

# Mental Health Analysis



Presented by: Team 7  
Haeyeon Jeong  
Sai Rachana Kandikattu  
Abiram Getie  
Yonathan Shimelis



How can we develop models to assess mental stress,  
and which factors are most influential in predicting  
mental health outcomes?

# Contents

- INTRODUCTION
- DATASET OVERVIEW
- DATA PREPROCESSING
- SMART QUESTIONS
- MODELS AND EVALUATIONS
- CONCLUSION



# Dataset Overview

Rows

292,364

Columns

17

Demographic	Health-Related
<ul style="list-style-type: none"><li>• Gender</li><li>• Country</li><li>• Occupation</li><li>• Self_Employed</li></ul>	<ul style="list-style-type: none"><li>• Family_History</li><li>• Treatment</li><li>• Mental_Health_History</li><li>• Growing_Stress</li></ul> <ul style="list-style-type: none"><li>• Mood_Swings <b>(Ordinal)</b></li><li>• Coping_Struggles</li><li>• Social_Weakness</li><li>• Mental_Health_Interview</li></ul>
Behavioral or Activity-Related	Time
<ul style="list-style-type: none"><li>• Days_Indoors <b>(Ordinal)</b></li><li>• Changes_Habits</li><li>• Work_Interest</li><li>• Care_Options</li></ul>	<ul style="list-style-type: none"><li>• Timestamp</li></ul>

# Data Preprocessing

Rows

-5,202



Columns

-3



**Missing Values  
Handling**

Drop rows with NA

Self\_Employed

**Drop Irrelevant  
Columns**

Self\_Employed  
Timestamp  
Country

**Feature  
Transformation**

Categorical → Numerical

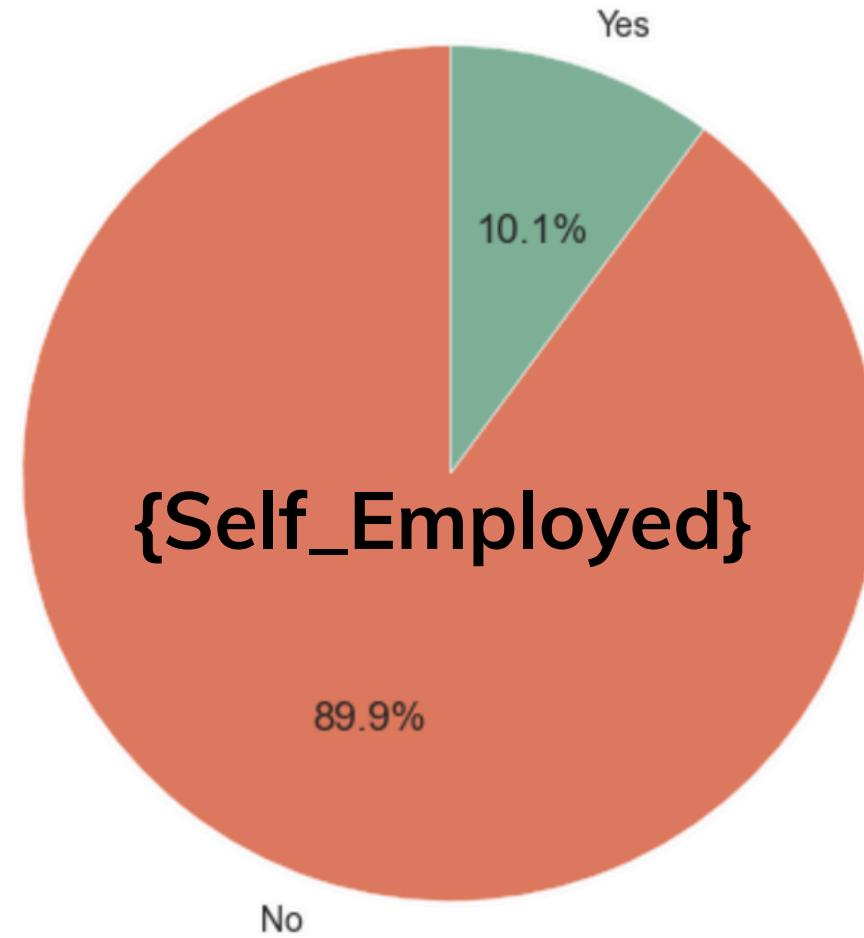
**One-Hot  
Encoding**

Occupation

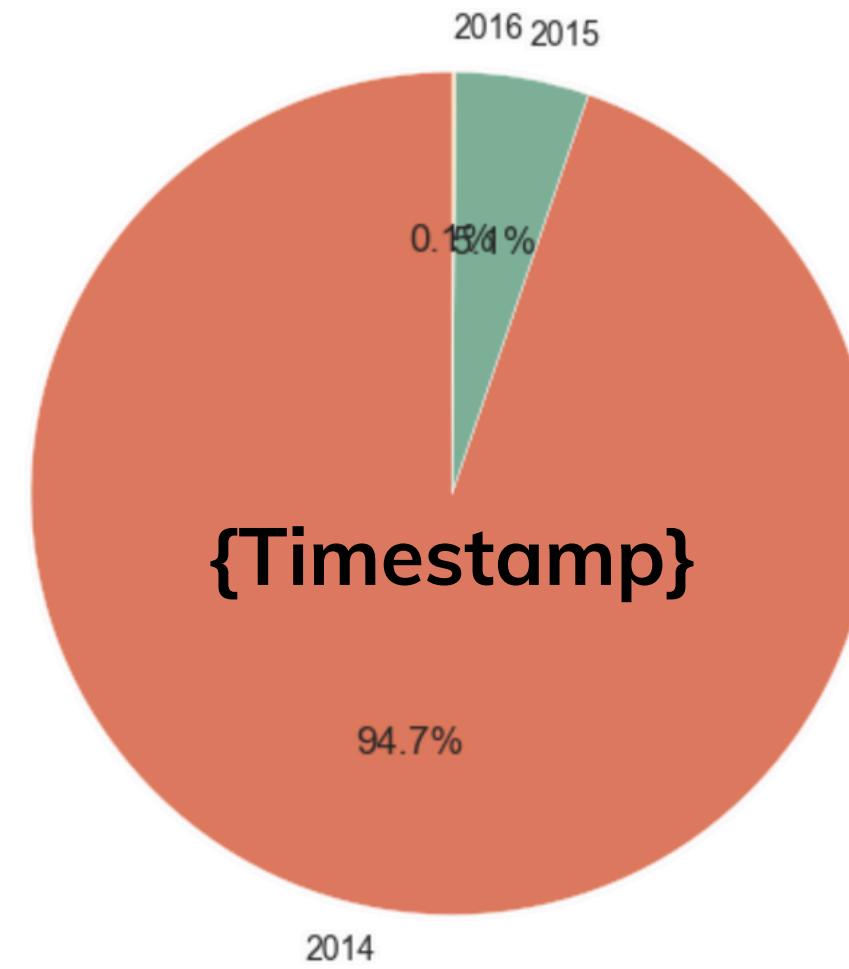
# EDA & Statistical Testing

$H_0$  : {The feature} and Growing\_Stress are independent.

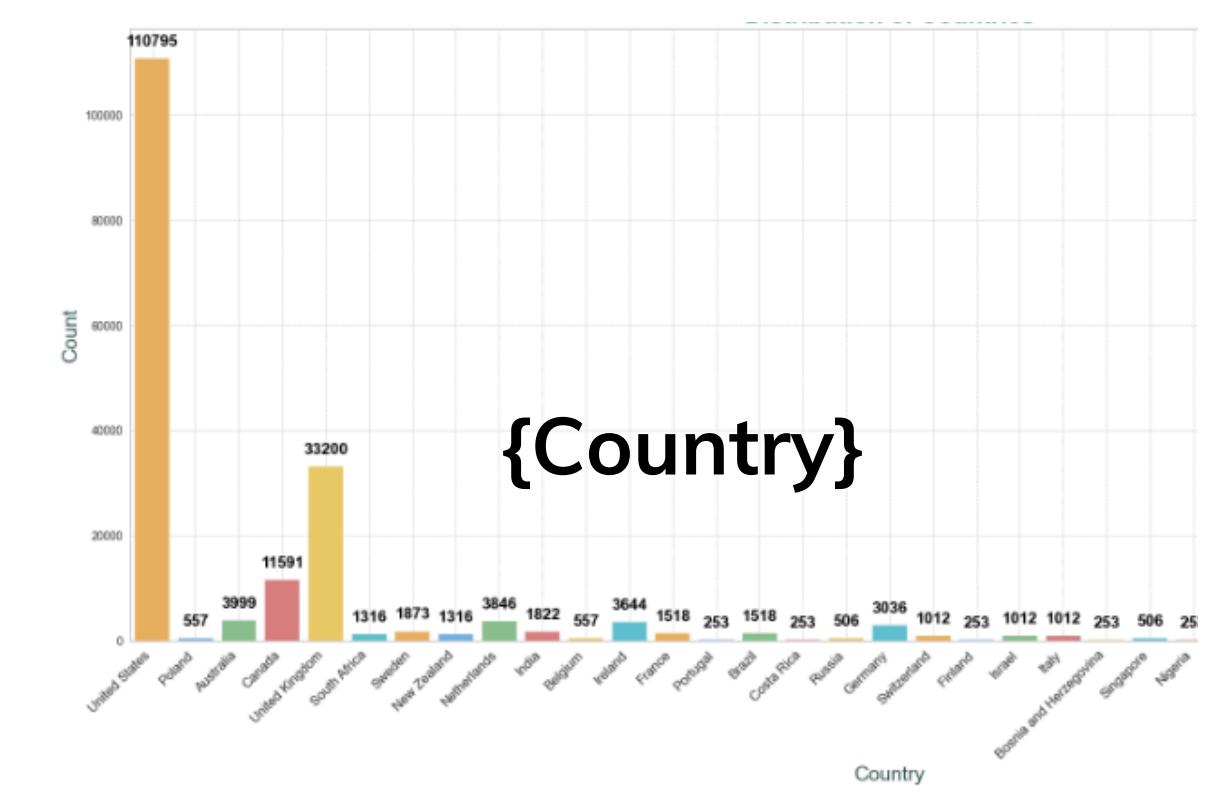
$H_1$  : {The feature} and Growing\_Stress are not independent.



p-value = 0.15



p-value = 0.93



p-value = 0.99

# Target Questions

Q1. What are the **top 5 factors** that influence the **growing stress**?

