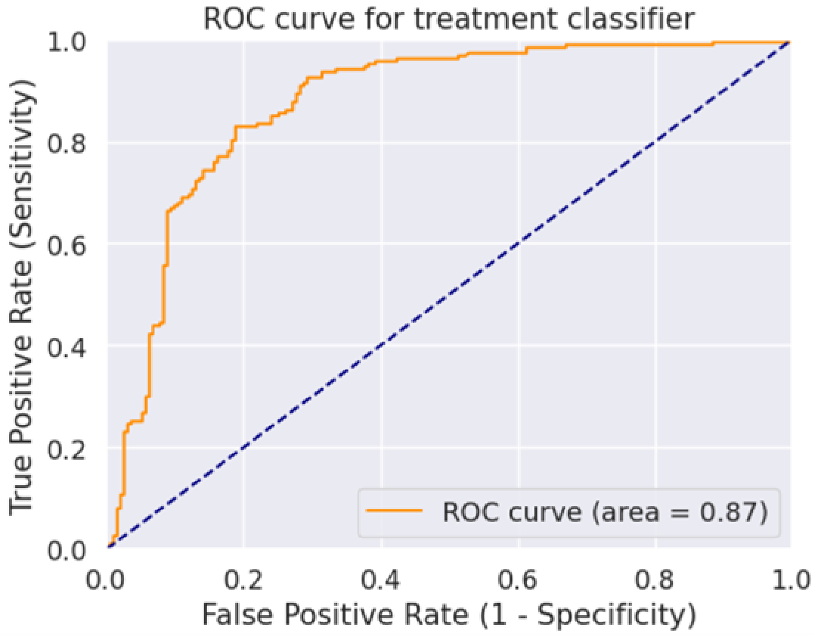


Predictions were created on the test set using the best method (AdaBoostClassifier)

Comparison with Baseline Methods

The model developed showcased promising performance. The model achieves an area under the receiver operating characteristic curve of 0.87 and a precision score of 0.76, predicting crises with an accuracy of

81% at a specificity of 72%. Predictions are clinically valuable in terms of either managing caseloads or mitigating the risk of crisis.



The curve goes from the bottom left corner to the top left corner, then curves towards the top right corner of the graph. This indicates a good model performance. The AUC for the ROC curve is 0.87, which is considered a good AUC value.

Challenges and Solutions

Some of the challenges encountered include limited access to large and high-quality datasets, feature selection, class imbalance and model overfitting. These were overcome by exploring publicly available depression-related datasets, utilizing correlation analysis to identify the most informative features, applying oversampling or under sampling techniques to balance the class distribution and implementing regularization techniques to prevent overfitting.

Conclusion

A predictive model using machine learning algorithms was successfully developed to identify employees at risk of mental health crises. Various machine learning models were developed and evaluated including logistic regression, decision trees, random forests and ensemble methods such as boosting. Promising results were achieved in terms of accuracy, precision and AUROC. The AdaBoost Classifier achieved the highest accuracy of 81.22% in detecting individuals at risk. The model demonstrates the potential for early detection of mental health crises in workplace settings.