Understanding and Predicting Depression to Enhance Mental Health Interventions

Group 8

Pham Van Hung Do Ba Huy Tran Vi Khang Le Dang Khoa
Do Phuc Kien Nguyen Duc Lap Le Tran Bao Loi

University of Information Technology

December 22, 2024

Table of Contents

1. Introduction
2. Dataset
3. Exploratory Data Analysis
4. Method
5. Evaluation
6. Conclusion
7. Demo

Table of Contents

1. Introduction
2. Dataset
3. Exploratory Data Analysis
4. Method
5. Evaluation
6. Conclusion
7. Demo

Introduction

Context

- Mental health, particularly depression, is a growing global concern. Understanding contributing factors is essential for creating effective interventions.
- Timely identification of at-risk individuals is crucial to providing support before the situation worsens.

Introduction

Current challenges

Many organizations struggle to predict depression using the available data, making it difficult to intervene proactively and prevent the escalation of mental health issues.

Role of Depression Prediction

- Identify individuals at risk and prioritize support for them.
- Create targeted campaigns to raise awareness about key risk factors.
- Monitor and improve the effectiveness of mental health interventions based on data.

Introduction

Input

Individual characteristics: Personal details, Work/study status, Health and lifestyle, Psychological and financial factors, and Educational background.

Output

Depression status prediction:

- Label 0: A person is predicted **not** to be depressed.
- Label 1: A person is predicted to be depressed.

Table of Contents

1. Introduction
2. Dataset
3. Exploratory Data Analysis
4. Method
5. Evaluation
6. Conclusion
7. Demo

Dataset

Depression Survey/Dataset for Analysis

- The dataset was collected as part of a comprehensive survey designed to identify factors contributing to depression risk among adults.
- It was gathered through an anonymous survey conducted between January and June 2023, during a time when COVID-19 had significant impacts on mental health.

Dataset

```
RangeIndex: 140700 entries, 0 to 140699
Data columns (total 20 columns):
 #   Column                                 Non-Null Count   Dtype
---  ------                                 --------------   -----
 0   id                                     140700 non-null  int64
 1   Name                                   140700 non-null  object
 2   Gender                                 140700 non-null  object
 3   Age                                    140700 non-null  float64
 4   City                                   140700 non-null  object
 5   Working Professional or Student        140700 non-null  object
 6   Profession                             104070 non-null  object
 7   Academic Pressure                      27897 non-null   float64
 8   Work Pressure                          112782 non-null  float64
 9   CGPA                                   27898 non-null   float64
 10  Study Satisfaction                     27897 non-null   float64
 11  Job Satisfaction                       112790 non-null  float64
 12  Sleep Duration                         140700 non-null  object
 13  Dietary Habits                         140696 non-null  object
 14  Degree                                 140698 non-null  object
 15  Have you ever had suicidal thoughts ?  140700 non-null  object
 16  Work/Study Hours                       140700 non-null  float64
 17  Financial Stress                       140696 non-null  float64
 18  Family History of Mental Illness       140700 non-null  object
 19  Depression                             140700 non-null  int64
dtypes: float64(8), int64(2), object(10)
memory usage: 21.5+ MB
```

Table of Contents

1. Introduction
2. Dataset
3. Exploratory Data Analysis
4. Method
5. Evaluation
6. Conclusion
7. Demo

Exploratory Data Analysis: Missing Values

id	0
Name	0
Gender	0
Age	0
City	0
Working Professional or Student	0
Profession	36630
Academic Pressure	112803
Work Pressure	27918
CGPA	112802
Study Satisfaction	112803
Job Satisfaction	27910
Sleep Duration	0
Dietary Habits	4
Degree	2
Have you ever had suicidal thoughts ?	0
Work/Study Hours	0
Financial Stress	4
Family History of Mental Illness	0
Depression	0

Figure: Missing values

Exploratory Data Analysis: Target Distribution

- 0: Not depression.
- 1: Depression.