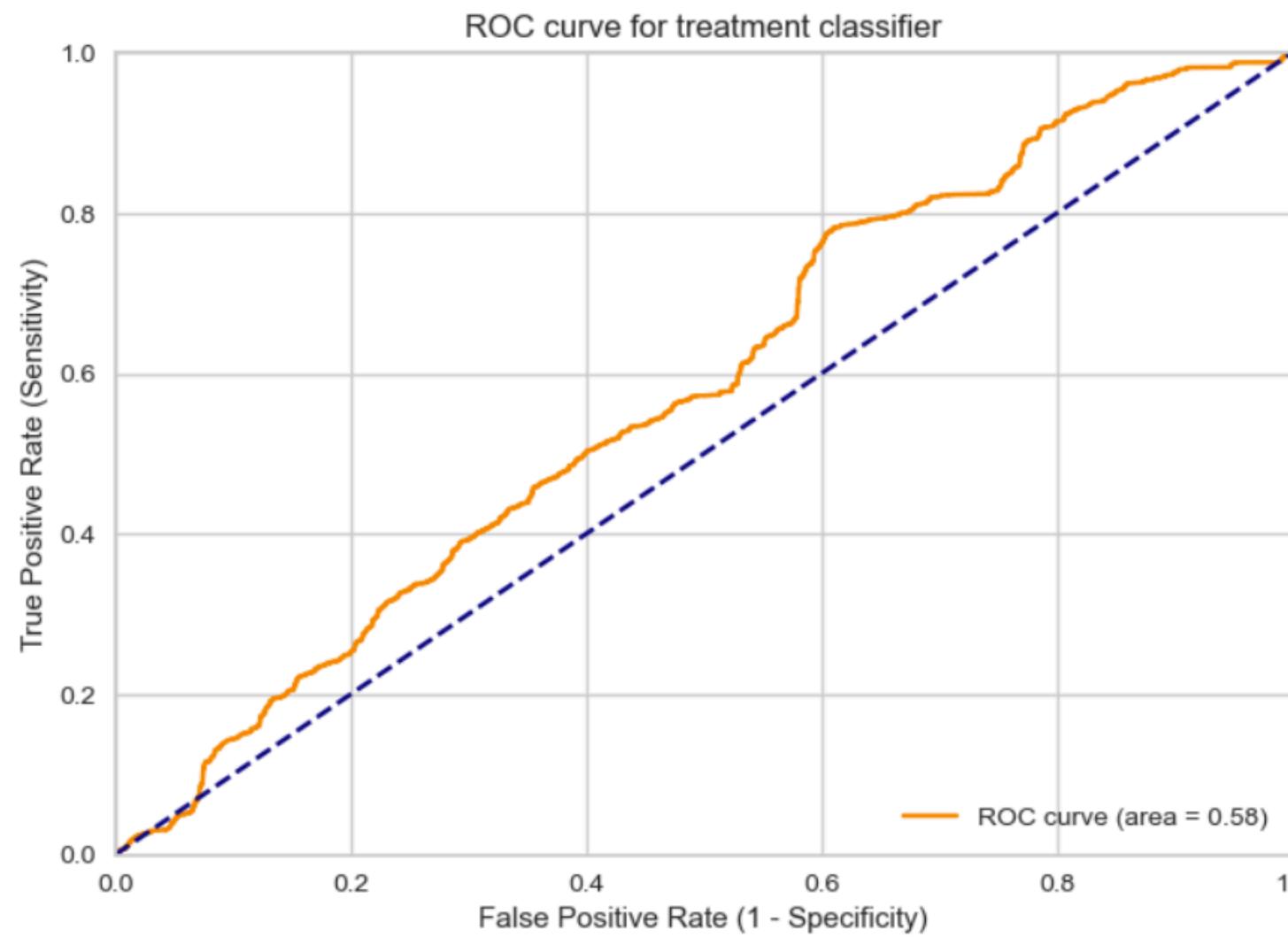
Q2. Do people with **family history** receive **treatment** or not?

Q3. What are the **factors** that impact the **growing stress for students**?

