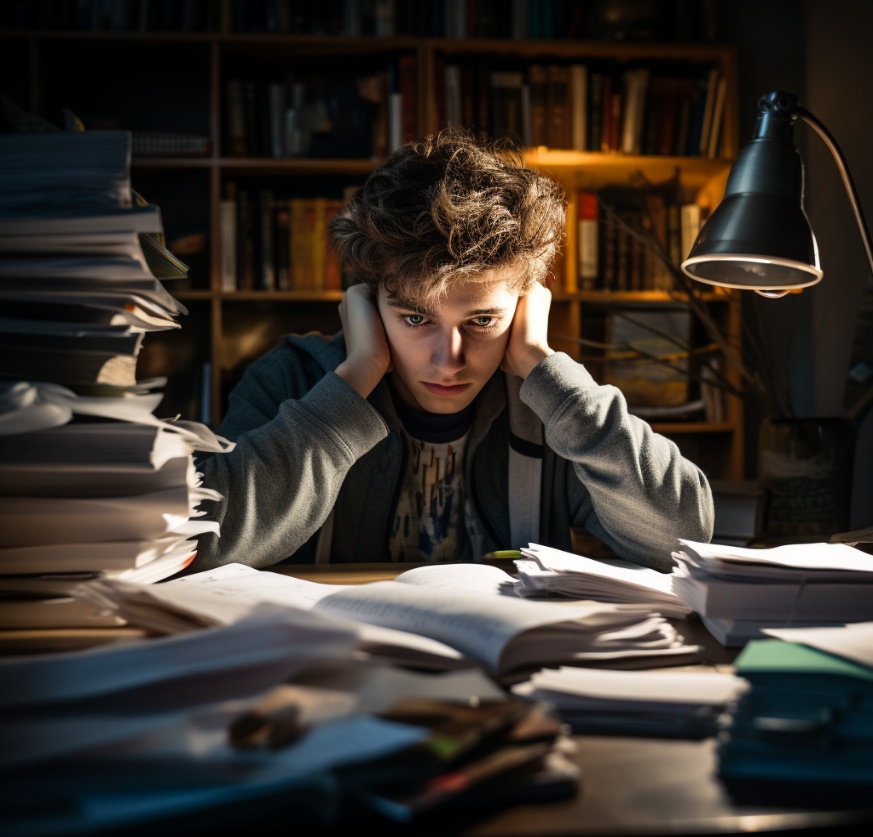
**Supervised Learning — Classification**

**Group-1(Pandas)**



**Niraj**

**Akshar**

**Vandita**

**Sashidhar**

**Index**

1. **Introduction**  
   1.1 Problem Definition  
   1.2 Why is it Worth Solving?
2. **Data Description**  
   2.1 Source of the Data  
   2.2 Type of Data Used
3. **Methodology**  
   3.1 Supervised Learning Overview
4. **Data Preparation**  
   4.1. What did I do with the data first?

4.2. How did I prepare the data?

4.3. What model did we use?

1. **Model Building**  
   5.1 What was your modeling process? Specifically, which algorithms and parameters did you use and why?
2. **Modeling Process**  
   6.1 What were your results

6.2 How did you evaluate the performance of your model? What metrics did you use?

1. **What did the charts show**  
   7.1 Bar Plots  
   7.2 Box Plots  
   7.3 Box Plots
2. **Conclusion**  
   8.1. What improvements would you like to make in future?

8.2. How do you think the solution could be used in real life?

8.3. What value do you think the solution will have to the client?

8.4. What did you learn through this project?

**1.Introduction**

* 1. **What is the problem?**  
     The problem we are trying to solve is predict whether a student is depressed or not based on their lifestyle and daily habits and GPA performance.
  2. **Why is it worth solving?**  
     Student mental health is serious as it shows many students are committing suicide attempt and getting suicidal thoughts because of stress and depression, Which can affect their school performance and life. If we can find patterns in their behavior, we might be able to help them early, that’s why it’s important.

1. **Data Description**
   1. **What is the source of your data ?**  
      The data is from a Kaggle on student Depression.

(<https://www.kaggle.com/datasets/adilshamim8/student-depression-dataset>)

* 1. **What kinds of data are you using?**

It’s classification data which contains numerical and categorical data.

* **Numbers:** age, hours of sleep, how many hours they use social media, and how many hours they study or exercise.
* **Text labels:** like whether they are depressed (Yes or No).

**3.Methodology**

**3.1 Supervised Learning Overview**

We are using supervised learning, which means the model learns from examples we give it. Classification is used when the thing we are trying to predict has categorical data like Depressed or Not Depressed.

**4.Data Preparation**

**4.1. What did I do with the data first?**

* We understood the business segments.
* We used Data Glance to check the data types and Missing values.
* We used Data Wrangling to clean the data such as Missing / Null.

**4.2. How did I prepare the data?**

* We cleaned the data by checking for missing or blank values by following wrangling steps.
* We changed Numerical labels of (1 or 0) Depressed or Not Depressed to Text Labels (Yes or NO).
* We also made sure that columns don’t have any missing values and compared the accuracy to use for modeling.

**4.3. What model did we use?**

* We used Supervised Learning.
* Our model was a classification model that shows whether the student is depressed or not by using (“YES or NO”)
* A screenshot of a computer

  AI-generated content may be incorrect.We used **Logistic Regression Classifier,** which is a simple and popular model for this kind of problem.



**5.Model Building**

**5.1. What was your modeling process? Specifically, which algorithms and parameters did you use and why?**

* We used Logistic regression classifier.
* It’s easy to understand the target value.
* It works well for problems where the answer is Yes or No especially for Categorical data type.

**6.** **Modeling Process**

**6.1. What were your results?**

Our model, trained on 80% of the data and tested on the remaining 20%, was able to accurately predict whether a student was experiencing depression most of the time.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.**6.2. How did you evaluate the performance of your model? What metrics did you use?**

* First, we looked at the accuracy like how many answers it got right.
* Then we used a confusion matrix to see how often it made mistakes.
* These tools helped us to understand if the model was useful or not.

**7.What did the charts show?**

**7.1 Bar Plots**

* In the **bar plot**, I saw that average age was about the same for students who are depressed and not depressed, no matter how many hours they use social media.

A screenshot of a computer

AI-generated content may be incorrect.

**7.2 Box Plots**

* The **box plot** showed that depressed students may use social media a bit more, but there is a lot of overlap.

A screenshot of a computer

AI-generated content may be incorrect.

**7.3. Box Plot**

* A screenshot of a computer

  AI-generated content may be incorrect.The **box plot** shows that which age group students are most depressed and students who are getting suicidal thoughts.

**8. Conclusions**

**8.1. What improvements would you like to make in future?**

* Try more advanced models like Random Forest.
* Use more features like how much sleep they get, how often they go to counseling, or how they cope with stress level.
* Get more data to make the model stronger.

**8.2. How do you think the solution could be used in real life?**

* Schools could use this kind of system to find students who need help before things get worse.
* Mental health apps could use this to send alerts or tips to students who are at risk.

**8.3. What value do you think the solution will have to the client?**

* It helps schools, counselors, and health professionals support students early.
* It could also be used by researchers who study mental health and behavior.

**8.4. What did you learn through this project?**

* We learned how to clean and prepare data by exploring Glance and Wrangling.
* We learned how to use simple machine learning models to make predictions.
* We learned how to check if a model is working or not using metrics and charts.