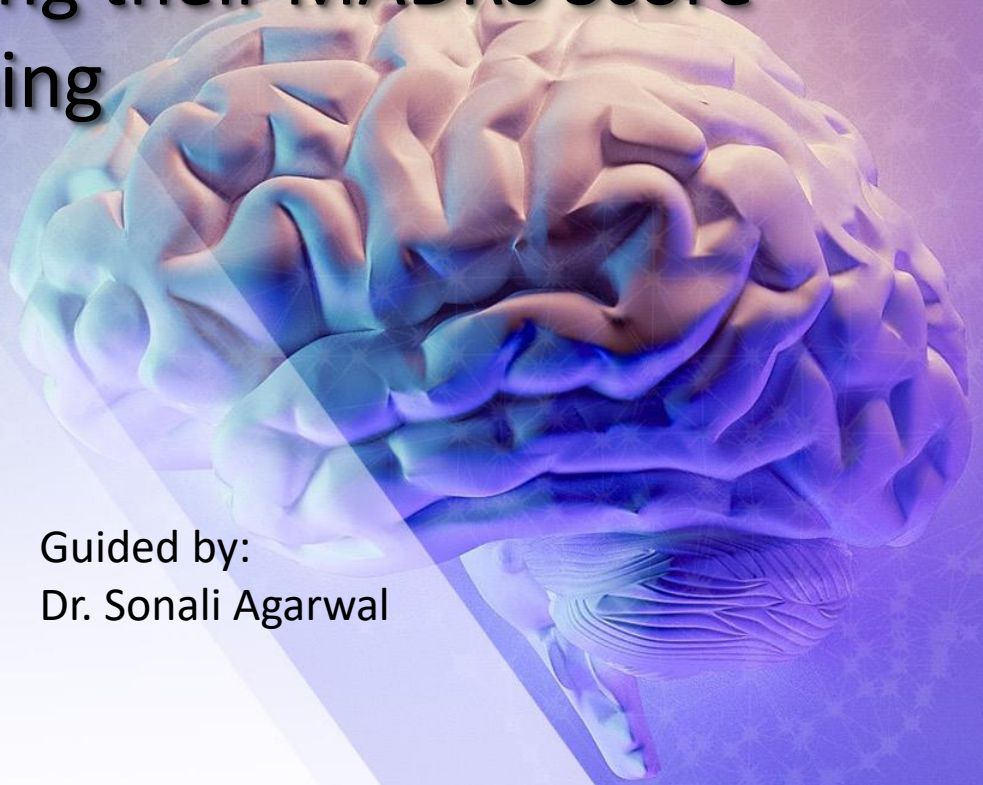
Classification of Depressed and Non-Depressed Subjects and Predicting their MADRS Score Using Machine Learning



Submitted by:
Shivam Kasat

Guided by:
Dr. Sonali Agarwal

Overview

- Objectives
- Using Awesome Backgrounds
- Engage your Audience
- Capture Audience Attention

Introduction

- What is Depression?
- What is motor activity data?
- How can we use this sensor based data?
- Use of Machine learning in Psychology.

Objectives

- Classification of depressed and Non-depressed subjects based on their motor activity data.
- Multi-class classification of subjects into No-depression, Mild-depression, severe-depression.
- Prediction of MADRS score of the subject which later can be used for classification.

Motivation

- The COVID-19 pandemic
- Change in lifestyle
- Increasing work load
- Depression in teenagers

Associated Challenges & Research Gap

- Dataset size
- Imbalanced dataset
- Improvisation in model performance
- Multi-class classification

Dataset Explained

- Dataset size
- Data type
- Data content

```
scores_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55 entries, 0 to 54
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   number      55 non-null     object 
 1   days        55 non-null     int64  
 2   gender      55 non-null     int64  
 3   age         55 non-null     object 
 4   afftype     23 non-null     float64
 5   melanch     20 non-null     float64
 6   inpatient   23 non-null     float64
 7   edu         53 non-null     object 
 8   marriage    23 non-null     float64
 9   work        23 non-null     float64
 10  madrs1      23 non-null     float64
 11  madrs2      23 non-null     float64
dtypes: float64(7), int64(2), object(3)
memory usage: 5.3+ KB
```

Work Plan

- Binary Classification
  - Dataset Preparation
  - Classification task
- Multiclass Classification
  - Dataset preparation
  - Classification task
- MADRS score prediction
  - Dataset Preparation
  - Prediction task

Binary Classification

- Dataset Preparation
- Data Modeling

Binary Classification Results

- Without Oversampling

	Accuracy	F1-Score	Precision	Recall
AdaBoost	0.71	0.56	0.56	0.56
XGBoost	0.74	0.58	0.61	0.56
CatBoost	0.78	0.63	0.65	0.60
GBM	0.72	0.52	0.57	0.48
1D CNN	0.63	0.65	0.66	0.60

Binary Classification Results

- With oversampling on training set

	Accuracy	F1-Score	Precision	Recall
AdaBoost	0.70	0.70	0.70	0.71
XGBoost	0.72	0.72	0.72	0.73
CatBoost	0.78	0.78	0.78	0.78
GBM	0.76	0.76	0.67	0.76

Multi-class Classification

- Dataset Preparation
- Data Modeling

Multi-class Classification Results

- Without Oversampling

	Accuracy	F1-Score	Precision	Recall
AdaBoost	0.82	0.82	0.82	0.82
XGBoost	0.86	0.86	0.87	0.86
CatBoost	0.85	0.85	0.85	0.85
GBM	0.84	0.84	0.85	0.84
1D CNN	0.83	0.84	0.85	0.84

Multi-class Classification Results

- With Oversampling on training set

	Accuracy	F1-Score	Precision	Recall
AdaBoost	0.89	0.89	0.89	0.89
XGBoost	0.87	0.87	0.87	0.87
CatBoost	0.89	0.89	0.89	0.89
GBM	0.88	0.88	0.88	0.88
1D CNN	0.83	0.84	0.83	0.85

MADRS Score Prediction

- With and without oversampling no major changes in results.

	MAE	MSE
AdaBoost	1.68	8.60
XGBoost	1.32	7.01
1D CNN	2.31	9.56

Conclusion

- Tree based ensemble models can be used to classify subjects as depressed and non-depressed, Also they can be used to classify subjects into no-depression, mild-depression & severe depression. CatBoost ensemble model can be preferred over others.
- Tree based ensemble model performed well and thus can be used to predict MADRS.
- The results can be improved further with more data, as we saw improvement in results with oversampling.
- This work can be used in smartwatche & smartphone applications to warn users about mental health.