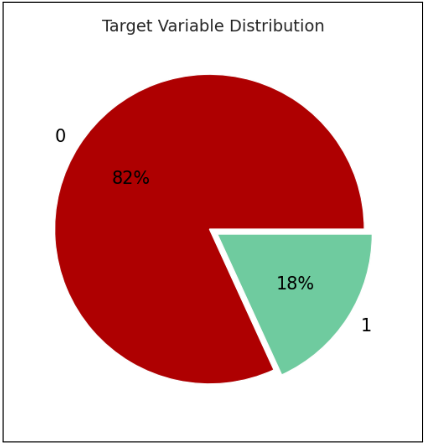
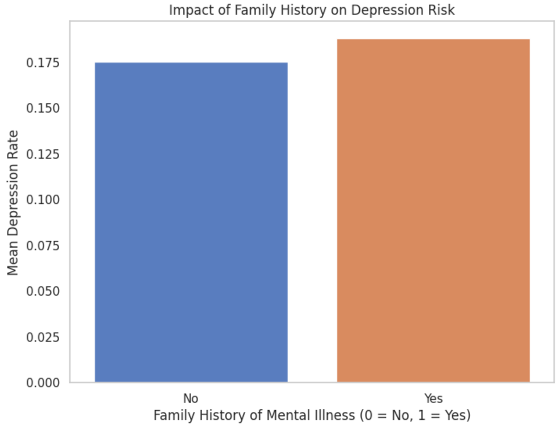
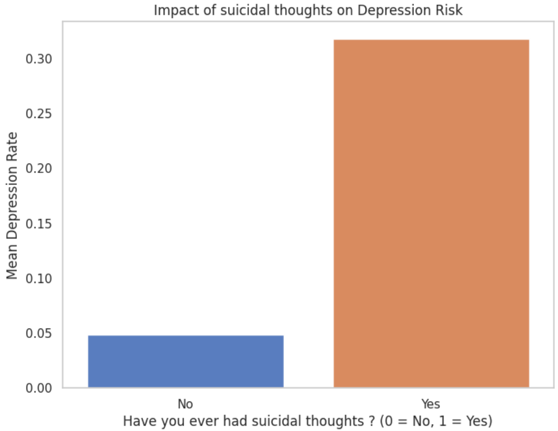


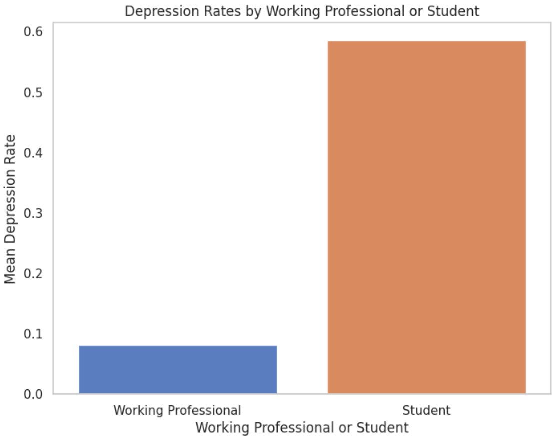
Figure: This is an example image.

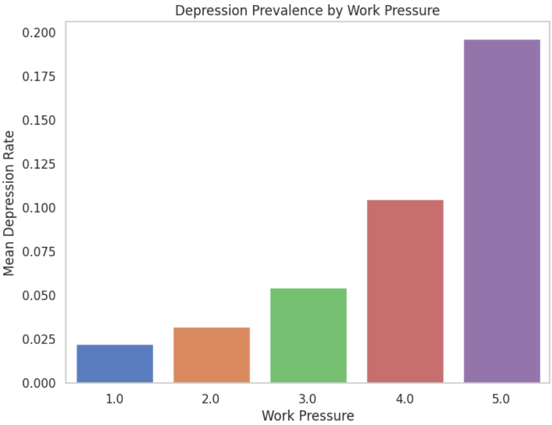
Exploratory Data Analysis: Features Distribution



