our Final Project

May
2025

Project :HealthCare (Depression)

Prepared by :

1)Mariam Mostafa Abdelaziz
2)Arwa Hamdy Mohammdy
3)Eslam Mohamed Abdelfattah
4)Mariam Mostafa Abdelaal

Prepared for : **DEPI**

MILESTONE 1

Deliverables:

• Dataset Exploration Report: A report that summarizes the data's characteristics, distribution of features, and any data quality issues discovered.

Size

```
[71]  df.shape

...   (27898, 27)
```

Feature Types

	Feature Name	Data Type
0	id	Numeric
1	Age	Numeric
2	Academic Pressure	Numeric
3	CGPA	Numeric
4	Study Satisfaction	Numeric
5	Work/Study Hours	Numeric
6	Financial Stress	Numeric
7	Depression	Numeric
8	Social Isolation	Numeric
9	Bullying	Numeric
10	Family Issues	Numeric
11	Uncertain Future	Numeric
12	Social Media Usage	Numeric
13	Drug/Smoking	Numeric
14	Daily Coffee Intake	Numeric
15	PHQ-9	Numeric
16	Cortisol_Level	Numeric
17	Gender	Categorical
18	City	Categorical
19	Sleep Duration	Categorical
20	Dietary Habits	Categorical
21	Degree	Categorical
22	Have you ever had suicidal thoughts ?	Categorical
23	Family History of Mental Illness	Categorical
24	Favorite Color	Categorical
25	Pet Ownership	Categorical
26	Music Genre Preference	Categorical

Statistical Summary

	id	Age	Academic Pressure	CGPA	Study Satisfaction	Work/Study Hours	Financial Stress	Depression	Social Isolation	Bullying	Family Issues	Uncertain Future	Social Media Usage	Drug/Smoking	Daily Coffee Intake	PHQ-9	Cortisol_Level
count	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27898.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000
mean	70442.149421	21.034766	3.141214	7.656104	2.943837	7.156984	3.139867	0.690011	2.995950	0.500090	2.997814	2.996989	3.006559	0.498763	1.991828	6.669008	4.196608
std	40641.175216	2.654828	1.381465	1.470707	1.361148	3.707642	1.437347	0.462497	1.413587	0.500009	1.413819	1.416844	1.414895	0.500007	1.406999	5.148011	2.572537
min	2.000000	16.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	1.000000	0.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000081
25%	35039.000000	19.000000	2.000000	6.290000	2.000000	4.000000	2.000000	0.000000	2.000000	0.000000	2.000000	2.000000	2.000000	0.000000	1.000000	3.000000	2.037070
50%	70684.000000	21.000000	3.000000	7.770000	3.000000	8.000000	3.000000	1.000000	3.000000	1.000000	3.000000	3.000000	3.000000	0.000000	2.000000	6.000000	4.062870
75%	105818.000000	23.000000	4.000000	8.920000	4.000000	10.000000	4.000000	1.000000	4.000000	1.000000	4.000000	4.000000	4.000000	1.000000	3.000000	9.000000	6.098431
max	140699.000000	25.000000	5.000000	10.000000	5.000000	12.000000	5.000000	1.000000	5.000000	1.000000	5.000000	5.000000	5.000000	1.000000	4.000000	20.000000	9.999471

Missing Values

Dealing with missing values by removing

```
id                                      0
Gender                                  0
Age                                     0
City                                    0
Academic Pressure                       0
CGPA                                    0
Study Satisfaction                      0
Sleep Duration                          0
Dietary Habits                          0
Degree                                  0
Have you ever had suicidal thoughts ?   0
Work/Study Hours                        0
Financial Stress                        3
Family History of Mental Illness        0
Depression                              0
Social Isolation                        0
Bullying                                0
Family Issues                           0
Uncertain Future                        0
Social Media Usage                      0
Drug/Smoking                            0
Favorite Color                          0
Pet Ownership                           0
Daily Coffee Intake                     0
Music Genre Preference                  0
PHQ-9                                   0
Cortisol_Level                          0
dtype: int64
```

Outliers

Dealing with outliers by Keeping them because of Their Low Percentage

	Feature	Outlier Percentage
0	id	0.00
1	Age	0.00
2	Academic Pressure	0.00
3	CGPA	0.03
4	Study Satisfaction	0.00
5	Work/Study Hours	0.00
6	Financial Stress	0.00
7	Social Isolation	0.00
8	Bullying	0.00
9	Family Issues	0.00
10	Uncertain Future	0.00
11	Social Media Usage	0.00
12	Drug/Smoking	0.00
13	Daily Coffee Intake	0.00
14	PHQ-9	3.66
15	Cortisol_Level	0.00
16	Depression	0.00

Duplicates

```
[10]  df.duplicated().sum()
      ✓ 0.0s

      0
```

Notes

An imbalance in class distribution was observed in the target column, where one class dominates the others. This may affect the performance of classification models and should be addressed during modeling (e.g., using smote techniques).

```
df["Depression"].value_counts(normalize=True) * 100
```

[7]

```
Depression
1    69.004947
0    30.995053
Name: proportion, dtype: float64
```

---

column such as ID do not carry useful information for modeling purposes. This column has been identified as non-informative and will be excluded from further analysis.

```
correlation_matrix = df.corr()
target_correlation = correlation_matrix['Depression']
print(target_correlation)
```

[33] ✓ 0.0s

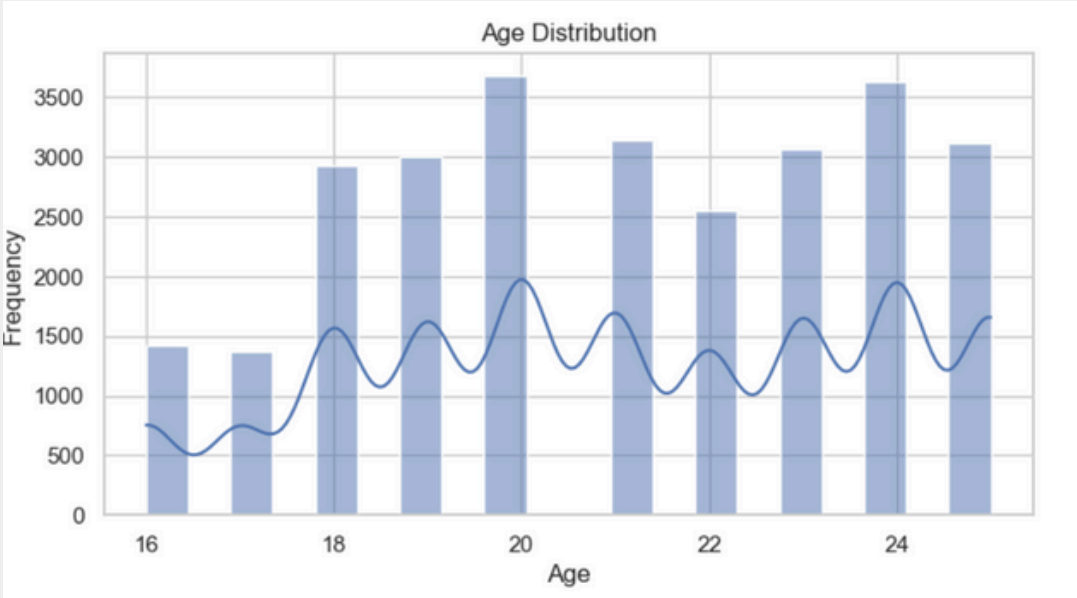
```
id                                      0.001596
```

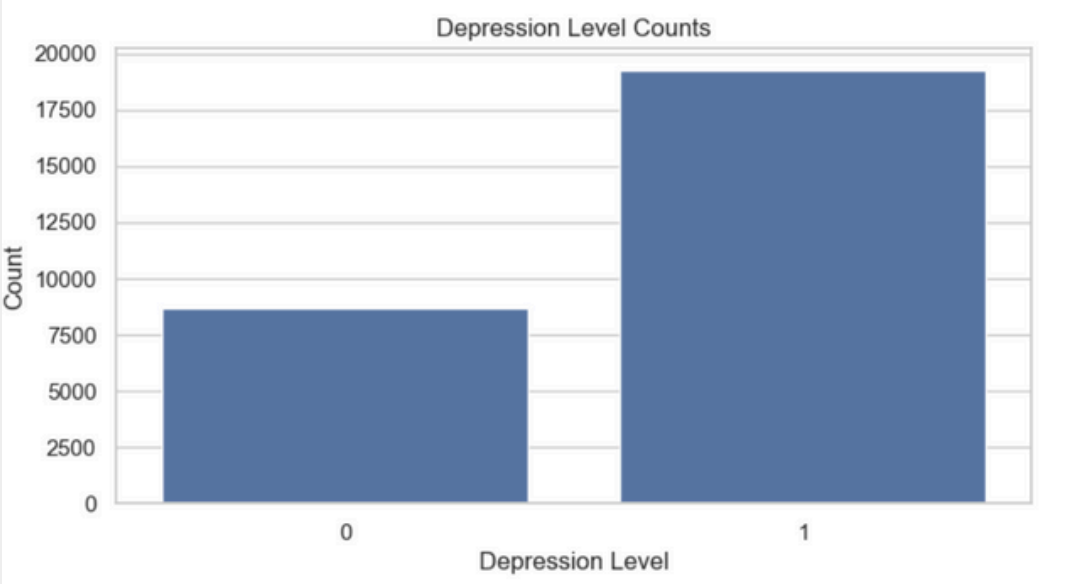
No timestamp or date-related column was found in the dataset, which limits time-based analysis or trend identification.

