Mental Health Model to Prevent Depression

Francesco Brigante 1987197, Giorgia Barboni 1885285, Federico Gerardi 1982783,
Zhuoya Shao 2108513

Abstract

Depression affects more than 280 million people globally, with young adults being particularly vulnerable. Our study developed a binary classification model to predict depression using a Kaggle dataset with features such as age, financial stress, and work pressures. Four preprocessing methods (marking missingness with -1, mean imputation, KNN imputation, and feature merging) and several models (Logistic Regression, XGBoost, K-NN, Naive Bayes, and others) were tested.
We also implemented PCA to evaluate the impact of dimensionality reduction on the performance of the models. We also utilized Explainable AI techniques to analyze feature importance and gain insights into the key factors influencing model predictions.
This model can aid in early detection and prevention, raise awareness, and support mental health professionals.

1 Introduction

Depression is potentially chronic and disabling, making it one of the most pressing global health issues. It affects approximately 280 million people worldwide. According to data from the CDC (Centers for Disease Control and Prevention) of 2019 [2], 21% of adults experiencing depressive symptoms in the last two weeks were between 18 and 29 years old. Our goal is to create a binary classification model to predict whether or not a person is depressed. This model aims to help psychologists and mental health experts in their work and to raise awareness of the problem. Effective early detection focuses on identifying and addressing risk factors before depression fully manifests, enabling timely intervention and potentially mitigating its long-term impact.

2 Datasets

We used a dataset from a Kaggle competition, consisting of a questionnaire about people's lives. The dataset contains approximately 140,000 rows with features such as `Age, Gender, Academic Pressure, Financial Stress, Work Pressure, Suicidal Thoughts, ...`
During preprocessing, we identified missing values in the original dataset, because both students (20% of the dataset population) and workers (80%) completed the questionnaire. For example, features like Academic Pressure and Academic Satisfaction are relevant only to students, while Work Pressure and Job Satisfaction pertain only to workers. To handle missing values, we compared different approaches:

- **Marking Missing Features Dataset**: Replace missing values with -1, indicating non-applicability.
- **Mean Dataset**: Fill missing values with their column mean.
- **KNN Dataset**: Use KNN to find feature neighbors and average their values to fill missing data.
- **Merged Features Dataset**: Merge contextually similar features (e.g., Work Pressure and Academic Pressure into a single "Pressure" feature).

After handling missing values, we applied the following preprocessing techniques:

- **Compression**: grouping less frequent categorical features into a new category called 'others'.
- **One-Hot Encoding**: converting categorical data into binary vectors.
- **Normalization**: applying Min-Max Scaling to scale numerical data into the range [0,1].

Each dataset was split into 70% training, 15% validation, and 15% test sets.

3 Models

3.1 Logistic Regression

We employed Logistic Regression to create a model that is fast, easy to use, and highly reliable for detecting potential depression cases. We tested this model on all generated datasets.

3.2 K-Nearest Neighbors (KNN)

KNN is effective for separable classes, so we tested it to verify class separability. We also tuned the hyperparameter k and evaluated the model on all generated datasets.

3.3 XGBoost

XGBoost is a scalable, end-to-end tree boosting system known for its efficiency with tabular datasets. We prioritized feature engineering and adjusted hyperparameters such as the number of trees and learning rate. This model was tested on all datasets [1].

3.4 Naive Bayes

Naive Bayes is based on the Bayesian theorem and is particularly effective for high-dimensional data. Although it assumes feature independence (which isn't the case for our specific task), its simplicity and fast training time made it worth testing [6].