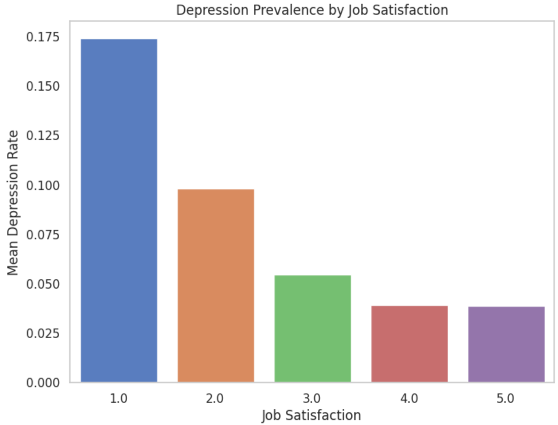


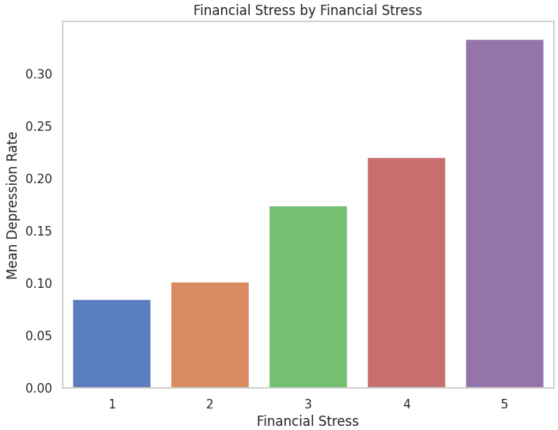
Exploratory Data Analysis: Features Distribution