# Q1. What are the top 5 factors that influence the growing stress?

## A) Logistic Regression

Number of Features Selected : 16



Accuracy: 54%

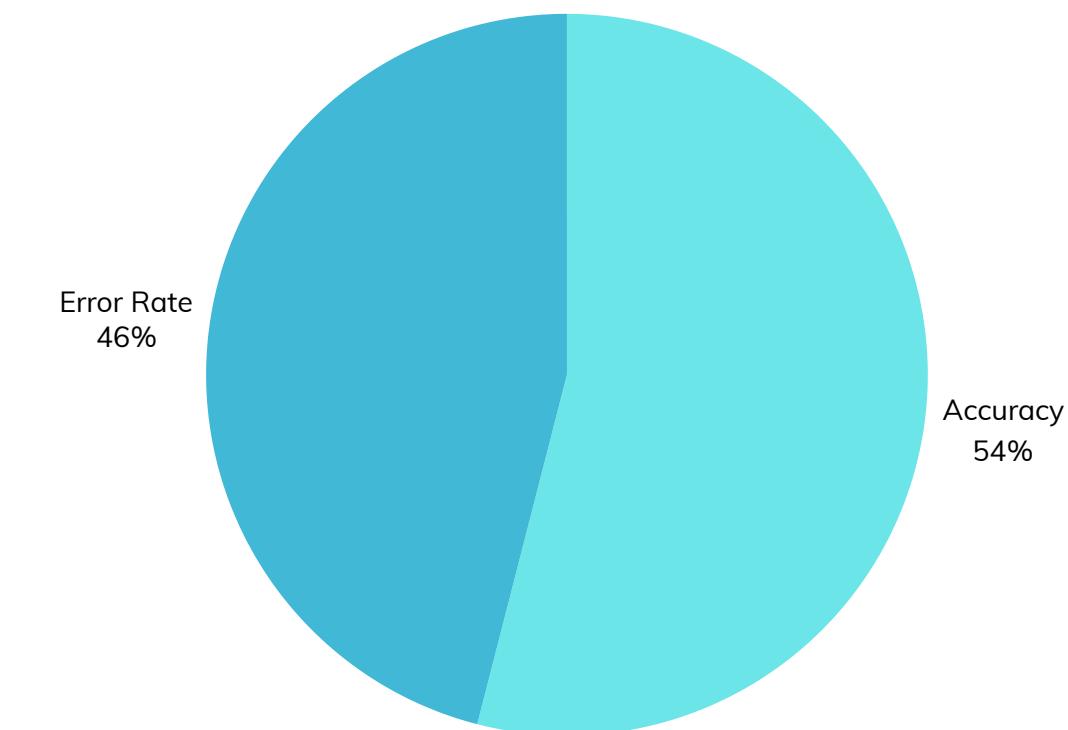
ROC: 0.58

Precision : 53%

Recall : 46%

F1- Score : 49%

Confusion Matrix		
		Predicted
Actual	0	1
	8463	9866
	7486	11973



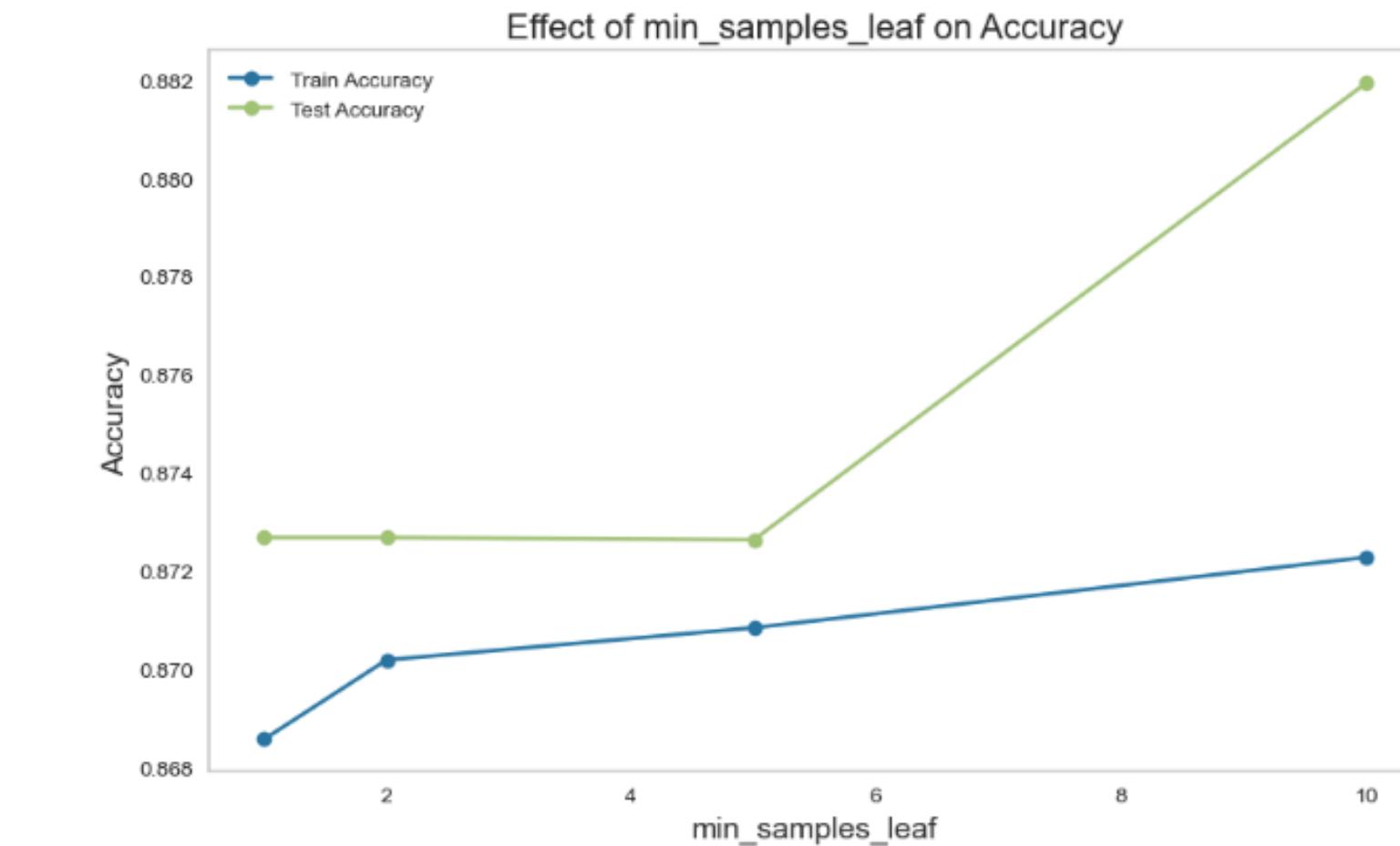
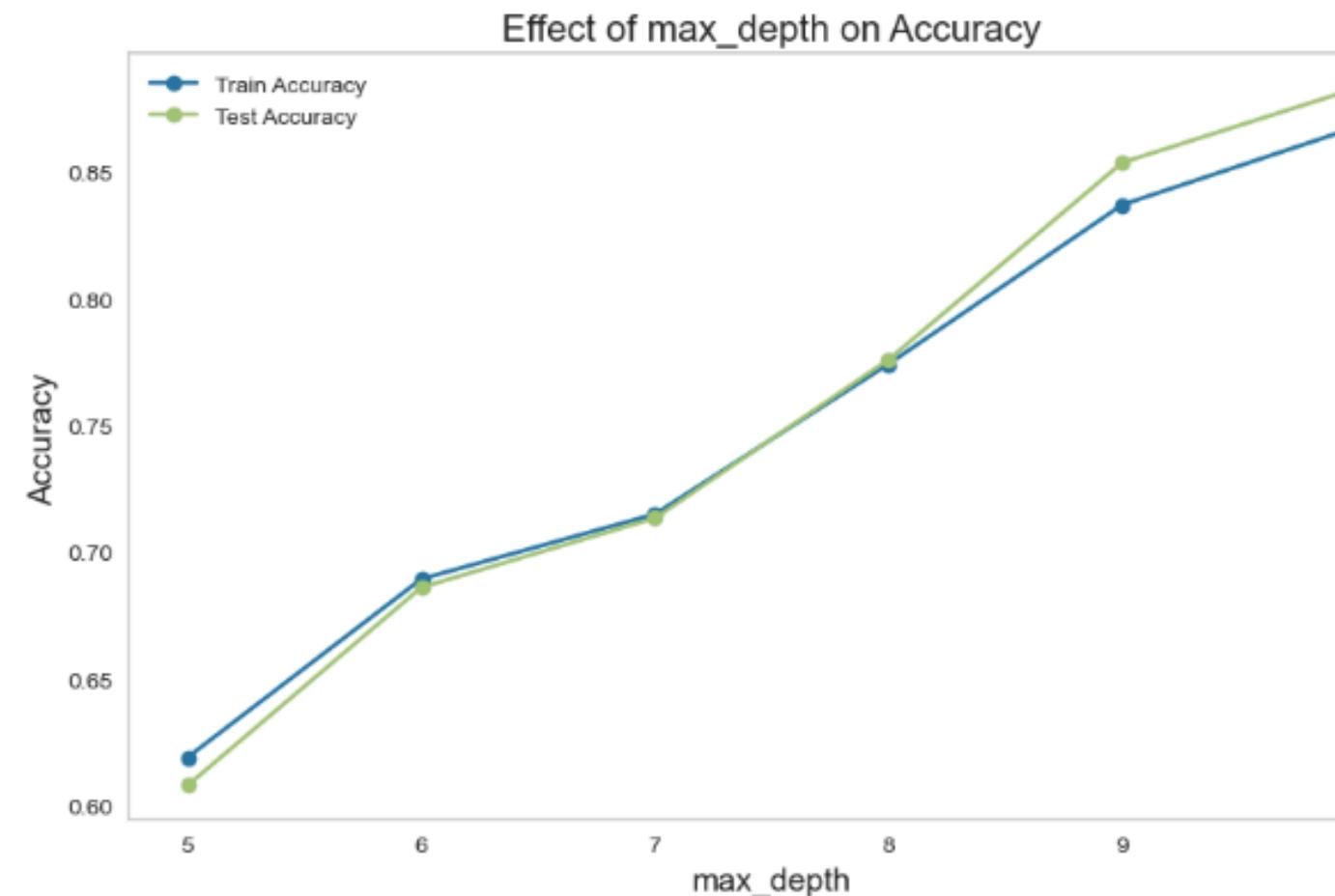
# Q1. What are the top 5 factors that influence the growing stress?