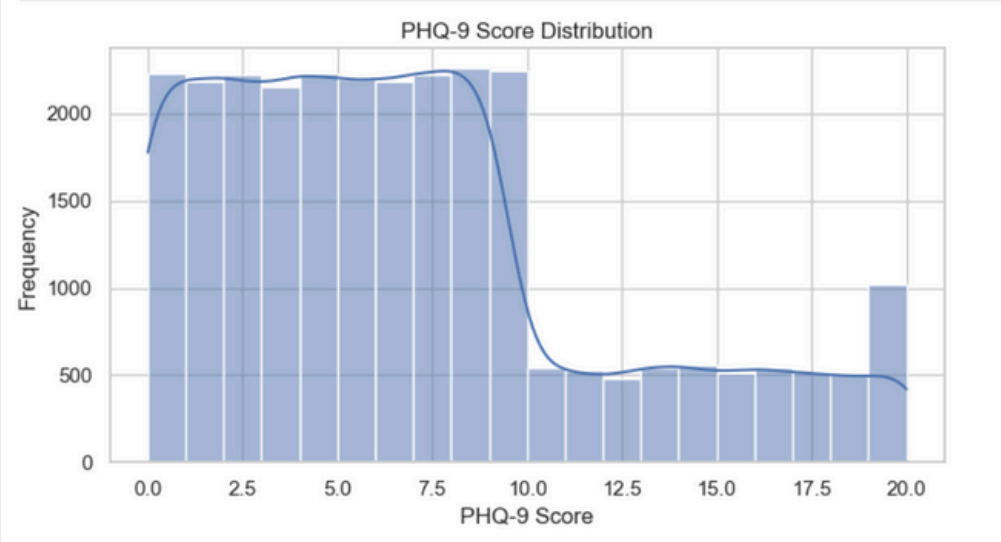
MILESTONE 2

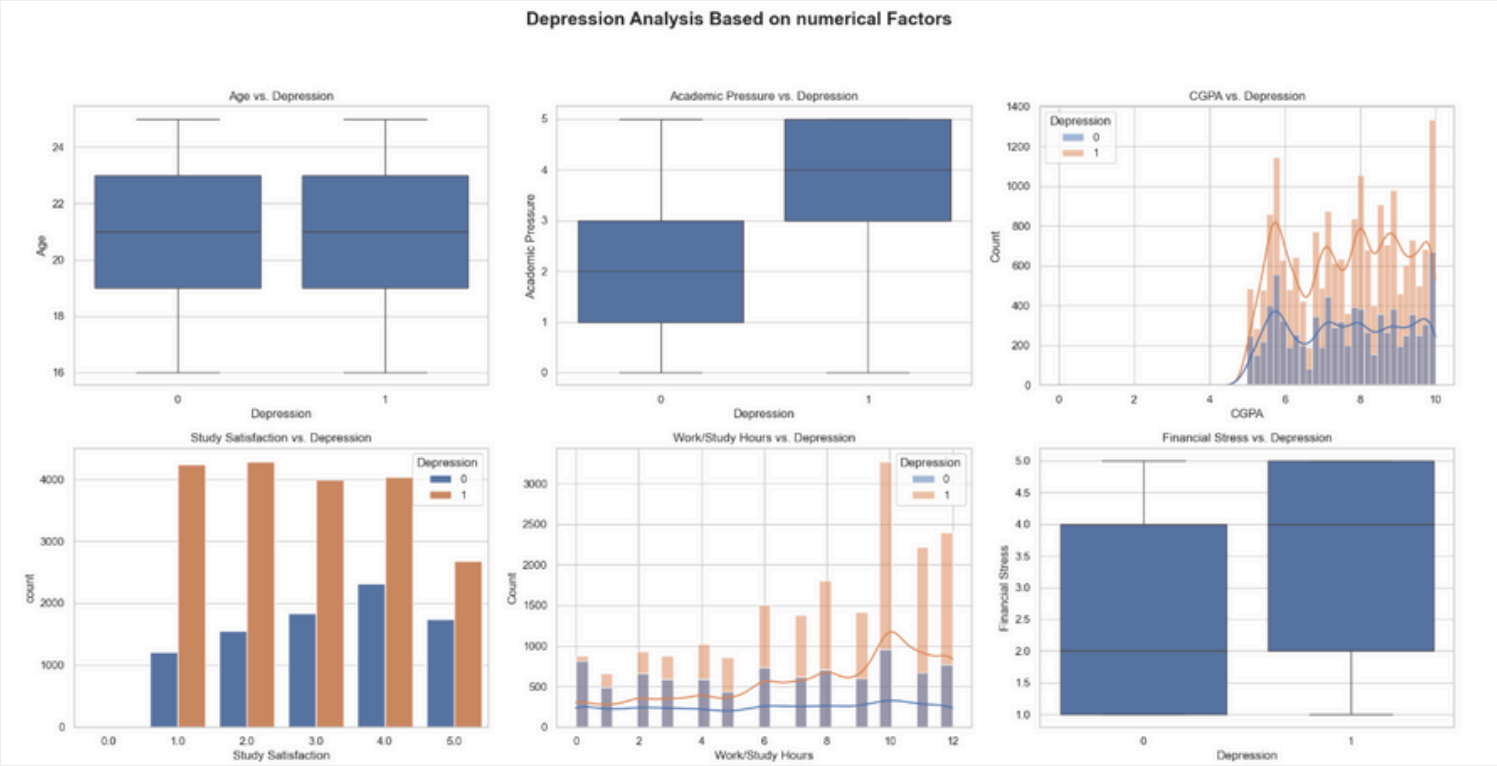
Deliverables:

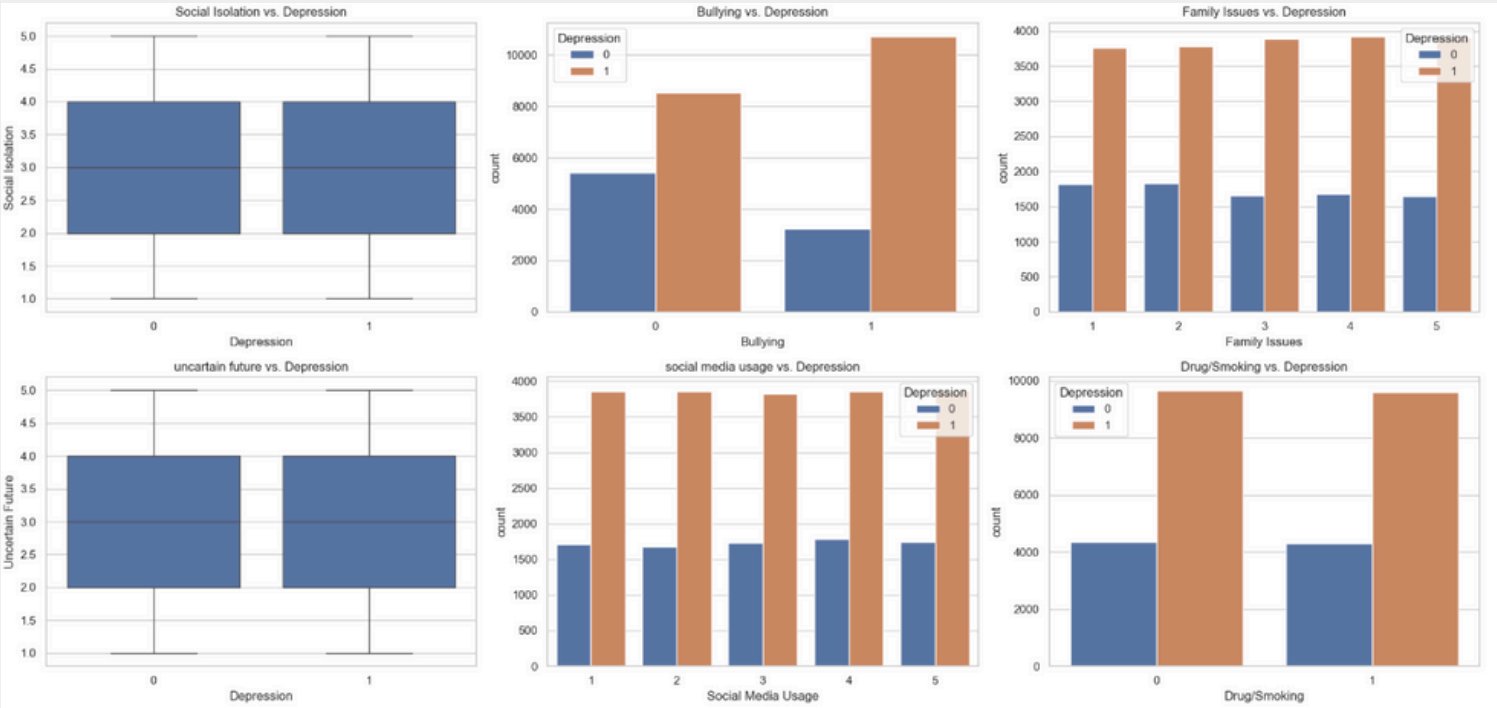
• Cleaned Dataset and Analysis Report: A detailed report outlining the data cleaning steps, analysis results, and insights gained from health metrics.