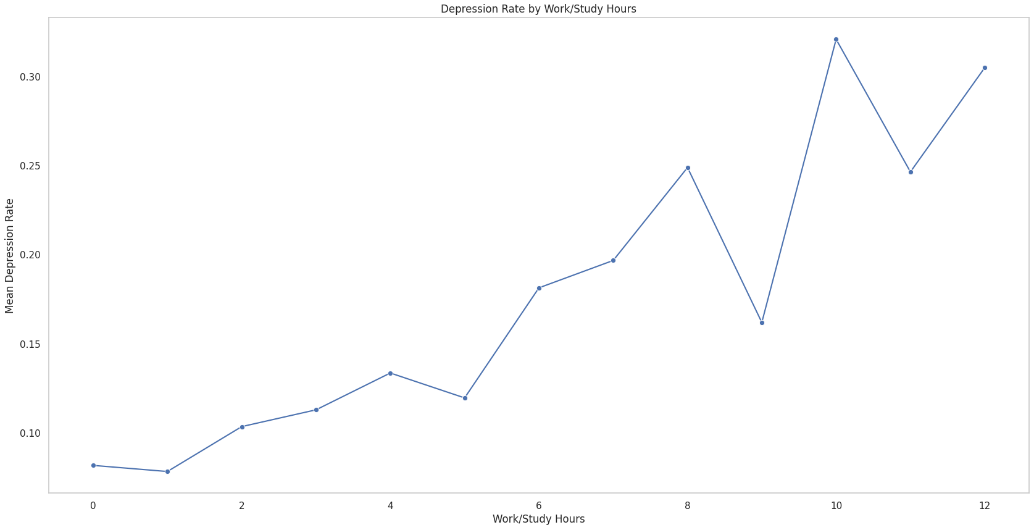


Exploratory Data Analysis: Features Distribution

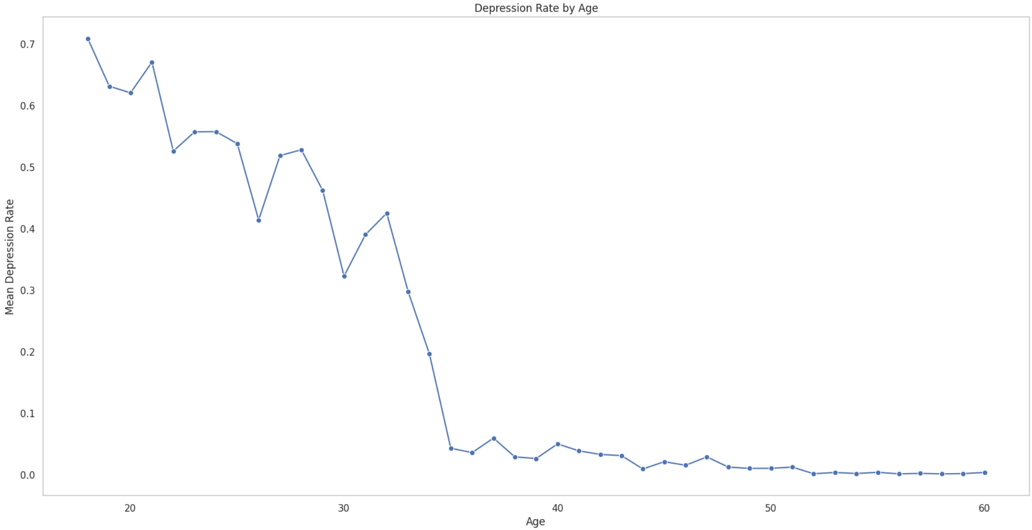




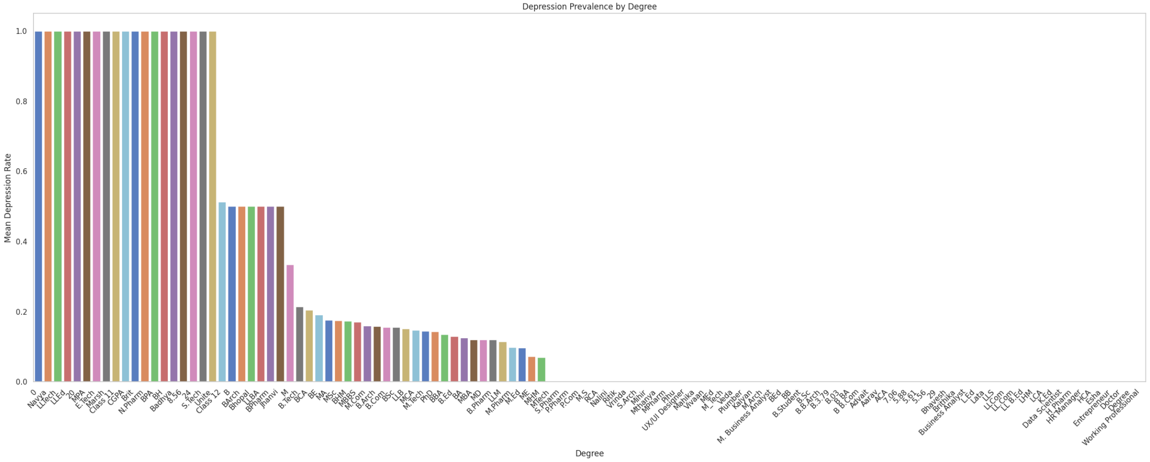
Exploratory Data Analysis: Features Distribution



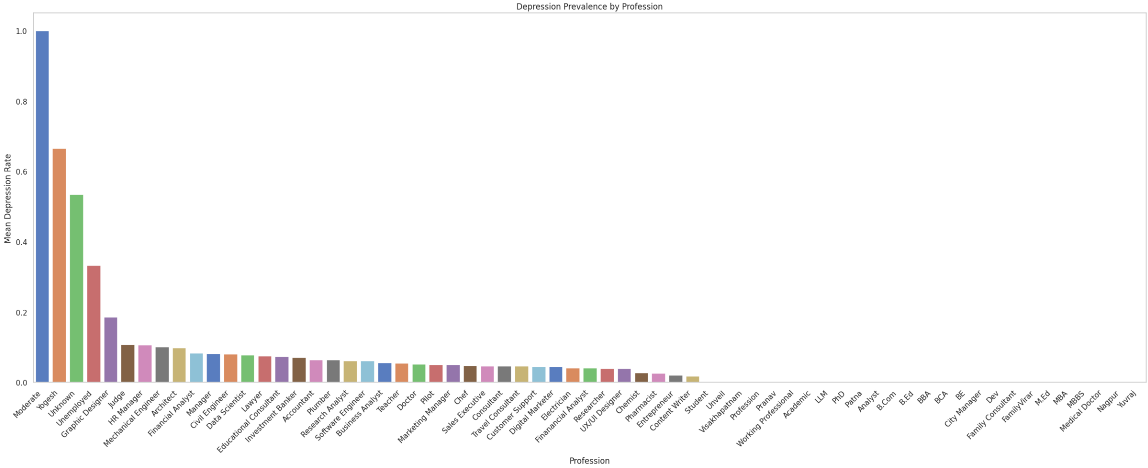
Exploratory Data Analysis: Features Distribution