## B) Decision Tree

Number of Features Selected : 16

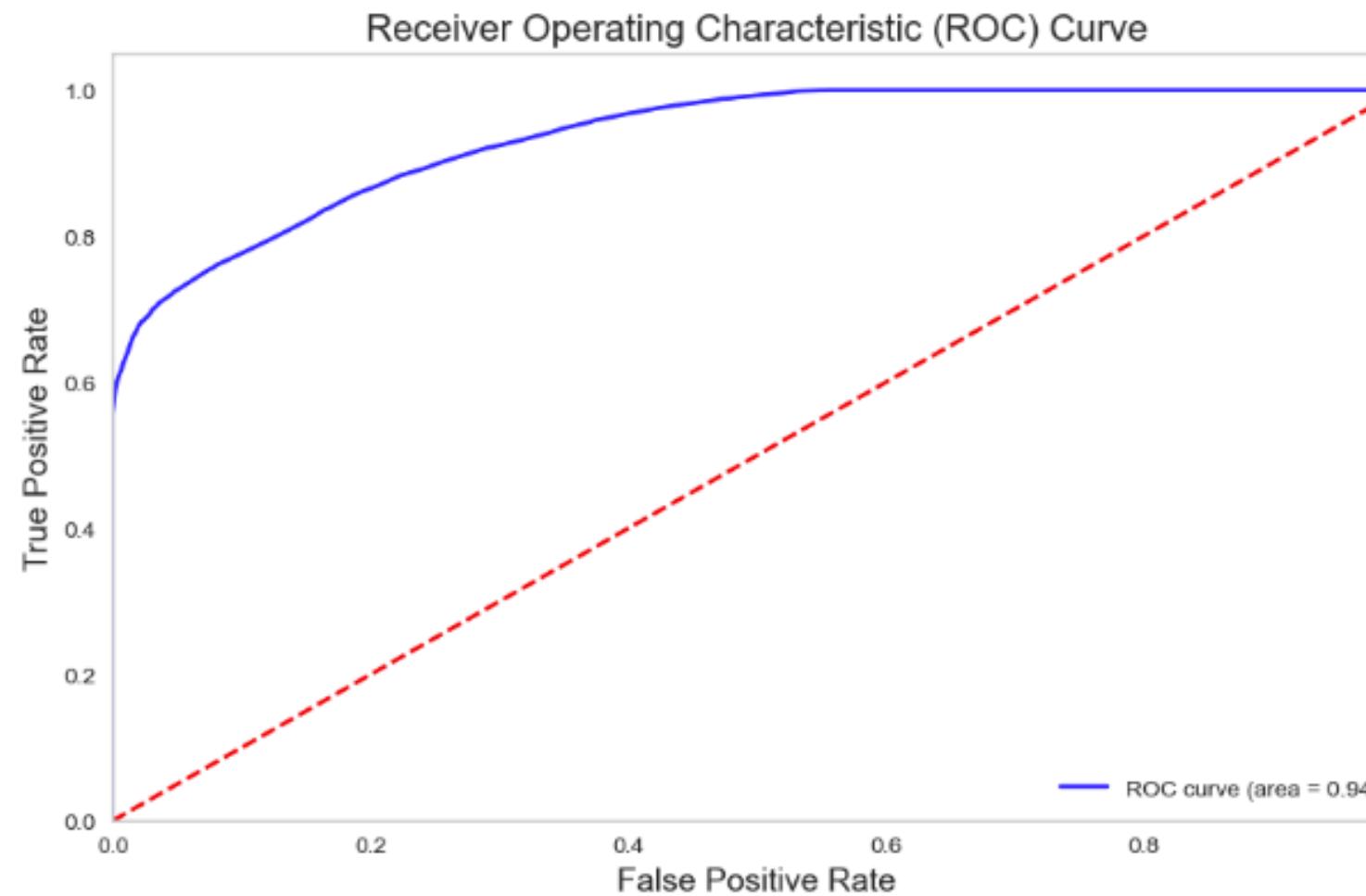
### Hyperparameter Tuning

Depth of the Tree : 9  
Min Samples Leaf : 10  
Min Samples Split : 10



# Q1. What are the top 5 factors that influence the growing stress?

## B) Decision Tree



Train Accuracy: 83.7%

Test Accuracy : 83.4%

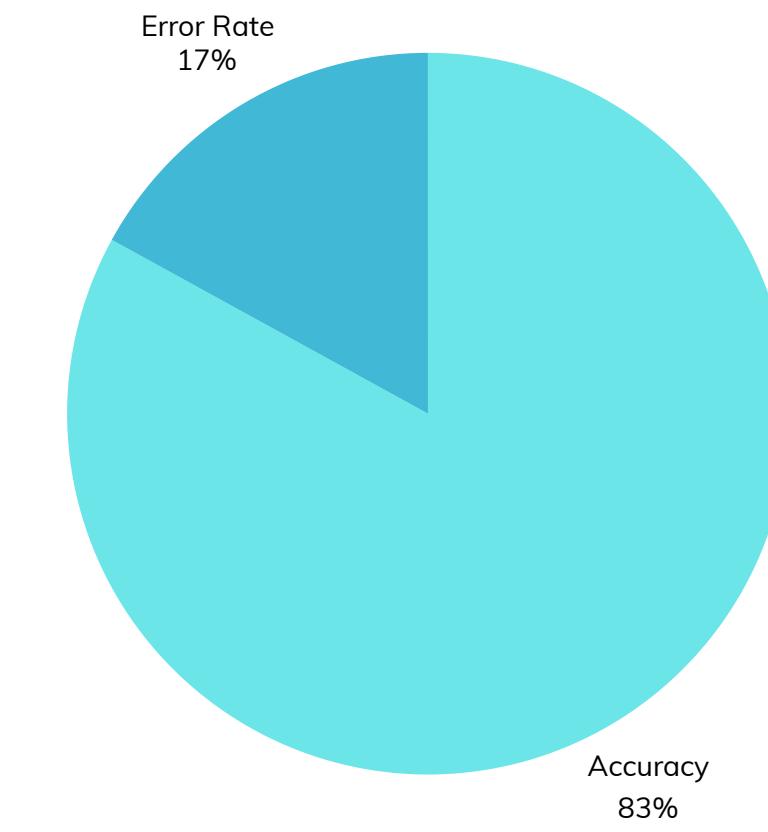
Precision: 80%

Recall : 87%

F1-score : 83%

ROC: 0.94

Actual	Predicted	
	0	1
0	15628	2393
1	3854	15913



# Q1. What are the top 5 factors that influence the growing stress?

## C) Random Forest

Number of Features Selected : 16

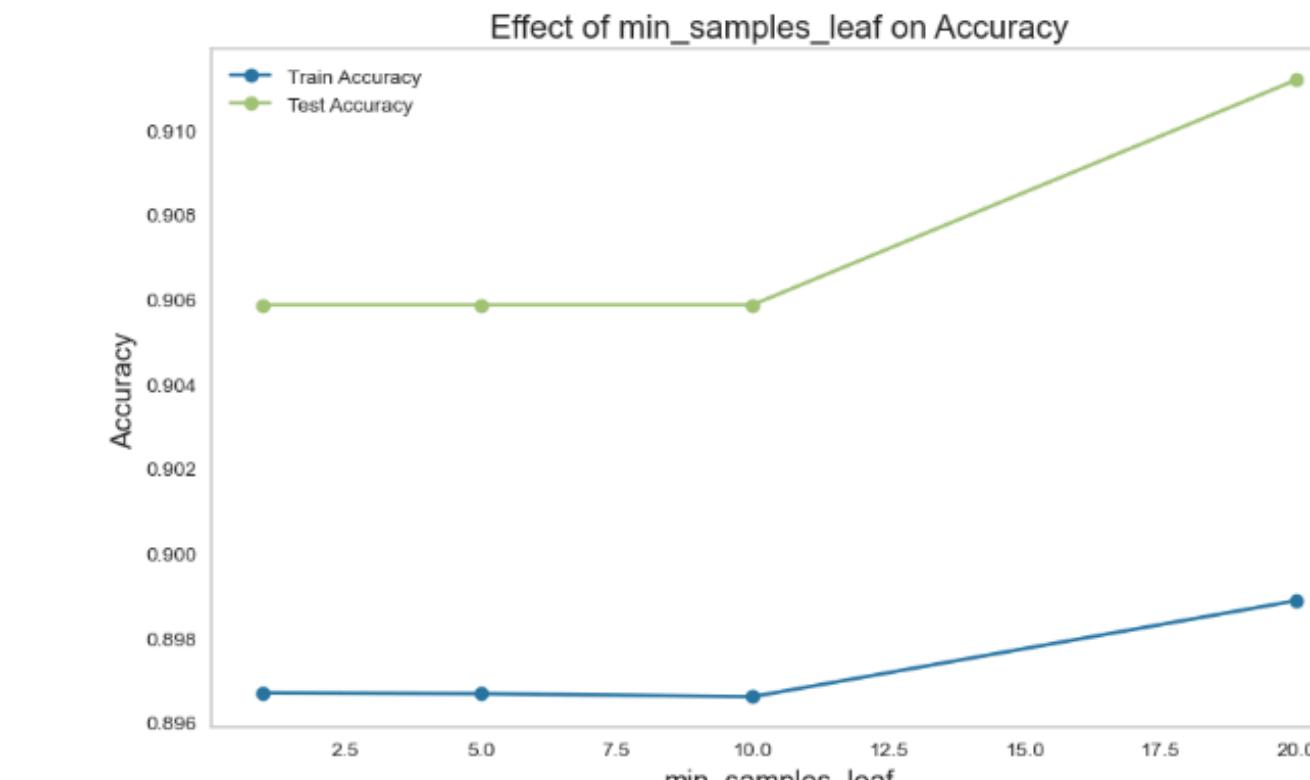
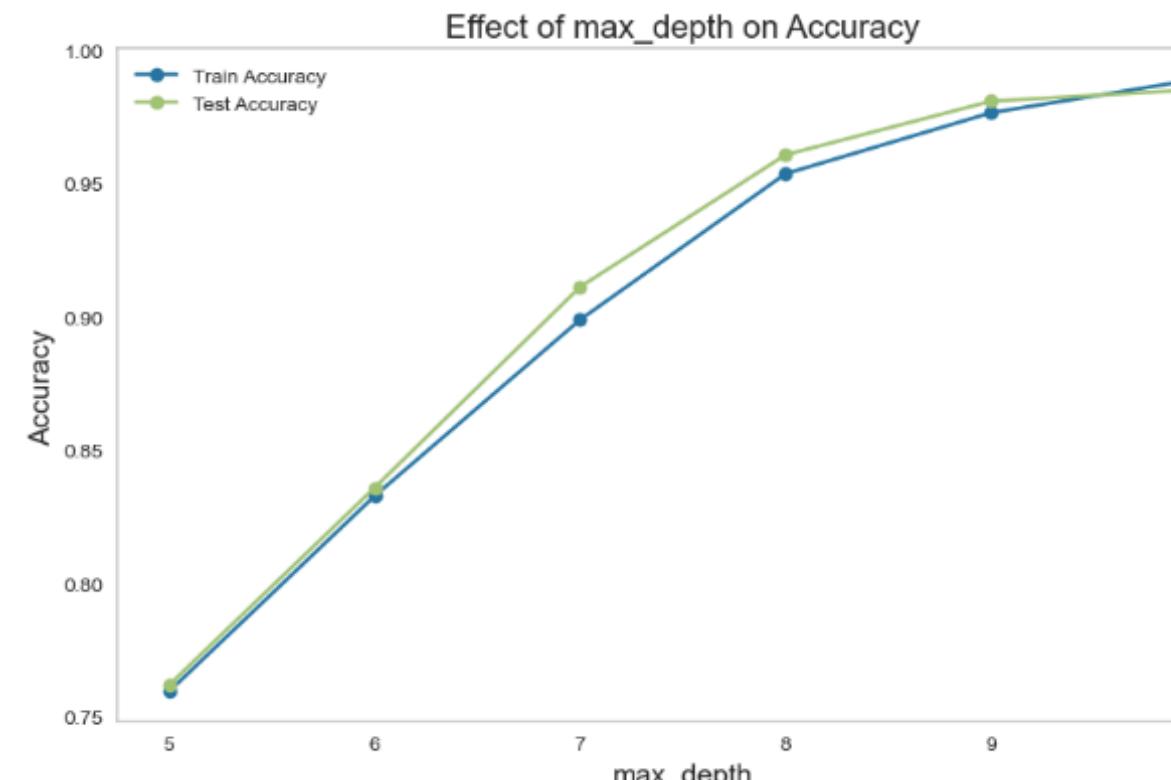
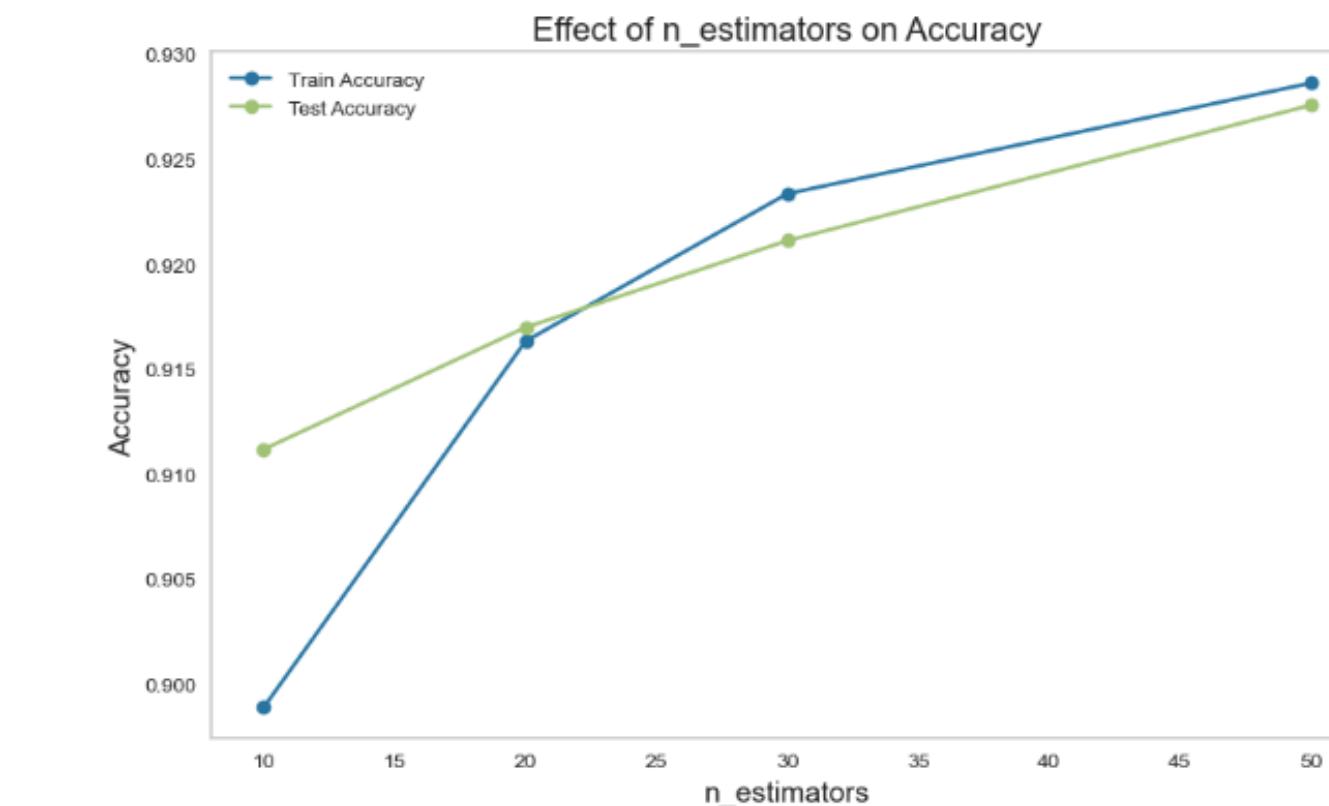
### Hyperparameter Tuning