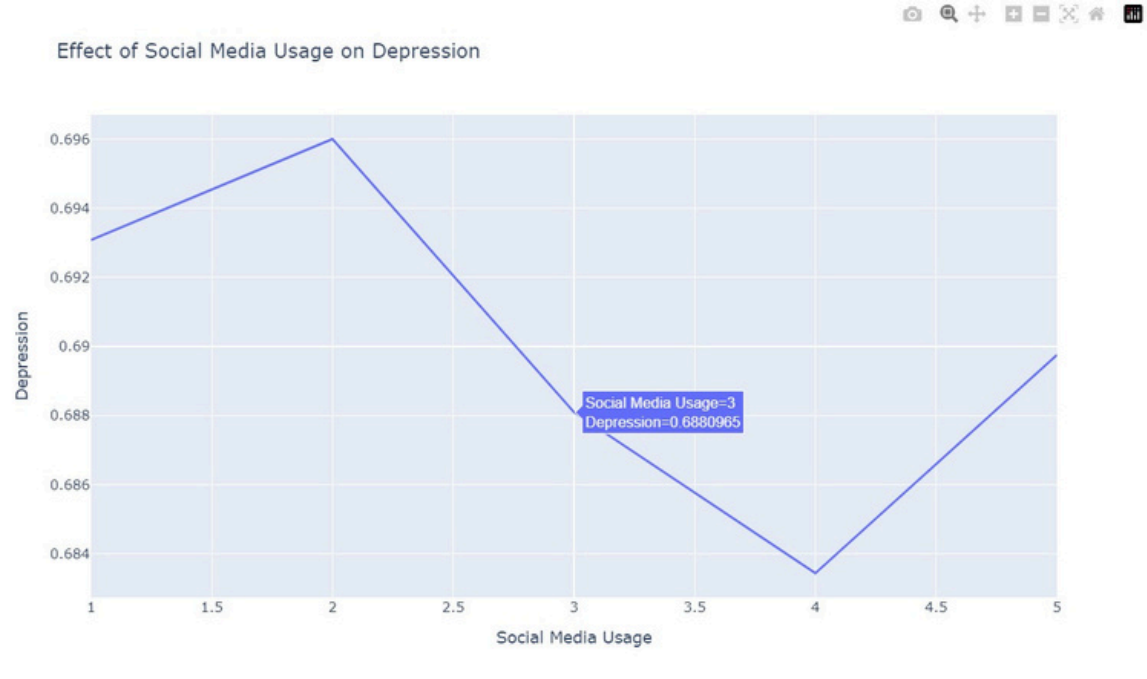


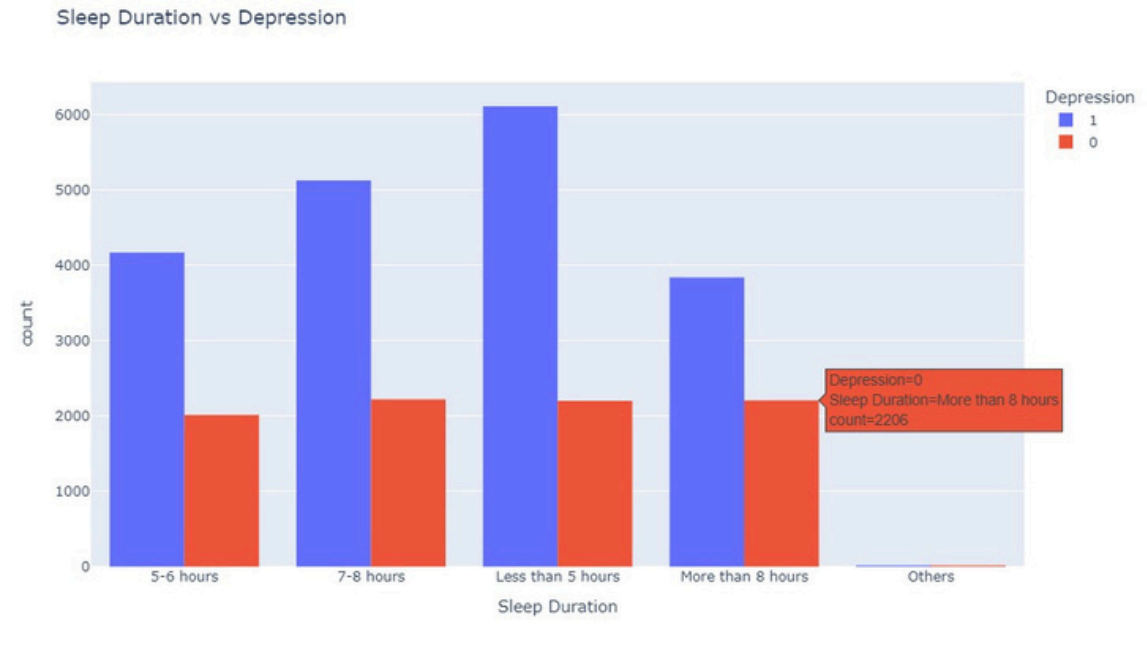


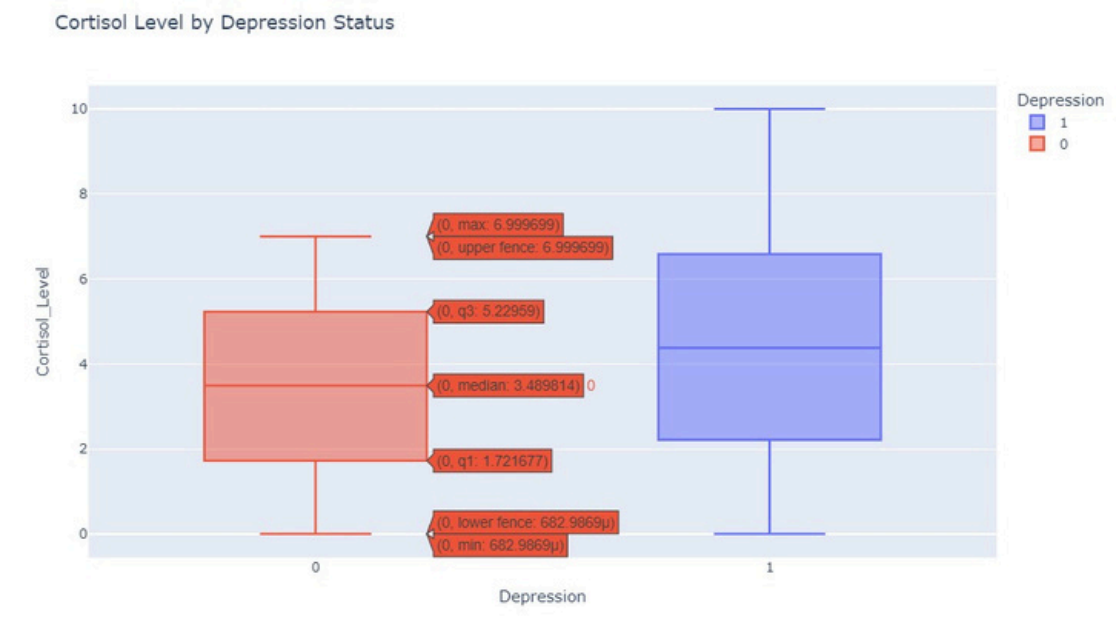



Depression Analysis Based on numerical Factors
Age vs. Depression
Age
Depression
Academic Pressure vs. Depression
Academic Pressure
Depression
CGPA vs. Depression
Count
CGPA
Depression
0
1
Study Satisfaction vs. Depression
count
Study Satisfaction
Work/Study Hours vs. Depression
Count
Work/Study Hours
Financial Stress vs. Depression
Financial Stress
Depression


Social Isolation vs. Depression
Social Isolation
Depression
Bullying vs. Depression
count
Bullying
Depression
0
1
Family Issues vs. Depression
count
Family Issues
uncartain future vs. Depression
Uncertain Future
Depression
social media usage vs. Depression
count
Social Media Usage
Drug/Smoking vs. Depression
count
Drug/Smoking


Effect of Social Media Usage on Depression
Depression
0.696
0.694
0.692
0.69
0.688
0.686
0.684
1
1.5
2
2.5
3
3.5
4
4.5
5
Social Media Usage
Social Media Usage=3
Depression=0.6880965


Sleep Duration vs Depression
count
6000
5000
4000
3000
2000
1000
0
5-6 hours
7-8 hours
Less than 5 hours
More than 8 hours
Others
Sleep Duration
Depression
1
0
Depression=0
Sleep Duration=More than 8 hours
count=2206


Cortisol Level by Depression Status
Cortisol_Level
10
8
6
4
2
0
0
1
Depression
Depression
1
0
(0, max: 6.999699)
(0, upper fence: 6.999699)
(0, q3: 5.22959)
(0, median: 3.489814)
(0, q1: 1.721677)
(0, lower fence: 682.9869µ)
(0, min: 682.9869µ)

Correlation Heatmap

	Gender	Age	City	Academic Pressure	CGPA	Study Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Family History of Mental Illness	Social Isolation	Bullying	Family Issues	Uncertain Future	Social Media Usage	Drug/Smoking	Favorite Color	Pet Ownership	Daily Coffee Intake	Music Genre Preference	PHQ-9	Cortisol_Level	Depression
Gender	1	0.0026	-0.022	-0.022	0.036	-0.016	0.012	0.063	0.011	-0.0013	0.013	-0.0055	-0.016	0.0063	-0.0019	0.0027	-0.005	-0.004	0.0031	-0.005	0.006	0.0087	-0.0074	0.0091	-0.0037	-0.0013
Age	0.0026	1	-0.015	-0.029	0.0099	-0.033	-0.003	0.0045	-0.087	-0.00074	-0.0048	0.0082	0.0015	0.0088	0.00039	-0.0013	-0.0018	0.0024	0.00018	-0.0031	0.0011	-0.005	-0.0059	-0.011	0.0042	-0.001
City	-0.022	-0.015	1	-0.021	-0.029	0.0071	-0.011	-0.011	0.0042	-0.0054	-0.0027	-0.004	0.0043	0.0024	-0.0048	0.0019	0.0041	-0.006	-0.0018	-0.0022	-0.0035	0.0057	-0.0036	-0.0011	-0.0054	-0.021
Academic Pressure	-0.022	-0.029	-0.021	1	-0.022	-0.11	-0.0064	0.091	-0.0028	0.26	0.096	0.15	0.03	0.0023	-0.00061	-0.0072	-0.0046	0.011	0.0044	0.00096	0.0035	-0.0062	-0.016	0.11	0.06	0.39
CGPA	0.036	0.0099	-0.029	-0.022	1	-0.044	-0.021	0.0013	-0.027	0.0085	0.0026	0.0059	-0.0038	-0.0014	0.0016	-0.01	-0.004	0.0032	0.0038	-0.0066	0.00049	-0.0096	0.0023	4.4e-05	0.0044	0.011
Study Satisfaction	-0.016	-0.033	0.0071	-0.11	-0.044	1	0.0037	-0.017	0.012	-0.084	-0.036	-0.065	-0.0039	-0.0048	-0.0015	0.0087	0.0016	-0.0066	-0.0018	-0.0037	-0.0088	0.0042	0.0047	-0.036	-0.033	-0.13
Sleep Duration	0.012	-0.003	-0.011	-0.0064	-0.021	0.0037	1	0.011	0.0041	-0.021	-0.039	-0.0021	0.00015	-0.0092	-0.005	-0.0031	-0.0039	-0.0014	0.0083	-0.0042	-0.005	-0.0017	-0.0073	-0.014	-0.0043	-0.015
Dietary Habits	0.063	0.0045	-0.011	0.091	0.0013	-0.017	0.011	1	-0.00045	0.11	0.03	0.09	0.0058	0.0033	-0.0014	-0.00017	-0.0018	0.01	0.01	0.0073	-0.0029	0.018	-0.012	0.038	0.036	0.17
Degree	0.011	-0.087	0.0042	-0.0028	-0.027	0.012	0.0041	-0.00045	1	0.0073	0.0048	0.0071	0.005	-0.014	-0.0021	-0.0048	-0.00093	-0.00058	0.0088	0.0057	0.0051	-0.0048	0.0088	0.0021	0.0071	-0.0016
Have you ever had suicidal thoughts ?	-0.0013	-0.00074	-0.0054	0.26	0.0085	-0.084	-0.021	0.11	0.0073	1	0.12	0.21	0.026	-0.003	-0.0038	-0.00033	-0.0087	0.001	0.0058	-0.0055	-0.0008	-0.0047	-0.011	0.12	0.074	0.44
Work/Study Hours	0.013	-0.0048	-0.0027	0.096	0.0026	-0.036	-0.039	0.03	0.0048	0.12	1	0.075	0.017	-0.0046	0.0091	0.0026	-0.0043	0.0021	0.0002	-0.0095	-0.0071	0.00048	0.0066	0.045	0.034	0.16
Financial Stress	-0.0055	0.0082	-0.004	0.15	0.0059	-0.065	-0.0021	0.09	0.0071	0.21	0.075	1	0.0085	-0.013	-0.0049	0.0016	-0.015	0.006	0.0057	0.0059	0.0053	-0.0091	-0.0053	0.075	0.056	0.29
Family History of Mental Illness	-0.016	0.0015	0.0043	0.03	-0.0038	-0.0039	0.00015	0.0058	0.005	0.026	0.017	0.0085	1	0.0023	-0.0052	-0.00012	-0.011	0.00096	-0.0035	-0.0013	0.0058	-0.0069	0.0039	0.01	0.0066	0.044
Social Isolation	0.0063	0.0088	0.0024	0.0023	-0.0014	-0.0048	-0.0092	0.0033	-0.014	-0.003	-0.0046	-0.013	0.0023	1	-0.0054	-0.0022	0.00087	0.0051	0.0029	-0.011	-0.006	0.0056	-0.0098	0.0017	-0.00092	0.0012
Bullying	-0.0019	0.00039	-0.0048	-0.00061	0.0016	-0.0015	-0.005	-0.0014	-0.0021	-0.0038	0.0091	-0.0049	-0.0052	-0.0054	1	-0.0026	0.0023	0.001	-0.015	-0.016	0.0005	0.012	-0.00012	0.044	0.033	0.17
Family Issues	0.0027	-0.0013	0.0019	-0.0072	-0.01	0.0087	-0.0031	-0.00017	-0.0048	-0.00033	0.0026	0.0016	-0.00012	-0.0022	-0.0026	1	0.0081	0.0091	0.0021	-0.0012	-0.0063	-0.0043	0.0048	0.0086	0.00014	0.026
Uncertain Future	-0.005	-0.0018	0.0041	-0.0046	-0.004	0.0016	-0.0039	-0.0018	-0.00093	-0.0087	-0.0043	-0.015	-0.011	0.00087	0.0023	0.0081	1	0.0051	9.6e-05	0.014	0.0036	0.0051	0.0017	0.003	0.006	-0.01
Social Media Usage	-0.004	0.0024	-0.006	0.011	0.0032	-0.0066	-0.0014	0.01	-0.00058	0.001	0.0021	0.006	0.00096	0.0051	0.001	0.0091	0.0051	1	-0.0042	0.0018	0.0029	-0.0039	-0.0064	0.0012	0.0043	-0.0059
Drug/Smoking	0.0031	0.00018	-0.0018	0.0044	0.0038	-0.0018	0.0083	0.01	0.0088	0.0058	0.0002	0.0057	-0.0035	0.0029	-0.015	0.0021	9.6e-05	-0.0042	1	0.0041	-0.01	0.0025	-8.3e-06	-0.0021	0.0013	0.00057
Favorite Color	-0.005	-0.0031	-0.0022	0.00096	-0.0066	-0.0037	-0.0042	0.0073	0.0057	-0.0055	-0.0095	0.0059	-0.0013	-0.011	-0.016	-0.0012	0.014	0.0018	0.0041	1	-0.00069	0.0084	-0.0048	-0.004	-0.0031	-0.0044
Pet Ownership	0.006	0.0011	-0.0035	0.0035	0.00049	-0.0088	-0.005	-0.0029	0.0051	-0.0008	-0.0071	0.0053	0.0058	-0.006	0.0005	-0.0063	0.0036	0.0029	-0.01	-0.00069	1	0.0039	-0.0074	-0.00048	-0.0054	0.004
Daily Coffee Intake	0.0087	-0.005	0.0057	-0.0062	-0.0096	0.0042	-0.0017	0.018	-0.0048	-0.0047	0.00048	-0.0091	-0.0069	0.0056	0.012	-0.0043	0.0051	-0.0039	0.0025	0.0084	0.0039	1	0.0044	-0.0014	0.0068	-0.0017
Music Genre Preference	-0.0074	-0.0059	-0.0036	-0.016	0.0023	0.0047	-0.0073	-0.012	0.0088	-0.011	0.0066	-0.0053	0.0039	-0.0098	-0.00012	0.0048	0.0017	-0.0064	-8.3e-06	-0.0048	-0.0074	0.0044	1	0.00071	0.00099	-0.0098
PHQ-9	0.0091	-0.011	-0.0011	0.11	4.4e-05	-0.036	-0.014	0.038	0.0021	0.12	0.045	0.075	0.01	0.0017	0.044	0.0086	0.003	0.0012	-0.0021	-0.004	-0.00048	-0.0014	0.00071	1	0.056	0.28
Cortisol_Level	-0.0037	0.0042	-0.0054	0.06	0.0044	-0.033	-0.0043	0.036	0.0071	0.074	0.034	0.056	0.0066	-0.00092	0.033	0.00014	0.006	0.0043	0.0013	-0.0031	-0.0054	0.0068	0.00099	0.056	1	0.18
Depression	-0.0013	-0.001	-0.021	0.39	0.011	-0.13	-0.015	0.17	-0.0016	0.44	0.16	0.29	0.044	0.0012	0.17	0.026	-0.01	-0.0059	0.00057	-0.0044	0.004	-0.0017	-0.0098	0.28	0.18	1

MILESTONE 3

Deliverables:

• Predictive Model Performance Report: A detailed report summarizing the performance of various models, evaluation metrics, and the final model selection.