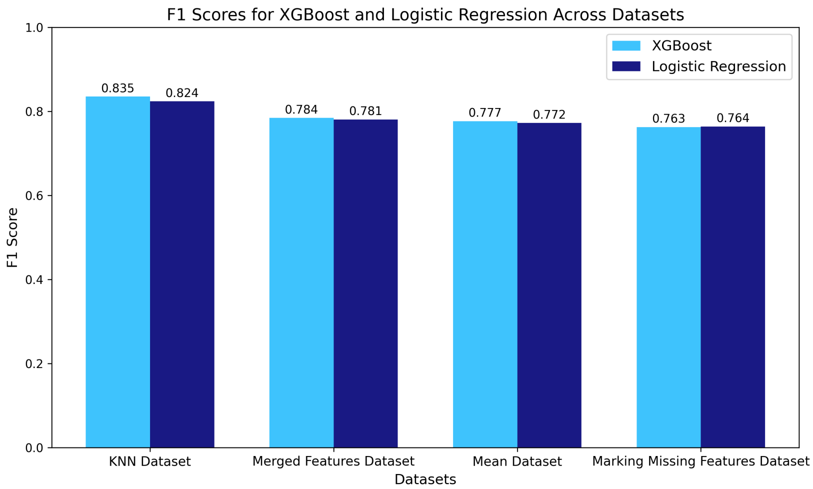
4 Models Evaluation

Given the imbalance present in the dataset, accuracy alone is not a reliable metric for assessing model performance. Instead, we focus on achieving an optimal balance between precision and recall. This approach aims to minimize the risk of incorrectly identifying depression cases as non-depression while avoiding unnecessary treatments.
To address these concerns, we selected F1-Score as the primary evaluation metric, since it provides a composite measure that rewards models with high sensitivity while posing a challenge to those with higher specificity [5].

Our results indicate that XGBoost and Logistic Regression consistently outperform the other models across all datasets. While KNN demonstrated reasonable performance, it fell short of the top-performing models. In contrast, Naive Bayes performed poorly, with results worse than random guessing.

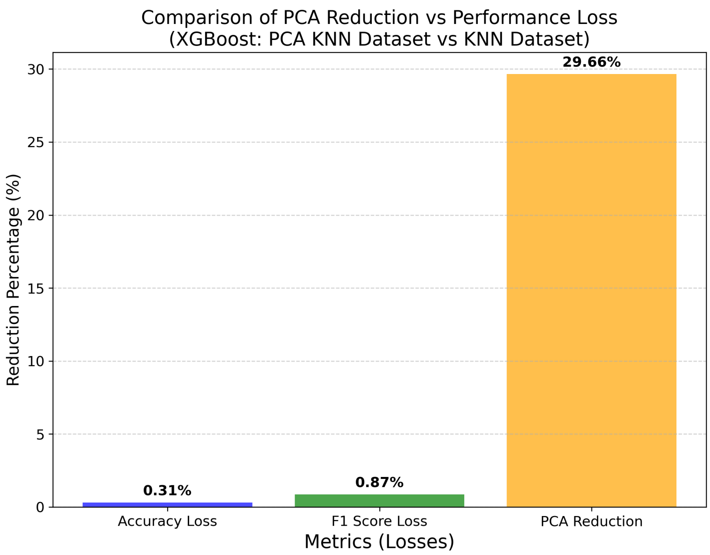
Model	Dataset	Accuracy	F1
XGBoost	KNN Dataset	0.941	0.835
Logistic Regression	KNN Dataset	0.937	0.824
XGBoost	Merged Features Dataset	0.926	0.784
Logistic Regression	Merged Features Dataset	0.925	0.781
XGBoost	Mean Dataset	0.924	0.777
Logistic Regression	Mean Dataset	0.923	0.772
Logistic Regression	Marking Missing Features Dataset	0.919	0.764
XGBoost	Marking Missing Features Dataset	0.918	0.763
K-Nearest Neighbors	KNN Dataset	0.917	0.757
K-Nearest Neighbors	Merged Features Dataset	0.91	0.729
K-Nearest Neighbors	Marking Missing Features Dataset	0.908	0.72
Naive Bayes	KNN Dataset	0.738	0.544
Naive Bayes	Marking Missing Features Dataset	0.722	0.523
Naive Bayes	Merged Features Dataset	0.683	0.495

The histograms below demonstrate that classifiers utilizing the KNN dataset as input consistently outperformed those employing other datasets. This suggests that the KNN dataset offers a smoother and more realistic approximation of real-world data compared to the alternatives.