Depth of the Tree : 7

Min Samples Leaf : 20

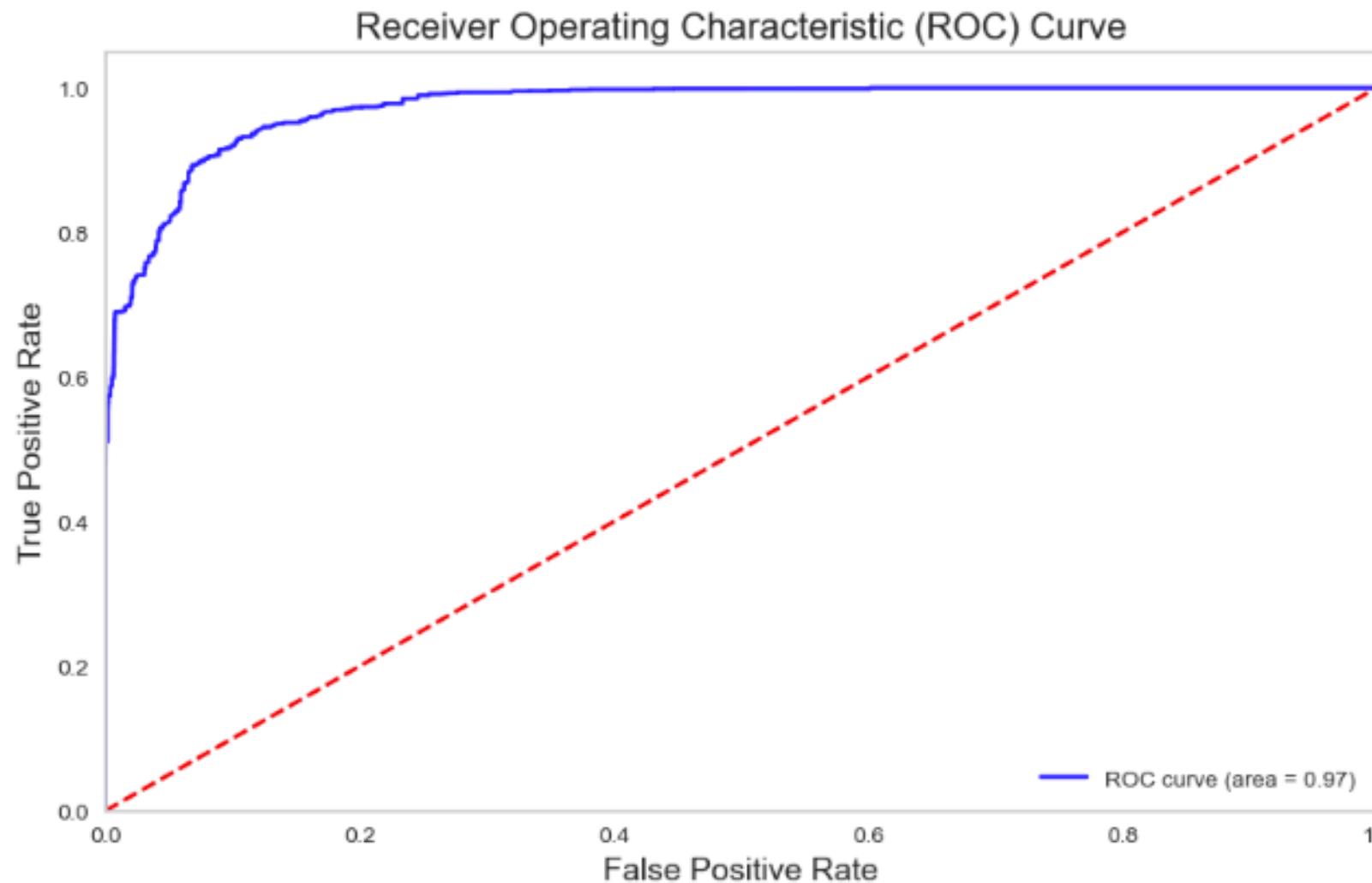
Min Samples Split : 20

N Estimators : 10



# Q1. What are the top 5 factors that influence the growing stress?

## C) Random Forest



Confusion Matrix		
		Predicted
Actual	0	1
	0	1924
1	16097	1388
		18379

Train Accuracy: 91.1%