Types of models evaluated:

1)CatBoost classifier
2)XGBoost classifier
3)Artificial Neural Networks (ANN)
4)Logistic Regression
5)LightGBM Classifier
6)Random Forest
7) KNN
8)Naive Bayes
9)Decision Tree
10)SVM
11)Gradient Boosting

Evaluation metrics used:

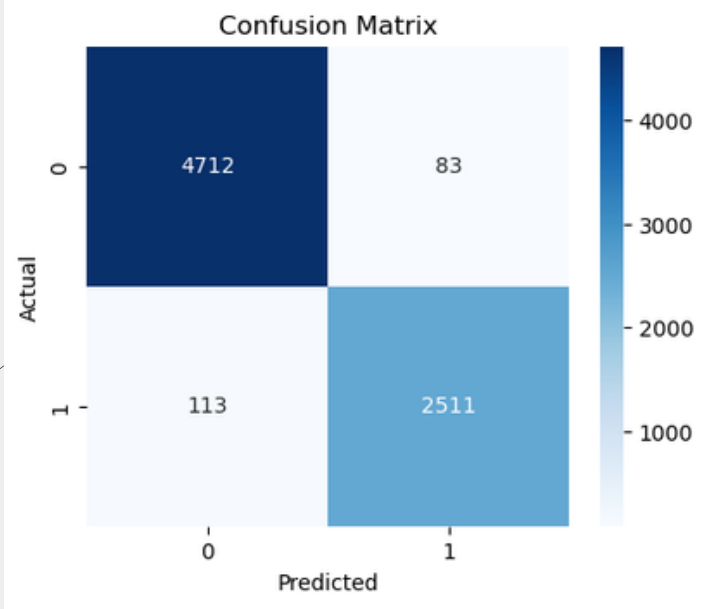
Depending on the problem type (classification , we use appropriate metrics such as:

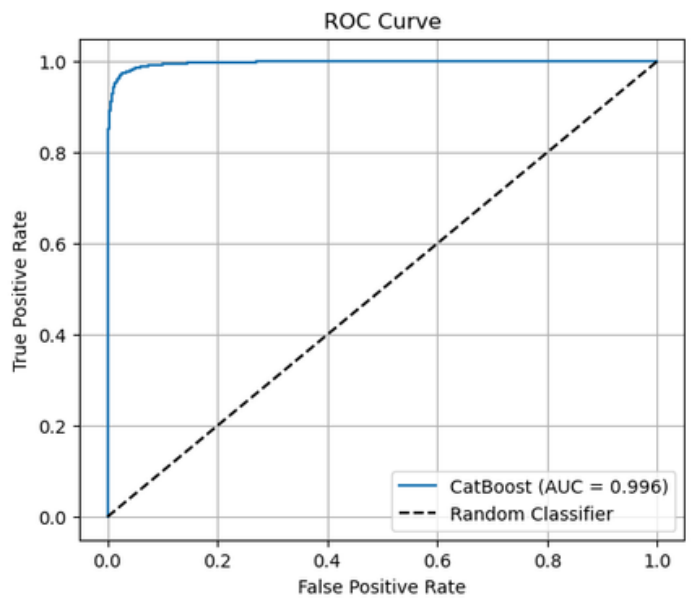
Accuracy
Precision
Recall
F1-score
ROC-AUC

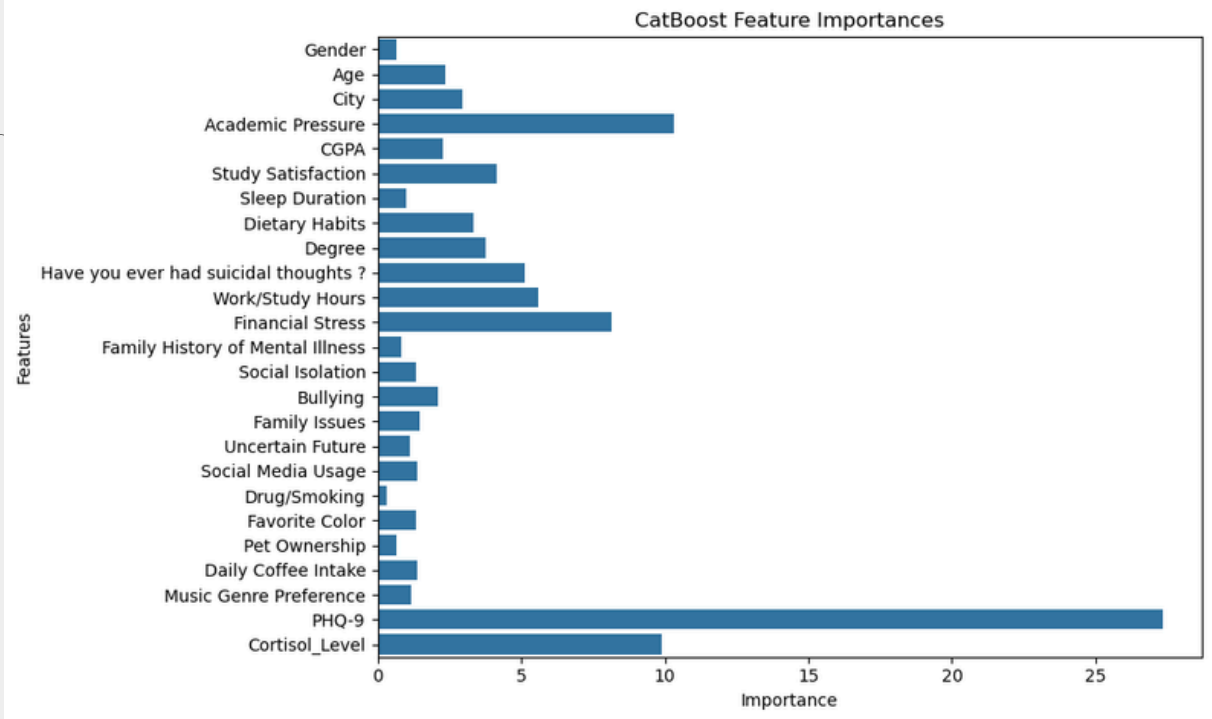
First Model : CatBoost

```
Complete Catboost Training Accuracy: 0.9981514643868061
Complete catboost Test Accuracy: 0.9735813451947701
```

```
Cross Validation Scores: [0.95066721 0.95268904 0.97978164 0.98483623 0.98443186]
Mean CV Accuracy: 0.9704811969268097
Standard Deviation: 0.015468299413336668
```