5 Principal Component Analysis

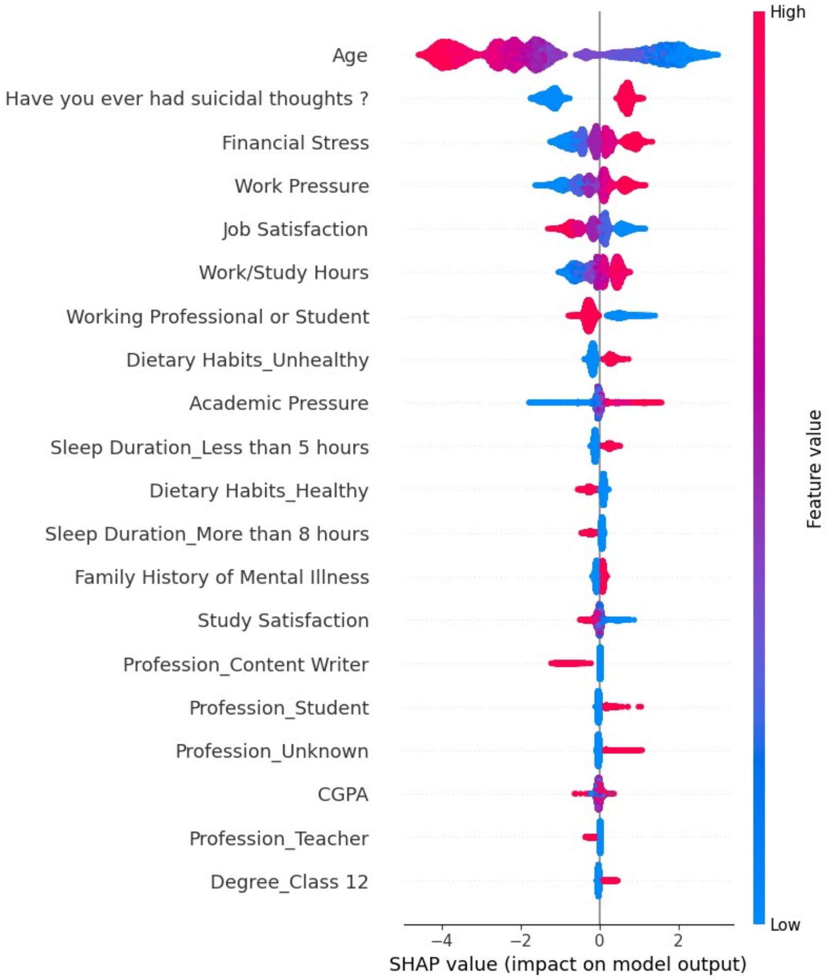
After identifying the optimal combination of model and dataset (XGBoost with the KNN dataset), we evaluated the impact of applying Principal Component Analysis. Retaining 95% of the total variance allowed us to reduce the dataset's features by 30%, resulting in only a minor performance decrease (-0.31% in Accuracy and -0.87% in F1 Score). These findings are interesting, as they demonstrate that the dataset's dimensionality can be significantly reduced with minimal impact on performance metrics.



6 Explainable AI

Interpretability allows to assess the model's decision-making process, which is crucial in healthcare to ensure that the predictions align with clinical knowledge.

We analyzed the predictions of our best-performing model (XGBoost with KNN-imputed dataset) using SHAP [4] to identify the factors that contributed the most to depression classification. Specifically we used TreeSHAP [3], a variant of SHAP efficient for tree-based machine learning models. SHAP assigns each feature an importance value for a specific prediction, which can be positive or negative, indicating how much the feature increases or decreases the likelihood of classifying someone as depressed compared to a baseline.



From the SHAP summary plot, the most influential features were:

- **Age**: Younger ages (blue) are associated with positive SHAP values (< 2.9), while older ages (red) are with negative SHAP values (> -4.5).
- **"Have you ever had suicidal thoughts?"**: Responding "Yes" (red) corresponds to positive SHAP values (around 1.1), whereas "No" (blue) to negative values (around -1.8).
- **Financial Stress**: Higher financial stress (red) is associated with positive SHAP values (< 1.2), while lower stress (blue) with negative values (> -1.4).
- **Work Pressure**: High work pressure (red) is associated with positive SHAP values (< 1.1), while low pressure (blue) with negative values (> -1.6).
- **Job Satisfaction**: Low job satisfaction (blue) is associated with positive SHAP values (< 1.2), while high satisfaction (red) with negative values (> -1.3).
- **Work/Study Hours**: Longer hours (red) have slightly positive SHAP values (< 0.9), while shorter hours (blue) have negative values (> -1.0).
- **Dietary Habits**: Unhealthy habits are linked to slightly positive SHAP values (< 0.6), while healthy habits are linked to slightly negative SHAP values (> -0.5).
- **Academic Pressure**: High academic pressure (red) shows positive SHAP values (< 1.6), while low pressure (blue) shows negative values (> -1.8).