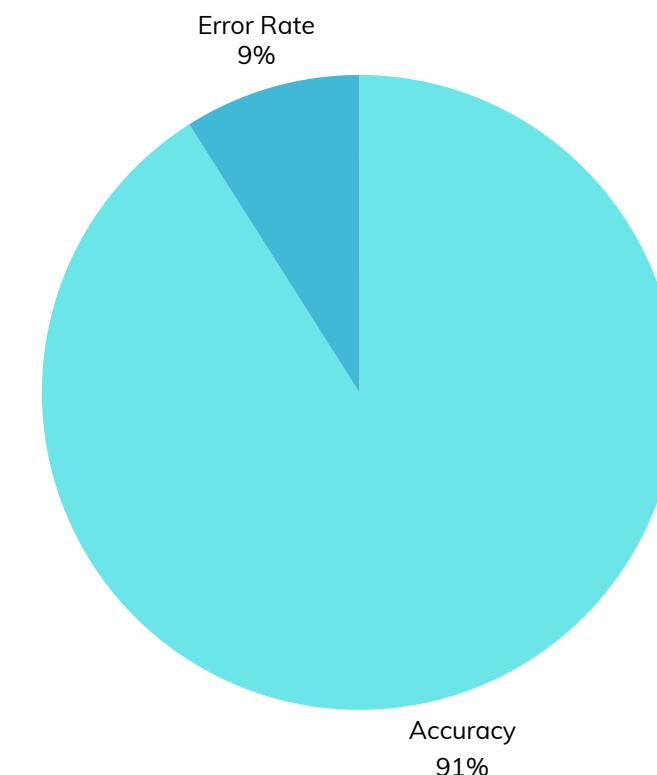
Test Accuracy : 91.2%

Precision: 92%

Recall : 89%

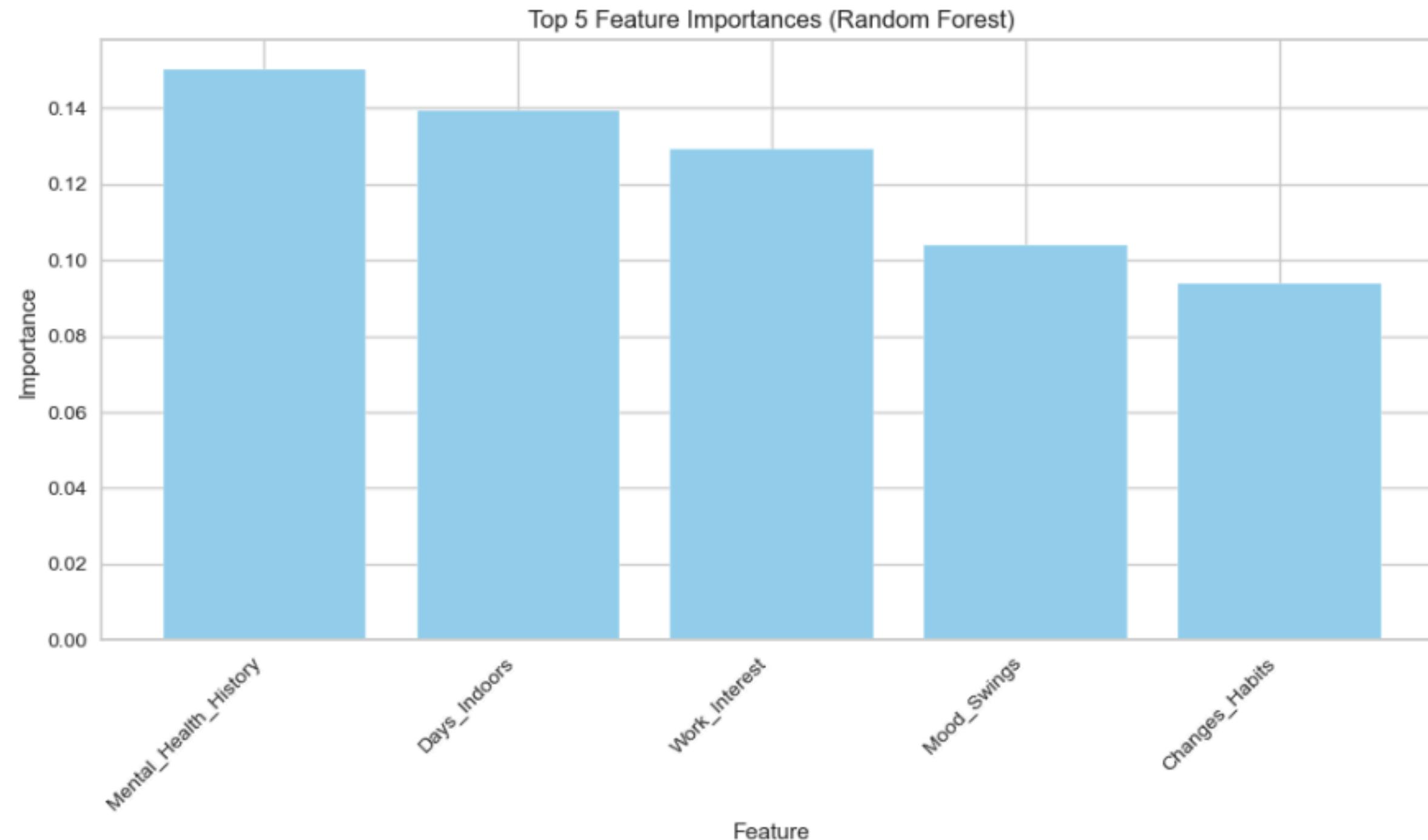
F1-score : 91%

ROC: 0.97



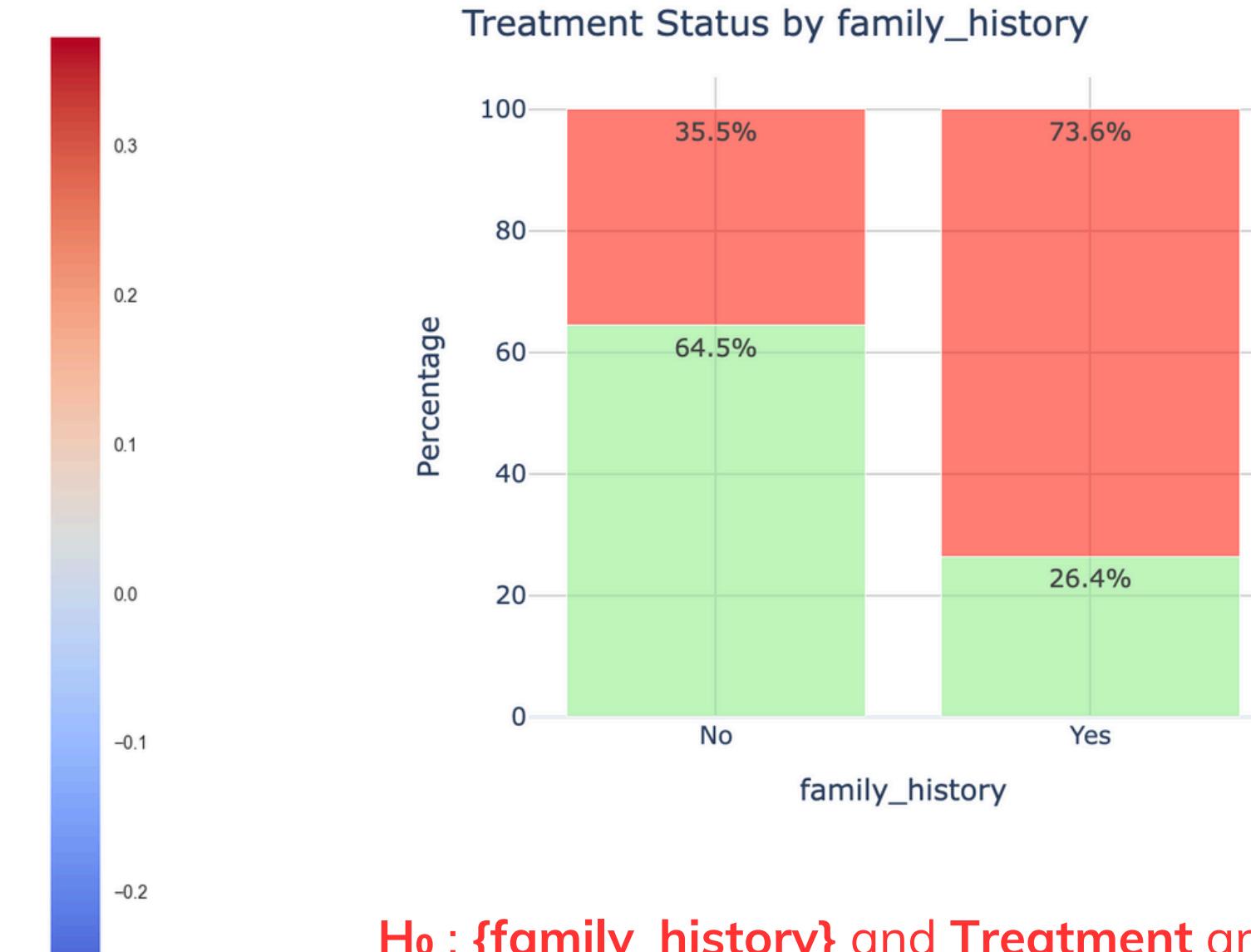
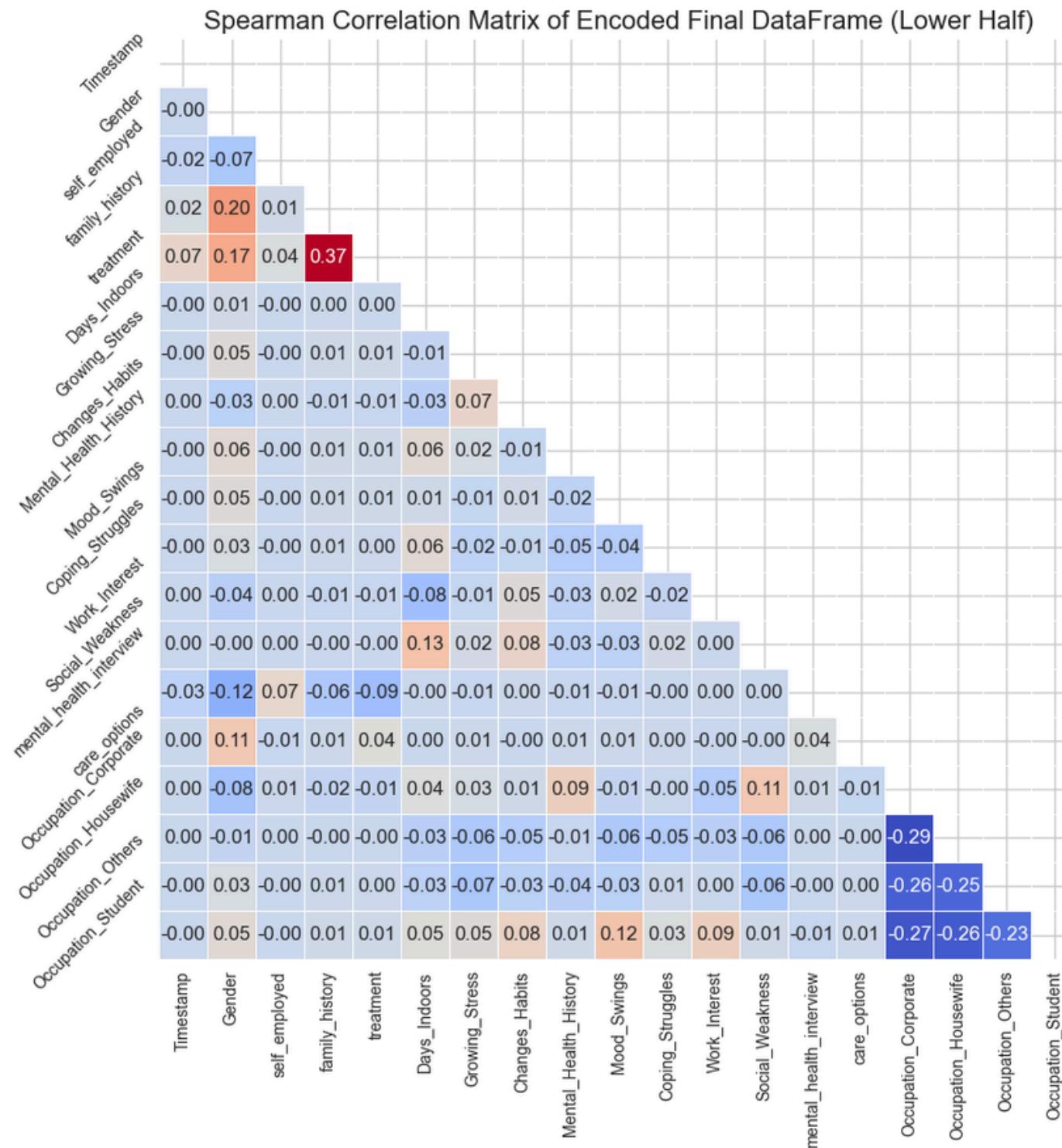
# Q1. What are the top 5 factors that influence the growing stress?