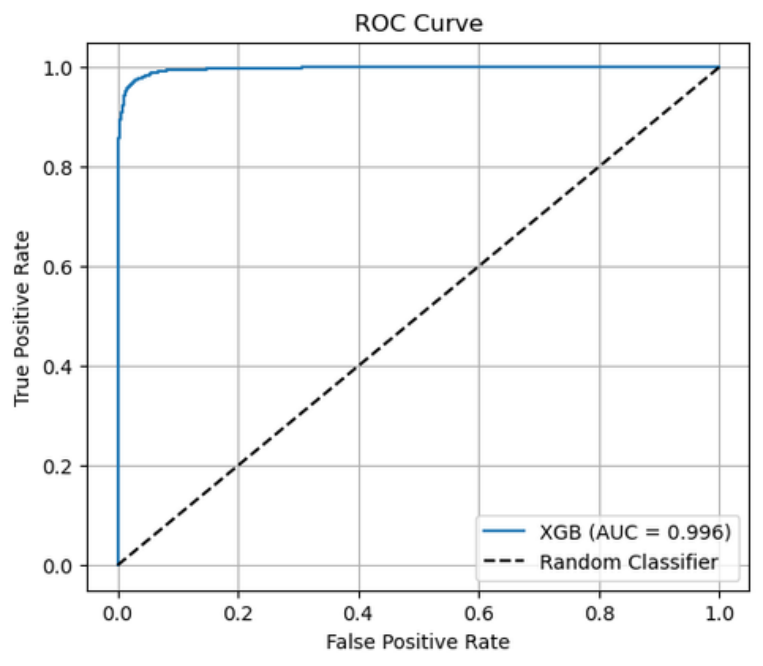


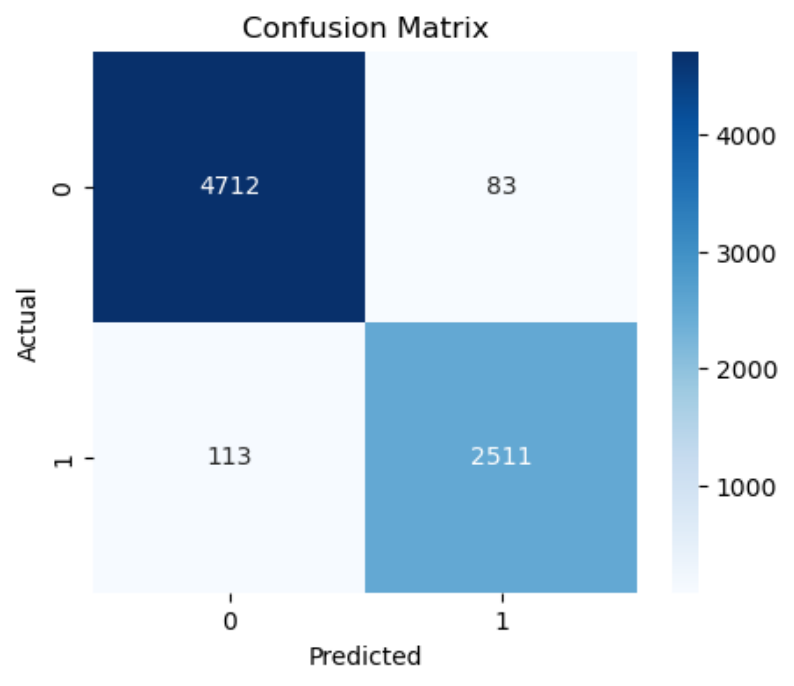


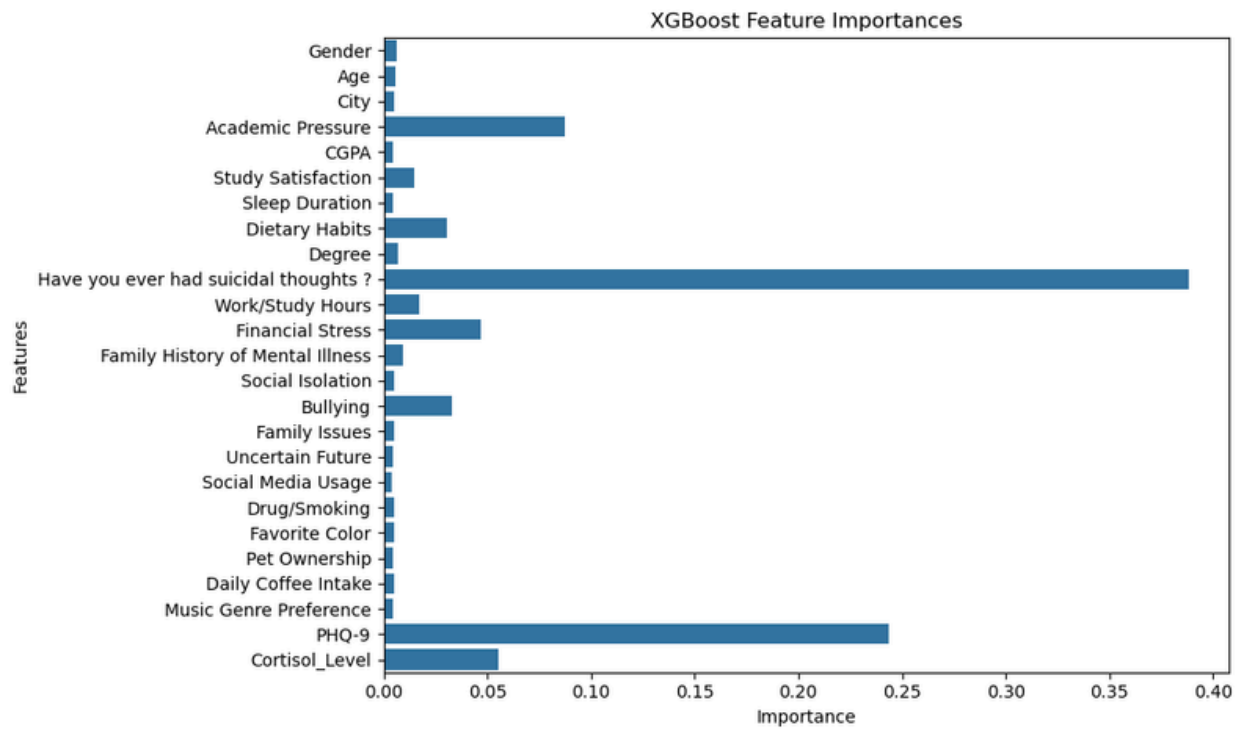
Second Model : XGBBoost

```
XGB Training Accuracy: 0.9826122118883946
XGB Test Accuracy: 0.973176978029384
XGBoost Accuracy: 0.973176978029384
```

```
Cross Validation Scores: [0.94035584 0.94985847 0.9828144  0.98544278 0.98584715]
Mean CV Accuracy: 0.9688637282652648
Standard Deviation: 0.01965615439543187
```



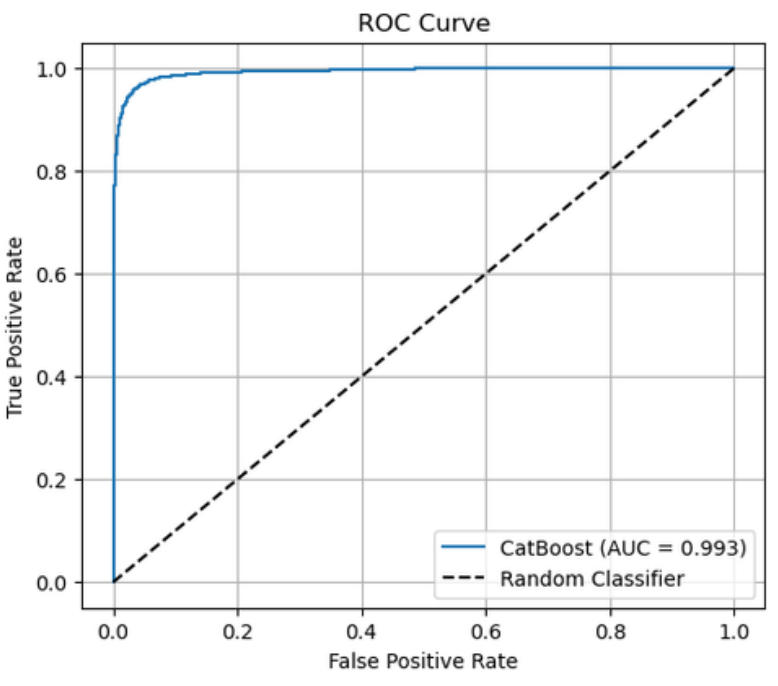


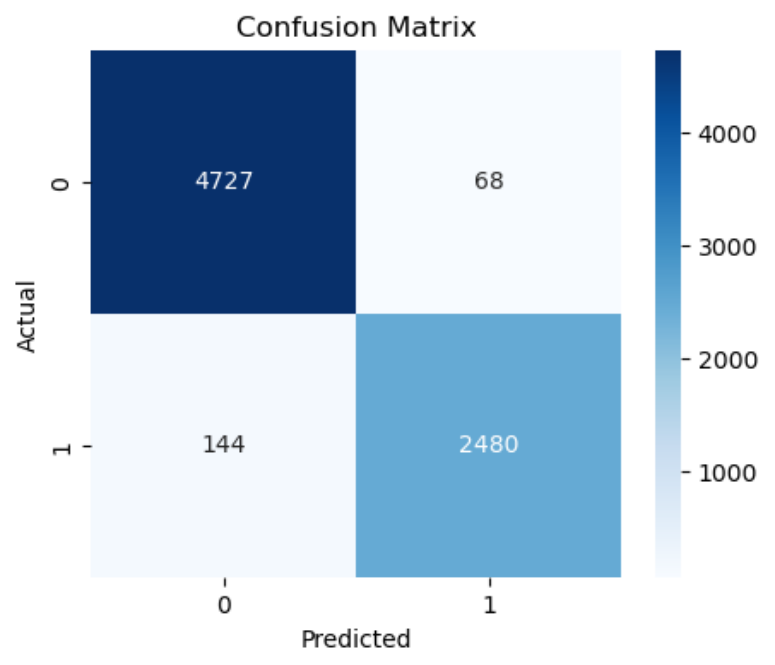


Third Model : Neural Network

```
ANN Training Accuracy: 0.9755646698630929
ANN Test Accuracy: 0.9637417441703734
```

```
Cross Validation Scores: [0.94439951 0.94096239 0.96825718 0.9617873  0.9741205 ]
Mean CV Accuracy: 0.9579053780832997
Standard Deviation: 0.013073885229566115
```



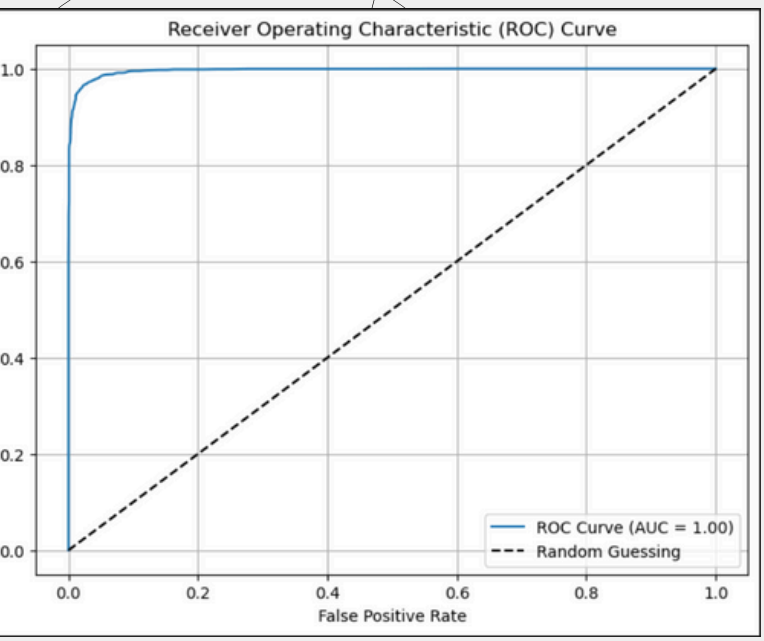


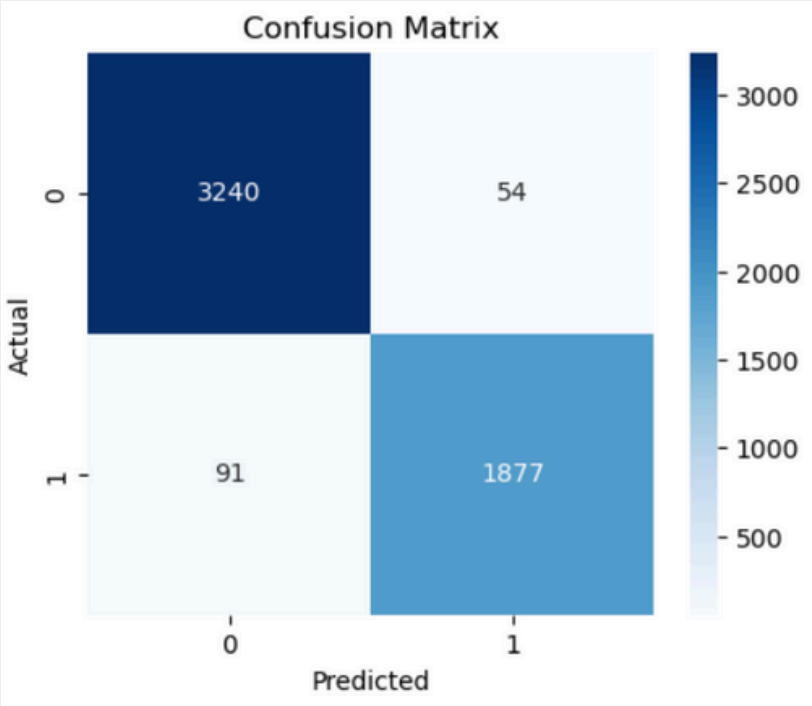
DOESN'T SUPPORT FEATURE IMPORTANCE

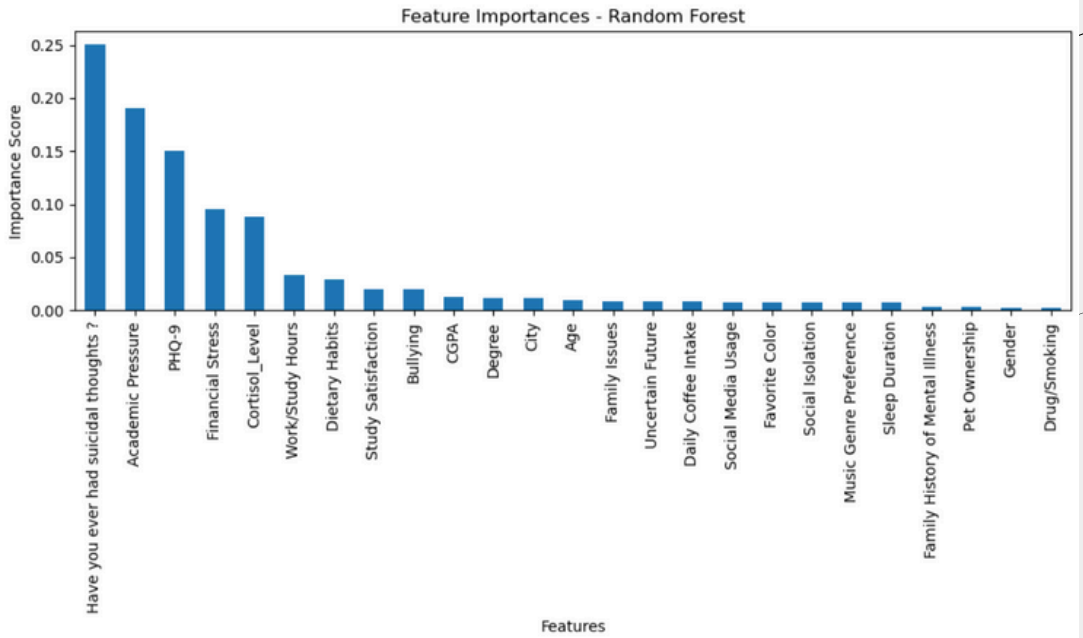
Fourth Model : Random Forest

```
Accuracy: 0.9724439376662866

Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.98      0.98      3294
           1       0.97      0.95      0.96      1968

    accuracy                           0.97      5262
   macro avg       0.97      0.97      0.97      5262
weighted avg       0.97      0.97      0.97      5262
```

```
Cross-validation scores: [0.97244394 0.97339415 0.96939745 0.97129823 0.96996769]
Average score: 0.9713002912722153
```