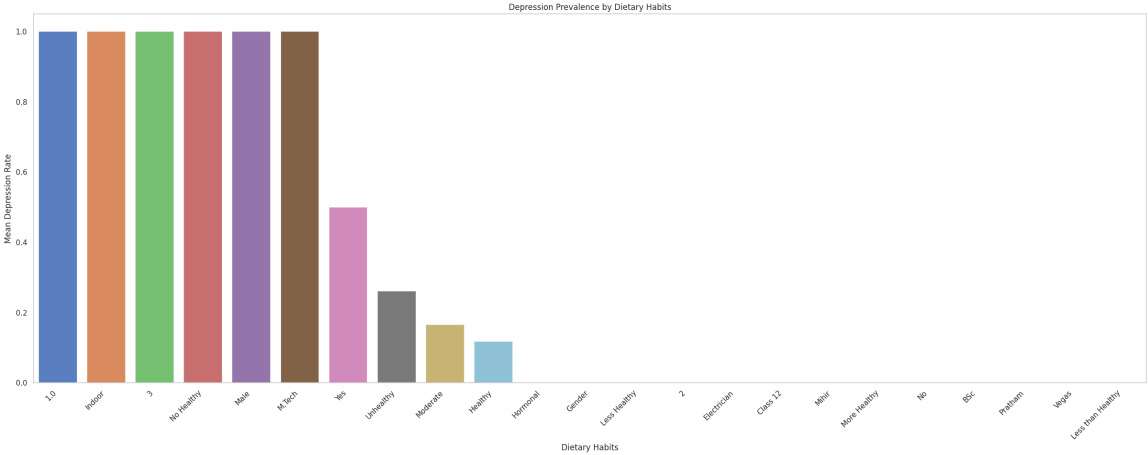
Exploratory Data Analysis: Features Distribution



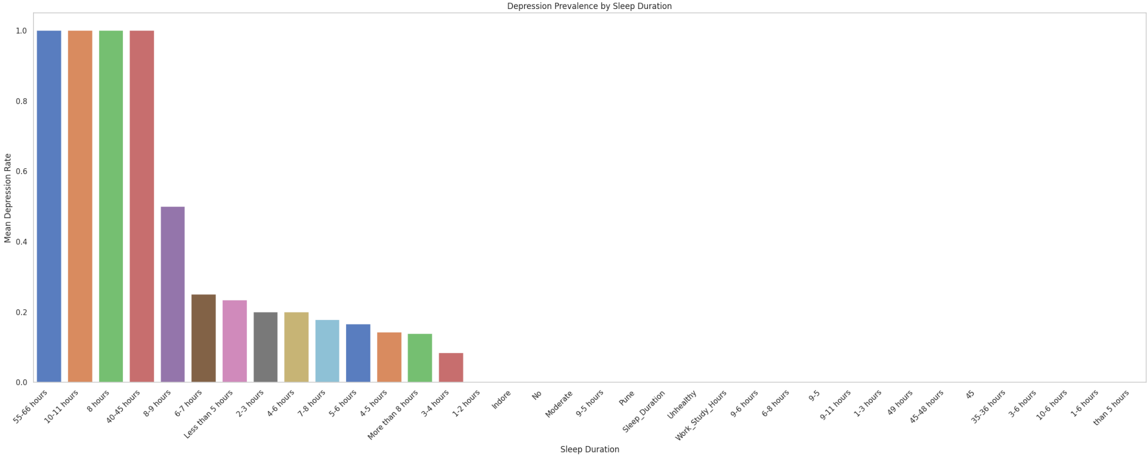
Exploratory Data Analysis: Features Distribution



Exploratory Data Analysis: Features Distribution



Exploratory Data Analysis: Features Distribution



Exploratory Data Analysis: Correlation

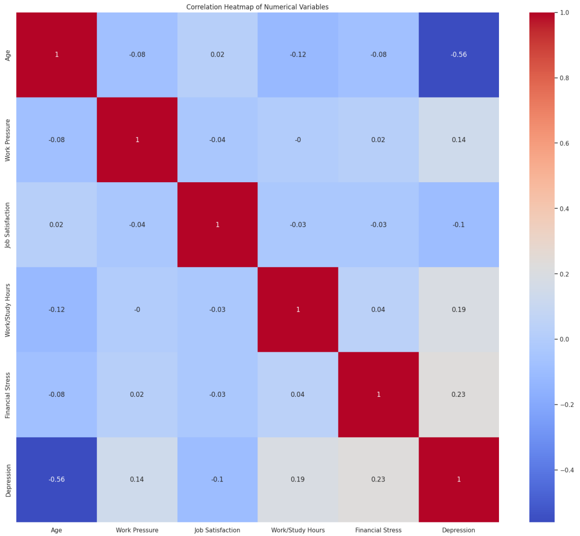


Table of Contents

1. Introduction
2. Dataset
3. Exploratory Data Analysis
4. Method
5. Evaluation
6. Conclusion
7. Demo

Pre-processing: Missing Values

Process missing values:

- Set threshold of 70 % for null values, remove columns exceeding it
- Filling null values with median for numerical columns
- Filling null values with "Unkown" or mode values for category

Apply Label Encoding to each categorical column.