C) Random Forest



# Q2. Do people with family history receive treatment or not?

## Background



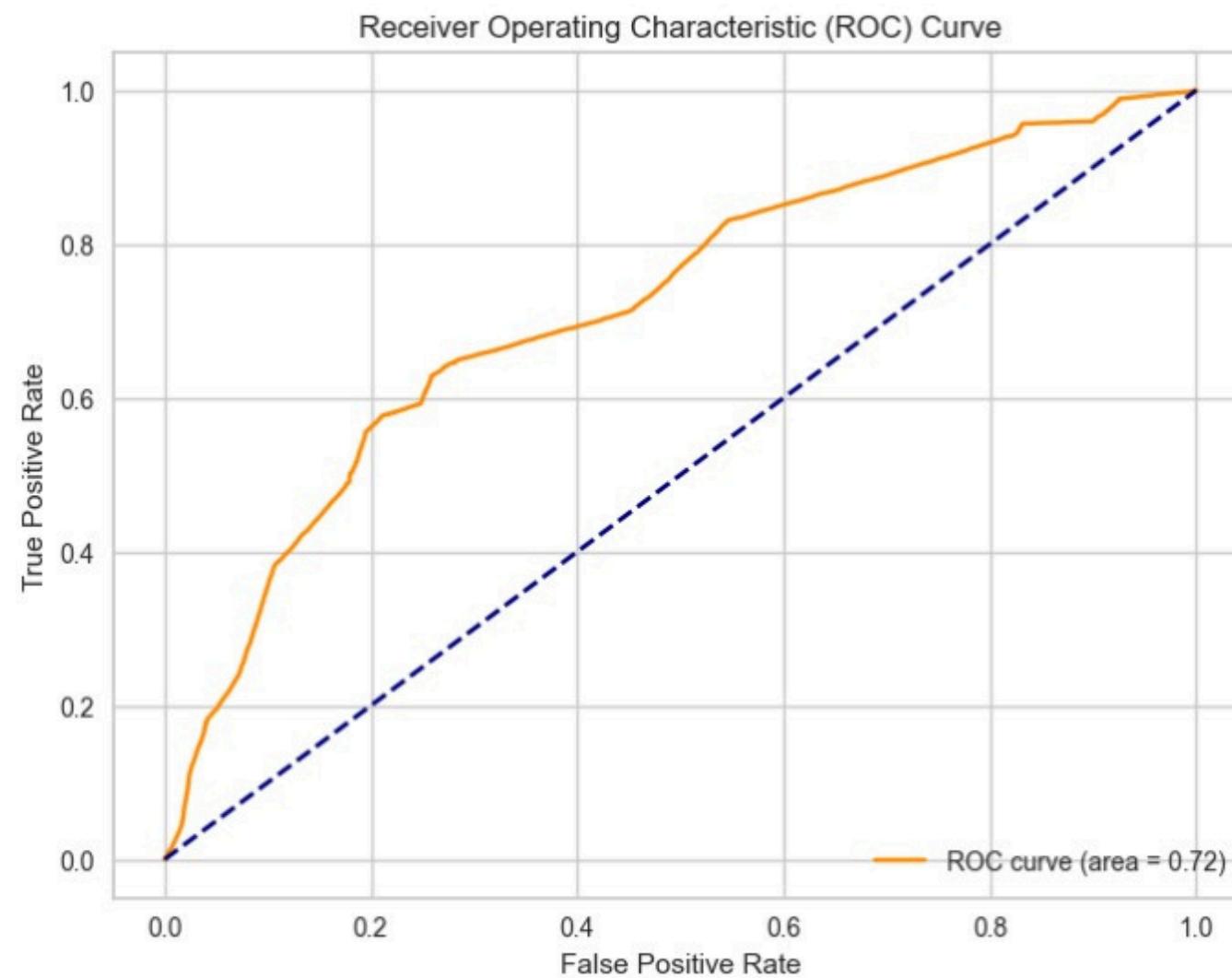
$H_0$  : {family\_history} and Treatment are independent.  
 $H_1$  : {family\_history} and Treatment are not independent.

p-value = 0.00

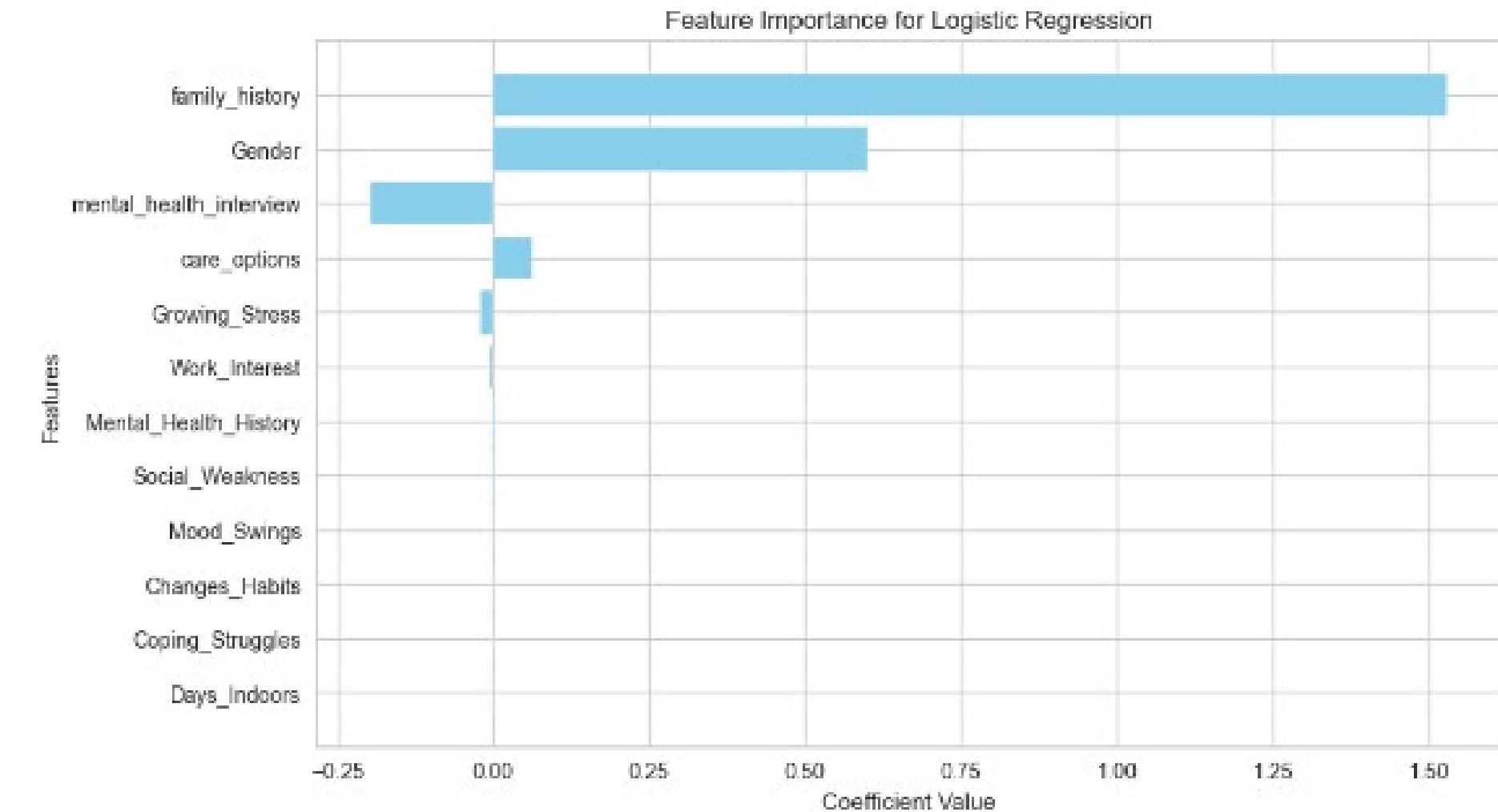
# Q2. Do people with family history receive treatment or not?