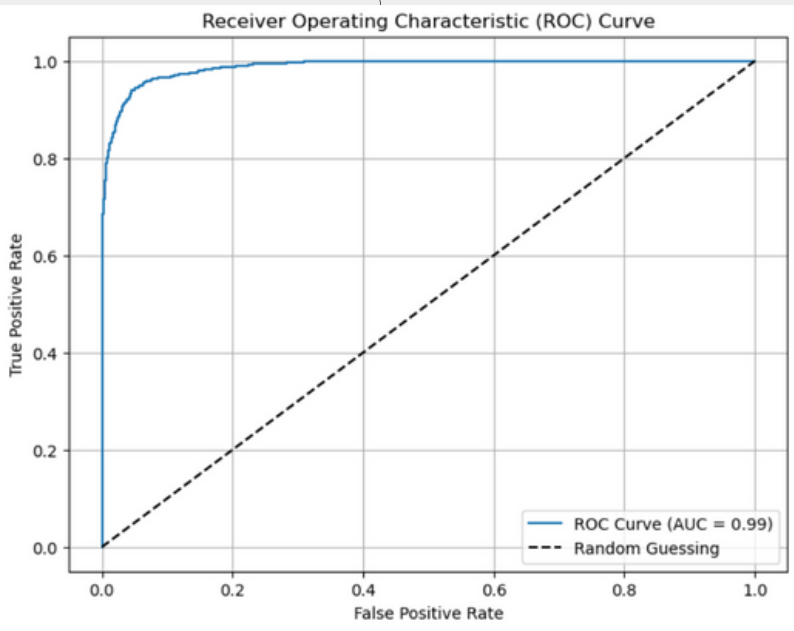


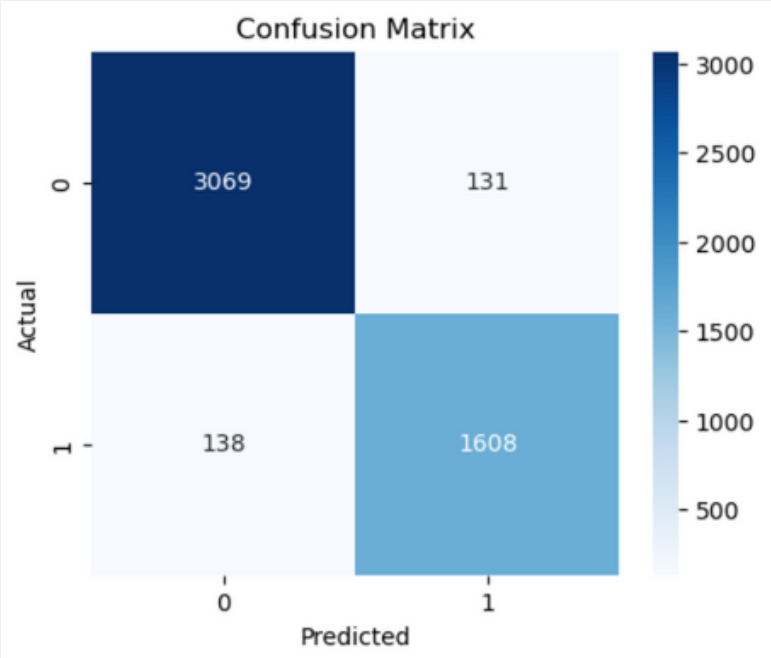


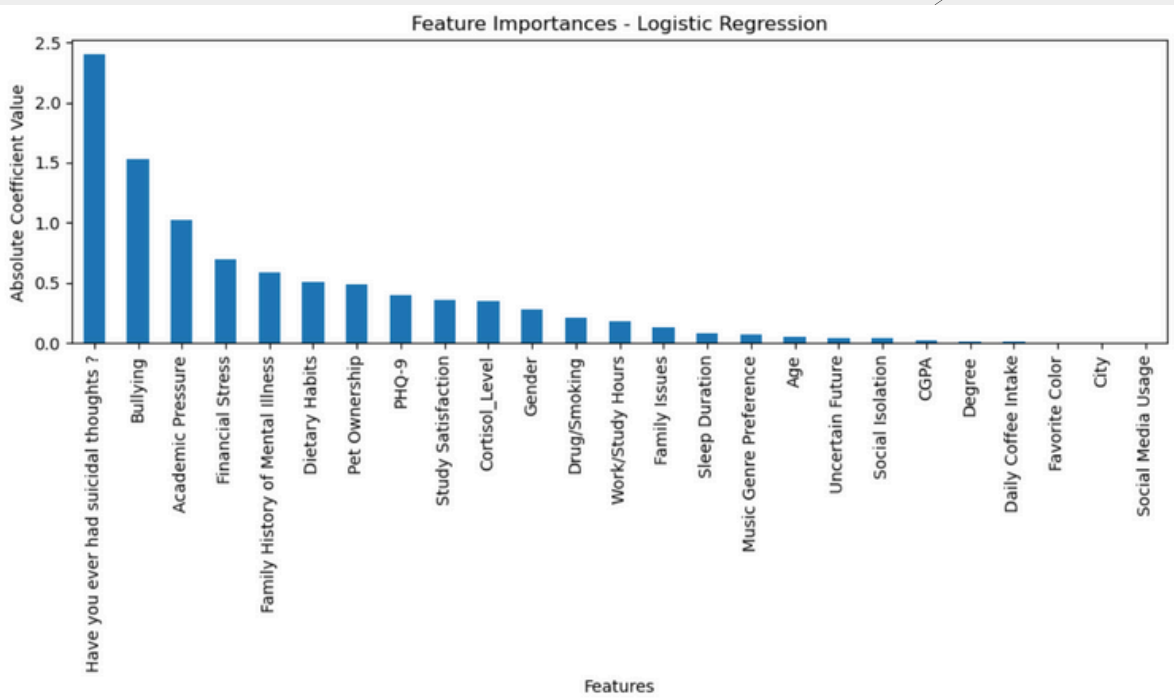
Fifth Model : Logistic Regression

```
Complete LogisticRegression Training Accuracy: 0.9401536595228468
Complete LogisticRegression Test Accuracy: 0.94561261625556
Confusion Matrix:
 [[3069  131]
 [ 138 1608]]

Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.96      0.96      3200
           1       0.92      0.92      0.92      1746

    accuracy                           0.95      4946
   macro avg       0.94      0.94      0.94      4946
weighted avg       0.95      0.95      0.95      4946
```

```
Cross-validation scores: [0.9458148  0.93974929 0.94419733 0.93752527 0.93813182]
Average score: 0.9410837040032349
```



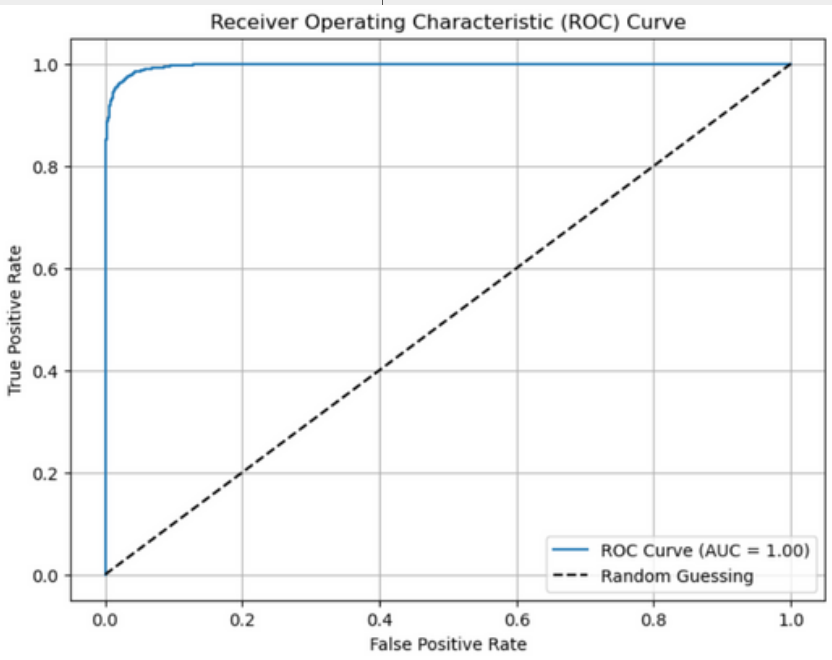


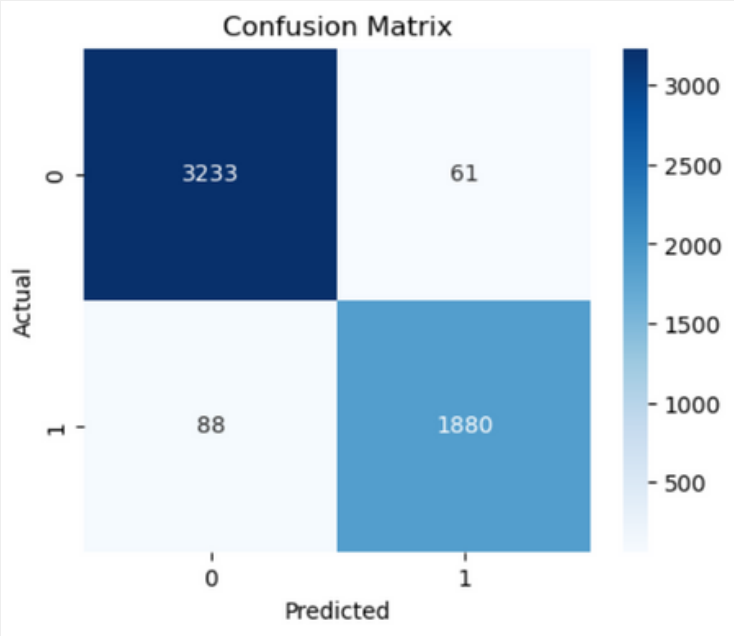


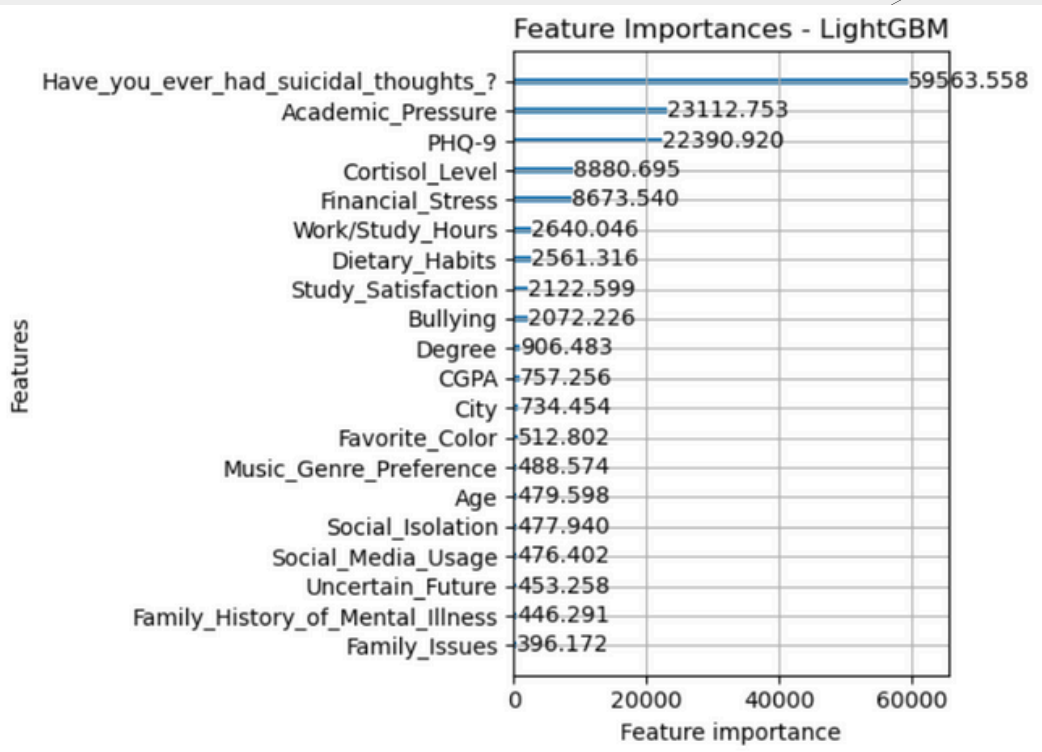
Sixth Model : LightGBM Classifier

```
Accuracy: 0.9716837704294945

Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.98      0.98      3294
           1       0.97      0.96      0.96      1968

    accuracy                           0.97      5262
   macro avg       0.97      0.97      0.97      5262
weighted avg       0.97      0.97      0.97      5262
```

```
Cross-validation scores: [0.97168377 0.97149373 0.970728   0.97148831 0.96825699]
Average score: 0.9707301586200703
```