Gender	Age	City	Working Professional or Student	Profession	Work Pressure	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Family History of Mental Illness	Depression
0	49	50	1	10	5	2	29	7	33	0	1	2	0	0
1	26	93	1	55	4	3	27	20	63	1	7	3	0	1
1	33	97	0	59	3	3	15	7	21	1	3	1	0	1
1	22	64	1	55	5	1	27	15	28	1	10	1	1	1
0	30	37	1	9	1	1	15	20	28	1	9	4	1	0

Figure: The dataset after transformation

Modeling

Methods
Decision Tree
Random Forest
LightGBM
Multi-Layer Perceptron

Table: Models name table

**Apply Standard Scaler for train data before using Neural Network*

Table of Contents

1. Introduction
2. Dataset
3. Exploratory Data Analysis
4. Method
5. Evaluation
6. Conclusion
7. Demo

Metrics and Evaluation: Confusion matrix

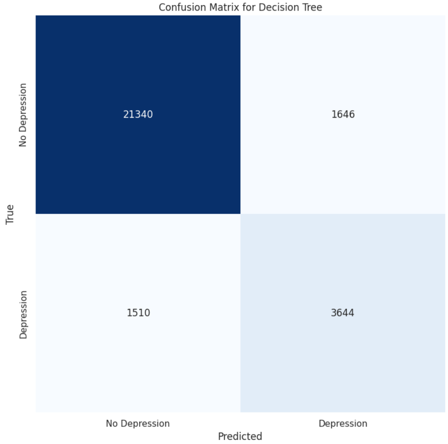


Figure: Confusion Matrix for Decision Tree

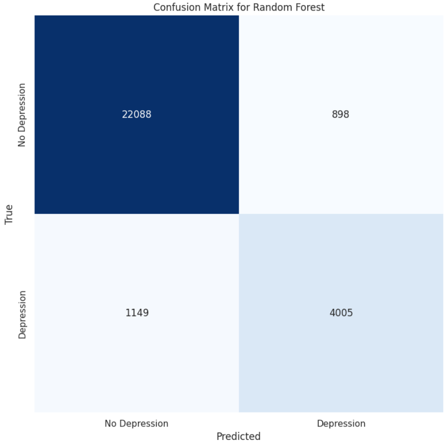


Figure: Confusion Matrix for Random Forest

Evaluation: Confusion matrix

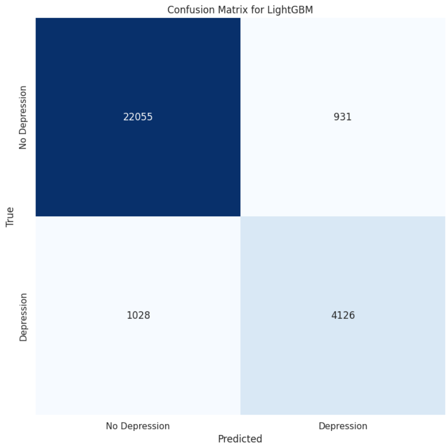


Figure: Confusion Matrix for LightGBM

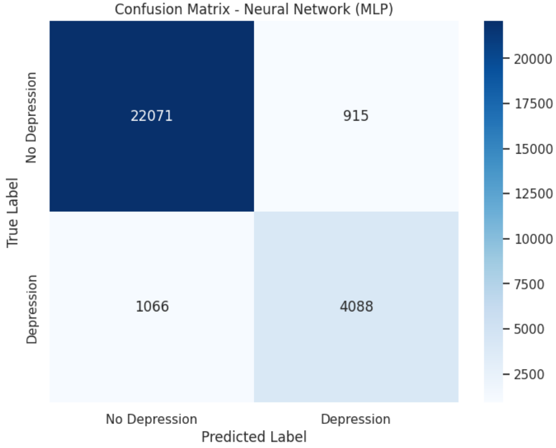


Figure: Confusion Matrix for Multi-Layer Perceptron

Metrics and Evaluation: ROC curve

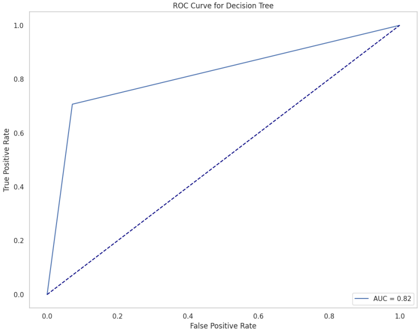


Figure: ROC Curve for Decision Tree

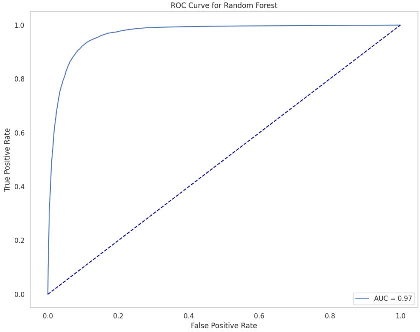


Figure: ROC Curve for Random Forest

Metrics and Evaluation: ROC curve

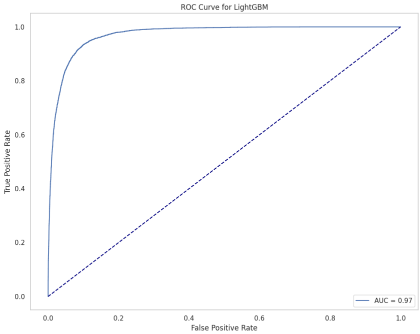


Figure: ROC Curve for LightGBM

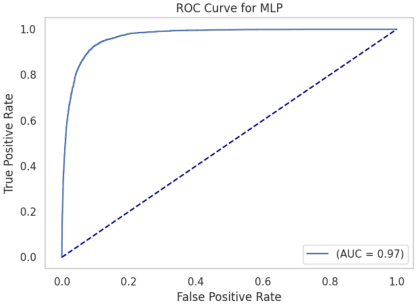


Figure: ROC Curve for MLP

Metrics and Evaluation: Precision, Recall, F1-score, Accuracy

Methods		**Precision**	**Recall**	**F1-score**	**Accuracy**
Decision Tree	No Depression	0.93	0.93	0.93	0.89
	Depression	0.69	0.71	0.70	
Random Forest	No Depression	0.95	0.96	0.96	0.93
	Depression	0.82	0.78	0.80	
LightGBM	No Depression	0.96	0.96	0.96	0.93