- **Sleep Duration**: Short sleep durations (less than 5 hours) are linked to slightly positive SHAP values (< 0.5). Long sleep durations (more than 8 hours) instead are linked to slightly negative SHAP values (> -0.5).

- **Family History of Mental Illness**: A positive family history (red) has slightly positive SHAP values (< 0.2), while no history (blue) shows negative values (> -0.2).

These findings are consistent with common clinical observations, further supporting the notion that our model effectively captures meaningful indicators of depression. The application of SHAP enhances interpretability, fostering confidence among clinicians and patients in the model's decision-making process.

7 Conclusions and Future Works

In this study, we developed a binary classification model for depression that achieved high accuracy and F1 score, especially when using XGBoost on the KNN-imputed dataset. Our aim was to create a tool for early detection of depression that mental health professionals could use for identifying individuals at risk. SHAP analysis showed that our model aligns with known risk factors, such as suicidal ideation, financial stress, and age. This supports the model's reliability and suggests it could be practical in real-world settings.

7.1 Key Findings

Dataset Preprocessing: Among the four preprocessing methods tested, the KNN-imputed dataset outperformed others, indicating that filling in missing data using similar "neighbors" is highly effective. Combining related features (like academic and work pressure) in the merged features dataset also proved to be effective, emphasizing the importance of domain knowledge.

Interpretability: SHAP gave us insights into the top predictors of depression and helped confirm that they matched clinical expectations, increasing our confidence that the model could be useful in practice to support healthcare professionals and advance research.

7.2 Future Directions

Dataset Integrations: To improve generalization, future research should prioritize the collection of data from diverse populations, with careful consideration of cultural and socioeconomic variations. Integrating survey data with complementary sources, such as metrics from wearable devices, has the potential to further enhance the model's predictive accuracy and robustness.

Explainable AI Enhancements: Integrating other interpretability techniques that complement SHAP could provide an even clearer understanding of how each factor influences the model's predictions.

Multi-class Classification: Training the model to predict depression severity (mild, moderate, severe) would offer clinicians more detailed information for customizing interventions.

By addressing the outlined future directions, we aim to create a more inclusive, transparent, and accurate tool that aligns with real-world healthcare needs, addressing critical health challenges like depression.

Members Roles

- Francesco Brigante, Zhuoya Shao: datasets creation and imputation of missing values, datasets preprocessing (compression of categorical features, one hot encoding, normalization), creation of PCA datasets.

- Federico Gerardi, Giorgia Barboni: model selection and testing, results analysis, feature importance using explainable AI.

References

[1] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *arXiv preprint arXiv:1603.02754* (2016). URL: https://arxiv.org/abs/1603.02754.

[2] Centers for Disease Control and Prevention. *Symptoms of Depression Among Adults: United States, 2019*. URL: https://www.cdc.gov/nchs/products/databriefs/db379.htm#:~:text=The%20percentage%20of%20adults%20who%20experienced%20any%20symptoms%20of%20depression,or%20severe%20symptoms%20of%20depression.

[3] Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. "Consistent Individualized Feature Attribution for Tree Ensembles". In: *arXiv preprint arXiv:1802.03888* (2018). URL: https://arxiv.org/abs/1802.03888.

[4] Scott M. Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems*. 2017. URL: https://arxiv.org/abs/1705.07874.

[5] David M. Powers. "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation". In: *Proceedings of the Workshop on Evaluation Methods for Machine Learning at AAAI-06*. 2006. URL: https://aaai.org/papers/ws06-06-006-beyond-accuracy-f-score-and-roc-a-family-of-discriminant-measures-for-performance-evaluation/.

[6] Qingping Tao et al. "Bayes and Naive Bayes Classifier". In: *arXiv preprint arXiv:1404.0933* (2014). URL: https://arxiv.org/abs/1404.0933.