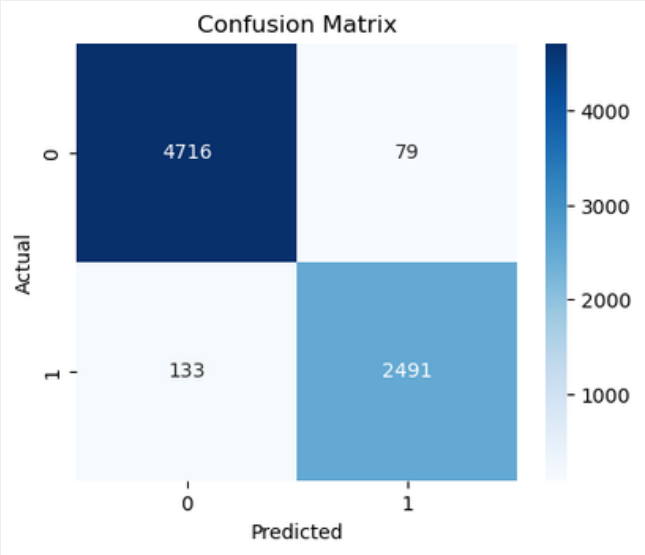


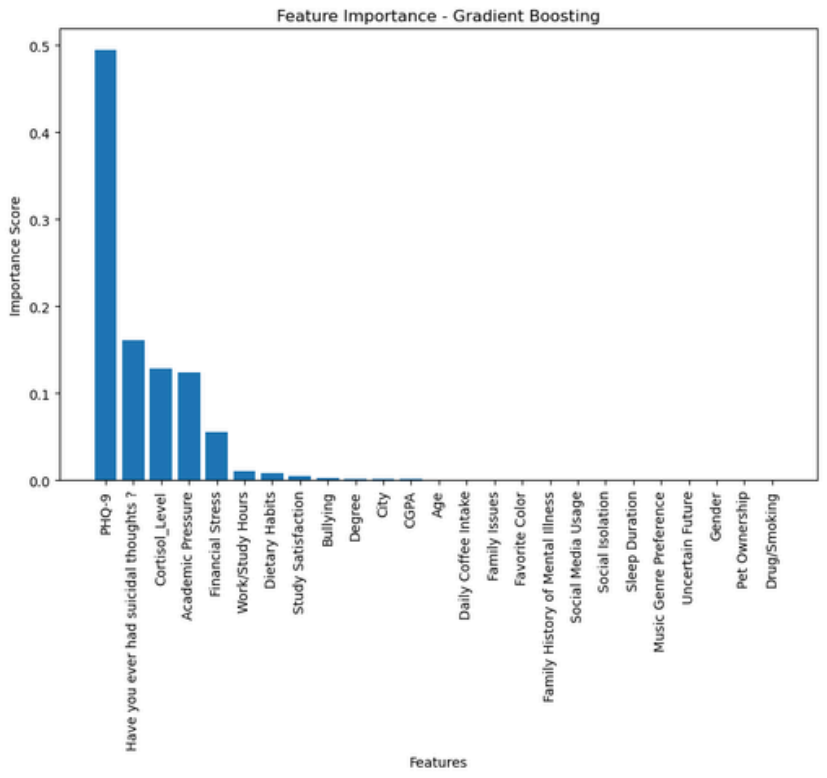




Seventh Model : Gradient boosting

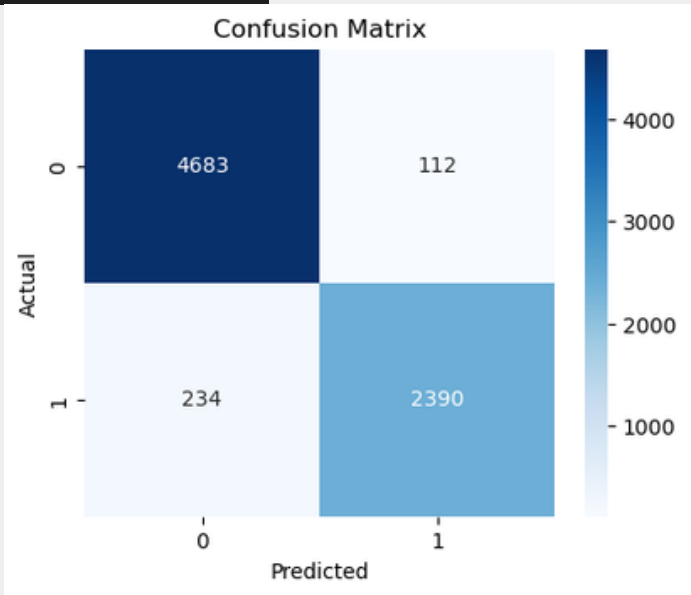
```
...  Gradient Boosting Test Accuracy: 0.9714247203127107
     Gradient Boosting Training Accuracy: 0.9753
     Gradient Boosting Test Accuracy: 0.9714
     Confusion Matrix:
      [[4716   79]
      [ 133 2491]]
     Classification Report:
                   precision    recall  f1-score   support

                0       0.97      0.98      0.98      4795
                1       0.97      0.95      0.96      2624

         accuracy                           0.97      7419
        macro avg       0.97      0.97      0.97      7419
     weighted avg       0.97      0.97      0.97      7419
```

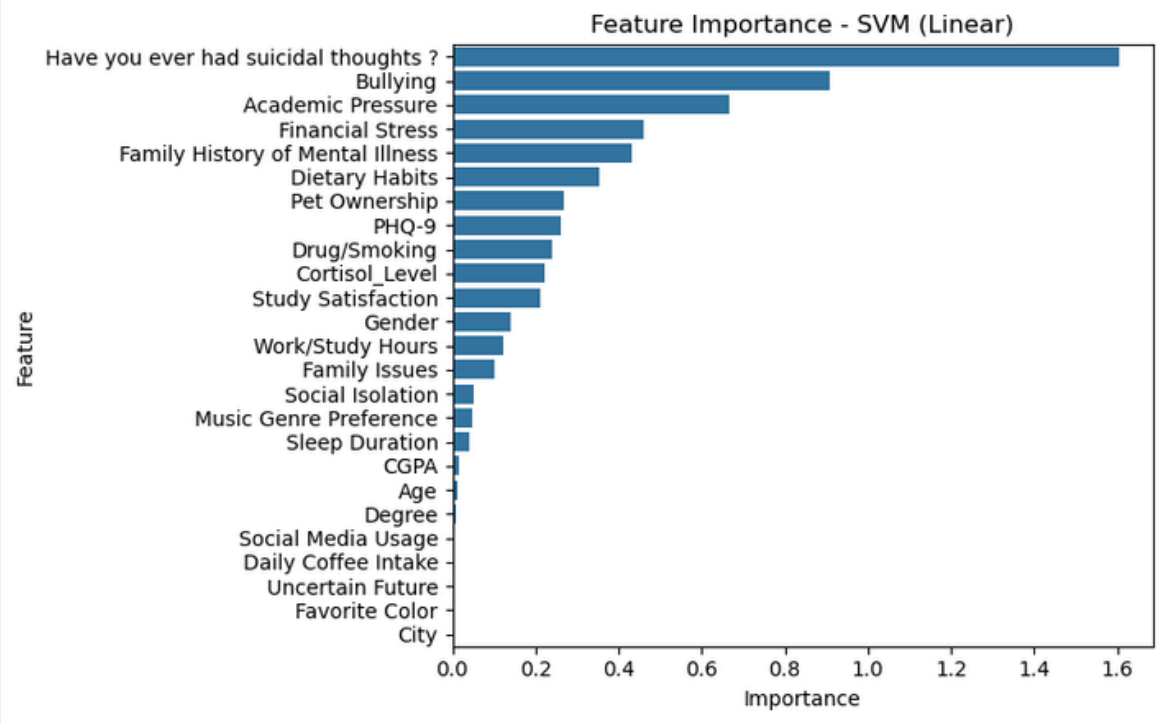




Eighth Model : SVM

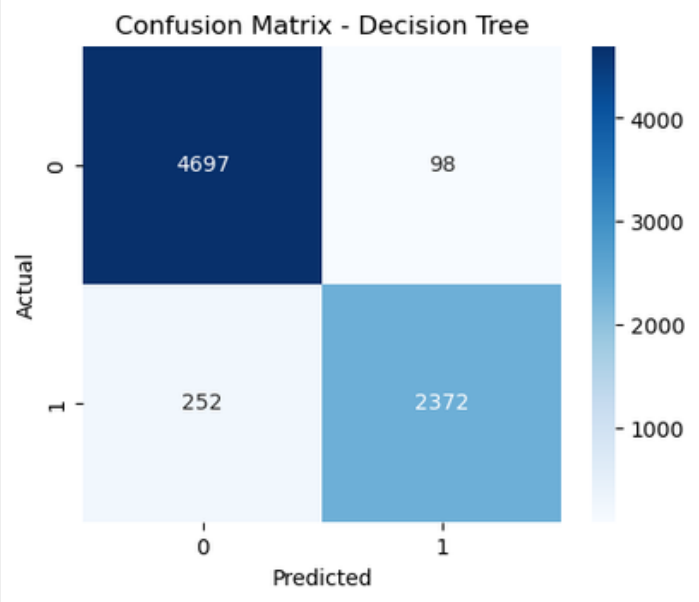
```
SVM Training Accuracy: 0.9551730113800474
SVM Test Accuracy: 0.9533629869254616

Confusion Matrix:
 [[4683  112]
 [ 234 2390]]

Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.98      0.96      4795
           1       0.96      0.91      0.93      2624

    accuracy                           0.95      7419
   macro avg       0.95      0.94      0.95      7419
weighted avg       0.95      0.95      0.95      7419
```