	Depression	0.82	0.80	0.81	
MLP	No Depression	0.95	0.96	0.96	0.93
	Depression	0.81	0.79	0.80	

Table of Contents

1. Introduction
2. Dataset
3. Exploratory Data Analysis
4. Method
5. Evaluation
6. Conclusion
7. Demo

Conclusion

Conclusion

- The goal is to use the data to explore factors causing depression and predict the likelihood of experiencing it. Create tools to help organizations diagnose more easily.
- Models such as Random Forest, LightGBM, and Multi-layer Perceptron have demonstrated strong potential in results.

Future Work

Conclusion

- **Data Augmentation**: Use data augmentation techniques to enhance the accuracy and generalizability of the predictive model.
- **In-depth Analysis:** Further investigate specific factors that strongly influence depression, such as social, economic, and environmental factors.

Table of Contents

1. Introduction
2. Dataset
3. Exploratory Data Analysis
4. Method
5. Evaluation
6. Conclusion
7. Demo

Demo

References

Walter Reade; Elizabeth Park
Exploring Mental Health Data

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu.
Lightgbm: A highly efficient gradient boosting decision tree

Marius-Constantin Popescu, Valentina E Balas, Liliana Perescu-Popescu, and Nikos Mastorakis.
Multilayer perceptron and neural networks.

Gilles Louppe.
Understanding random forests: From theory to practice, 2015.

Marius-Constantin Popescu, Valentina E Balas, Liliana Perescu-Popescu, and Nikos Mastorakis.
Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems

The End

Tasks Assignment Table

	Van Hung	Ba Huy	Vi Khang	Dang Khoa	Phuc Kien	Duc Lap	Bao Loi
Research and summarize knowledge	x	x	x	x	x	x	x
Design the slides using Latex				x		x	
Implement code demonstration					x		x
Presentation			x			x	
Write the report	x	x					x
Estimate percentage	14%	14%	14%	14%	14%	15%	15%