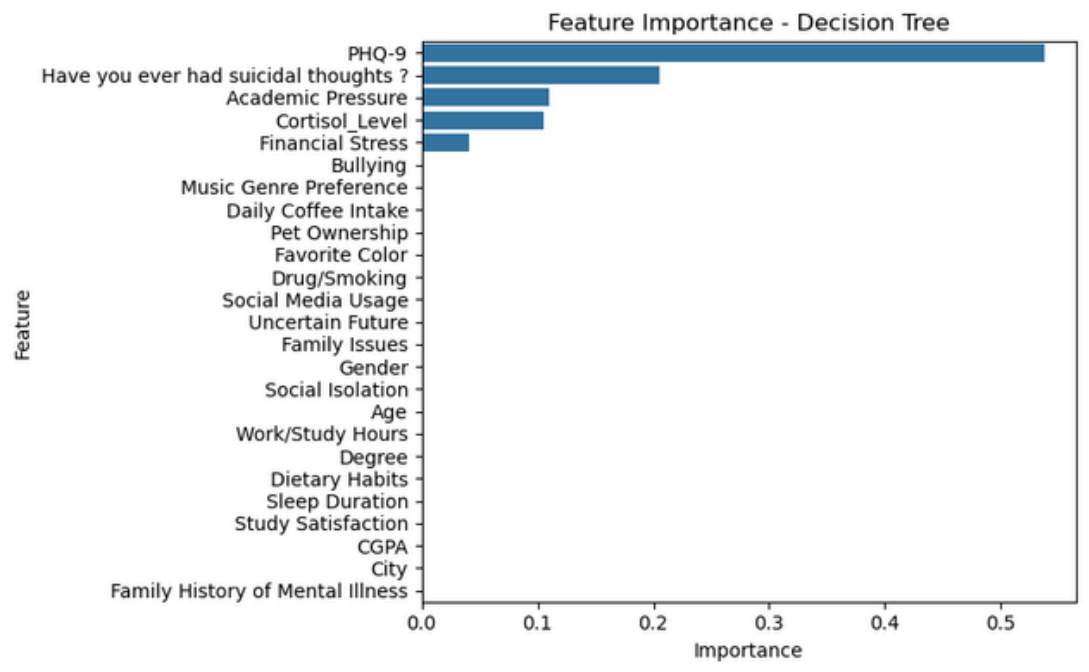




Ninth Model : Decision Tree

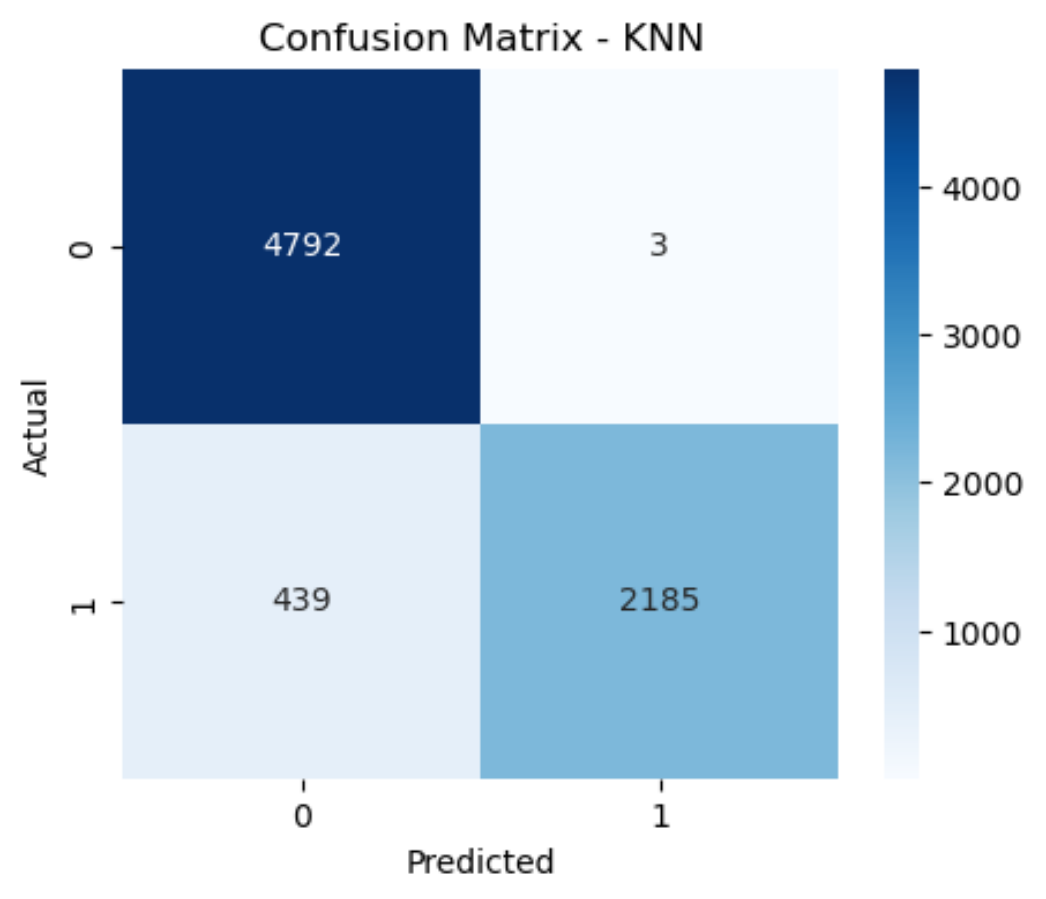
```
Decision Tree Training Accuracy: 0.9540
Decision Tree Test Accuracy: 0.9528

Confusion Matrix:
 [[4697   98]
 [ 252 2372]]

Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.98      0.96      4795
           1       0.96      0.90      0.93      2624

    accuracy                           0.95      7419
   macro avg       0.95      0.94      0.95      7419
weighted avg       0.95      0.95      0.95      7419
```





Tenth Model : KNN

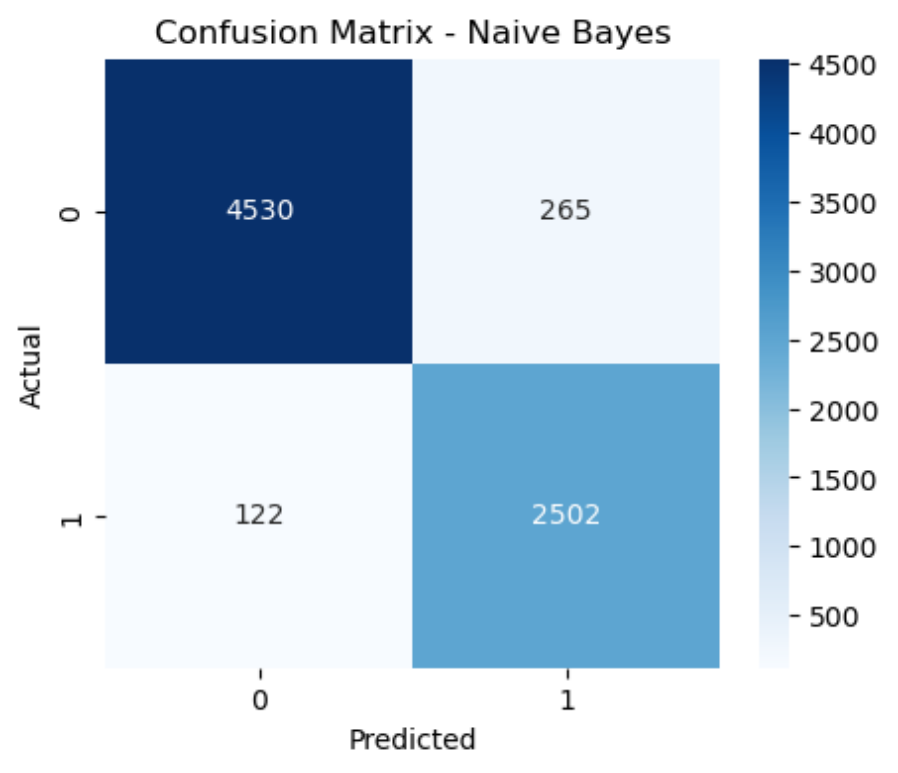
```
KNN Training Accuracy: 0.9648
KNN Test Accuracy: 0.9404

Confusion Matrix:
 [[4792    3]
 [ 439 2185]]

Classification Report:
              precision    recall  f1-score   support

           0       0.92      1.00      0.96      4795
           1       1.00      0.83      0.91      2624

    accuracy                           0.94      7419
   macro avg       0.96      0.92      0.93      7419
weighted avg       0.95      0.94      0.94      7419
```



DOESN'T SUPPORT FEATURE IMPORTANCE

Eleventh Model : NAIVE Bayes

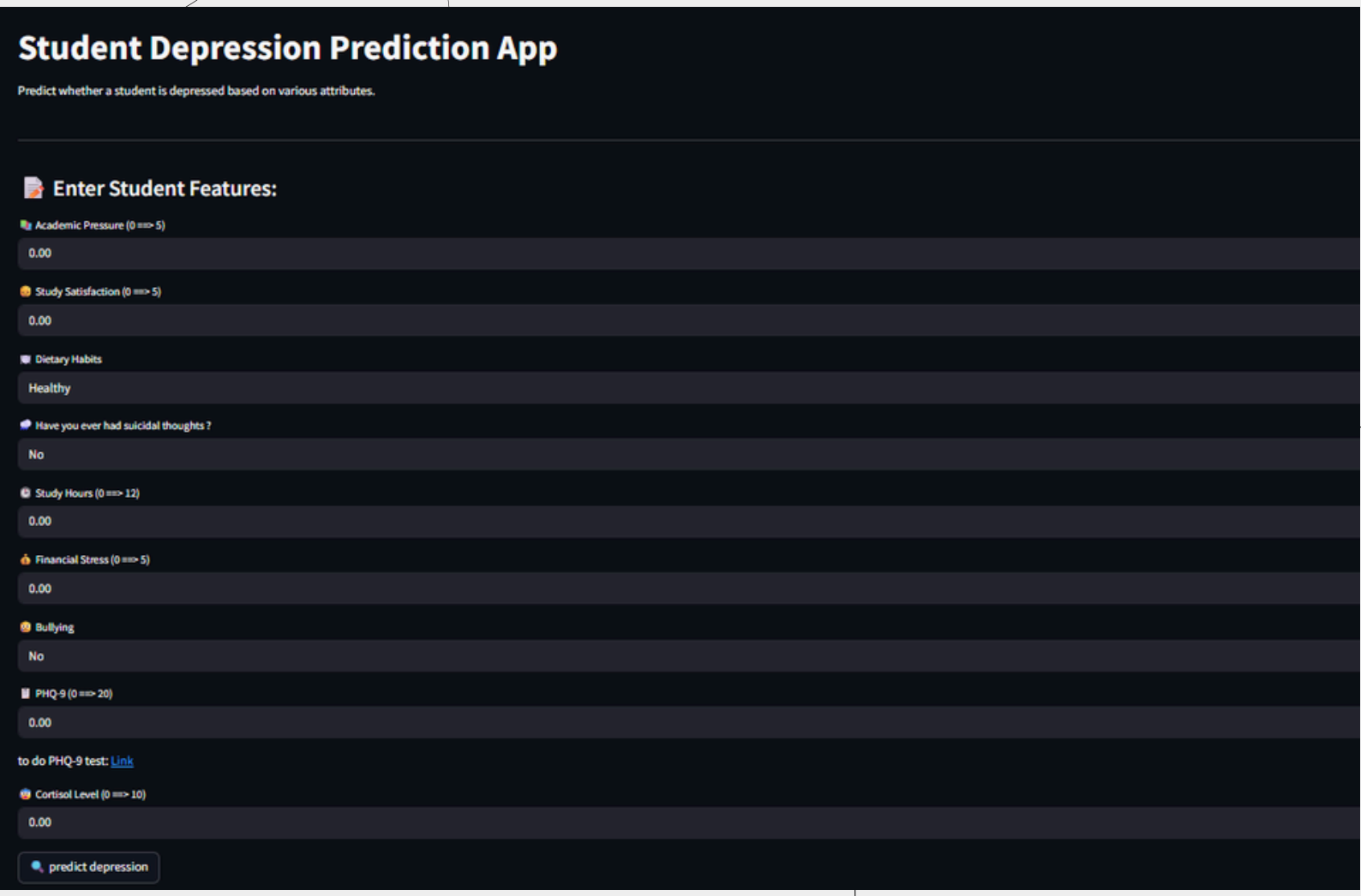
```
Naive Bayes Training Accuracy: 0.9506
Naive Bayes Test Accuracy: 0.9478

Confusion Matrix:
 [[4530  265]
 [ 122 2502]]

Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.94      0.96      4795
           1       0.90      0.95      0.93      2624

    accuracy                           0.95      7419
   macro avg       0.94      0.95      0.94      7419
weighted avg       0.95      0.95      0.95      7419
```



DOESN'T SUPPORT FEATURE IMPORTANCE

MILESTONE 4

To make the depression prediction model accessible and interactive, a web-based application titled "Student Depression Prediction App" was developed and deployed using Streamlit, a Python library for building data apps. The trained model used for prediction is based on the XGBoost algorithm, which provides high performance and accuracy for classification tasks. The model was serialized and loaded into the app using Pickle. The interface allows users to input key features such as academic pressure, study satisfaction, dietary habits, financial stress, bullying experiences, and more. These inputs are then processed in real time to predict the likelihood of depression. This deployment bridges the gap between data science and real-world application, making the tool practical for both students and mental health professionals.



MILESTONE 5

Our final Report & Presentation have been completed and uploaded elsewhere.

We are grateful to ENG. Eslam Elreedy for his efforts in helping us understand all the concepts clearly.
We would also like to express our thanks to DEPI for the great opportunity we have had.