## A) Logistic Regression

Number of Features Selected : 12



Accuracy: 68%  
ROC: 0.72

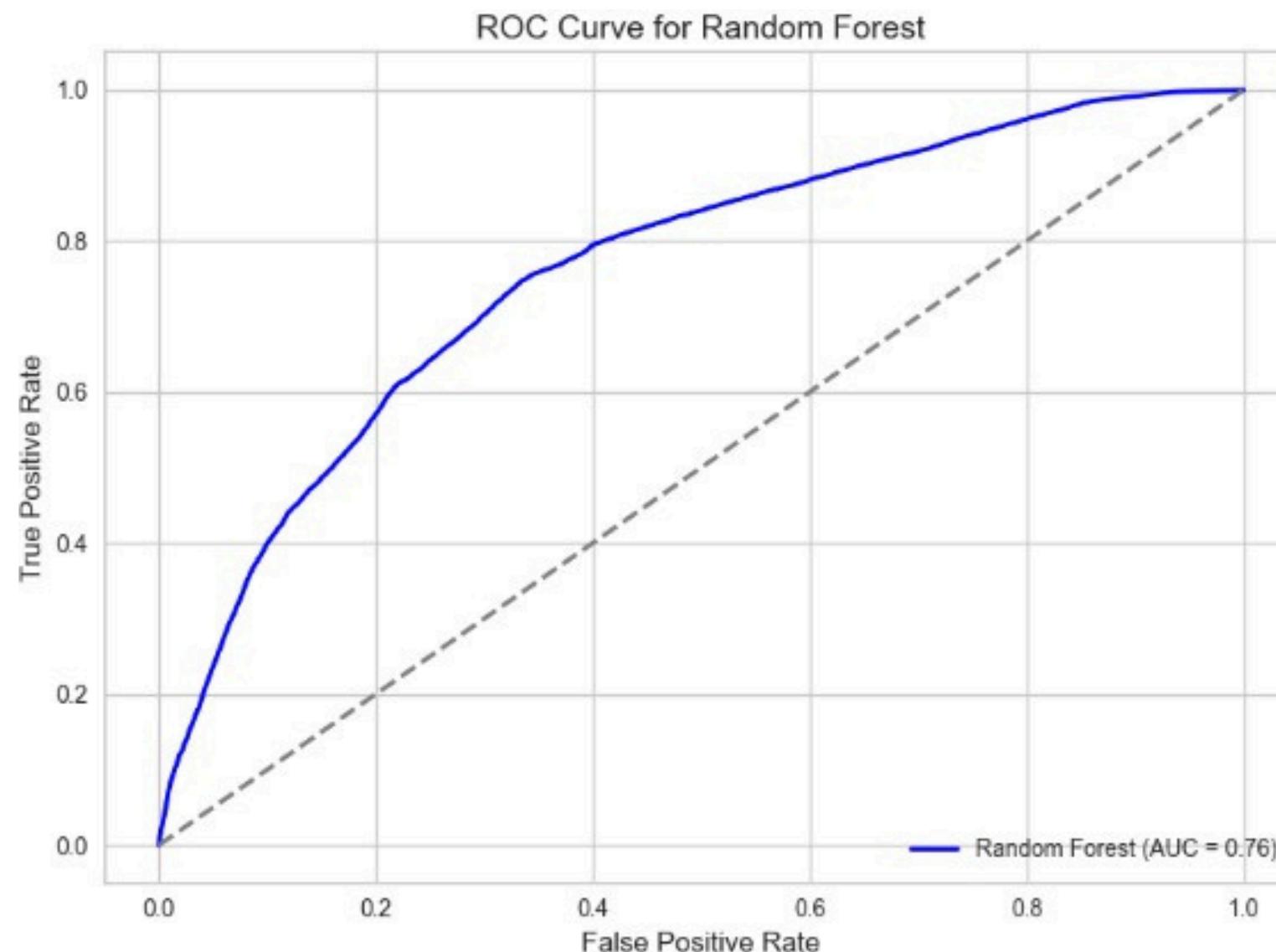


**family\_history**  
**Gender**  
**mental\_health\_interview**

# Q2. Do people with family history receive treatment or not?

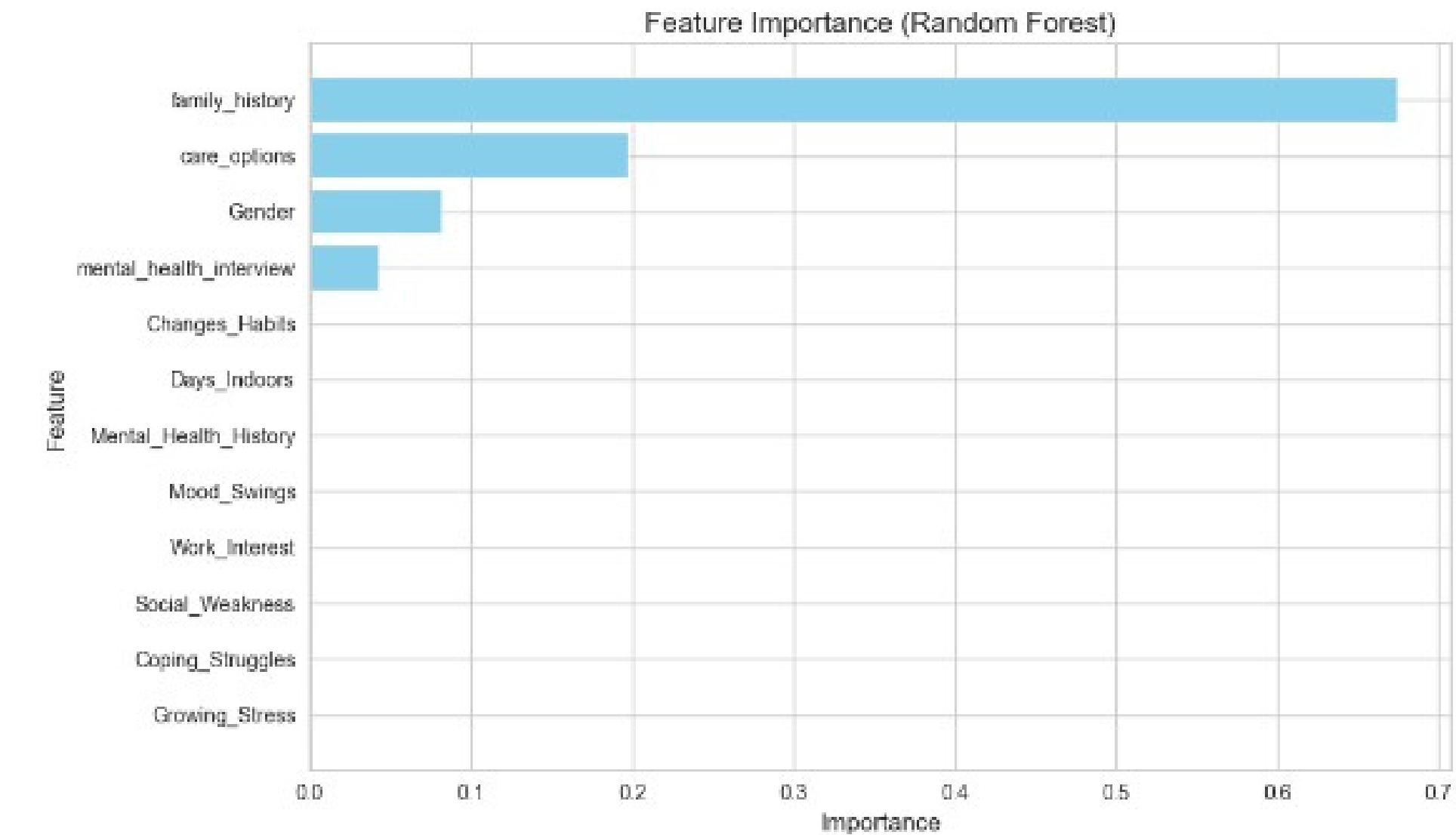
## B) Random Forest

Number of Features Selected : 12



Accuracy: 70%

ROC: 0.76



family\_history

care\_options

Gender

# **Q3. What are the factors that impact the growing stress for students?**

## **Steps**

### **1. Preprocessing**

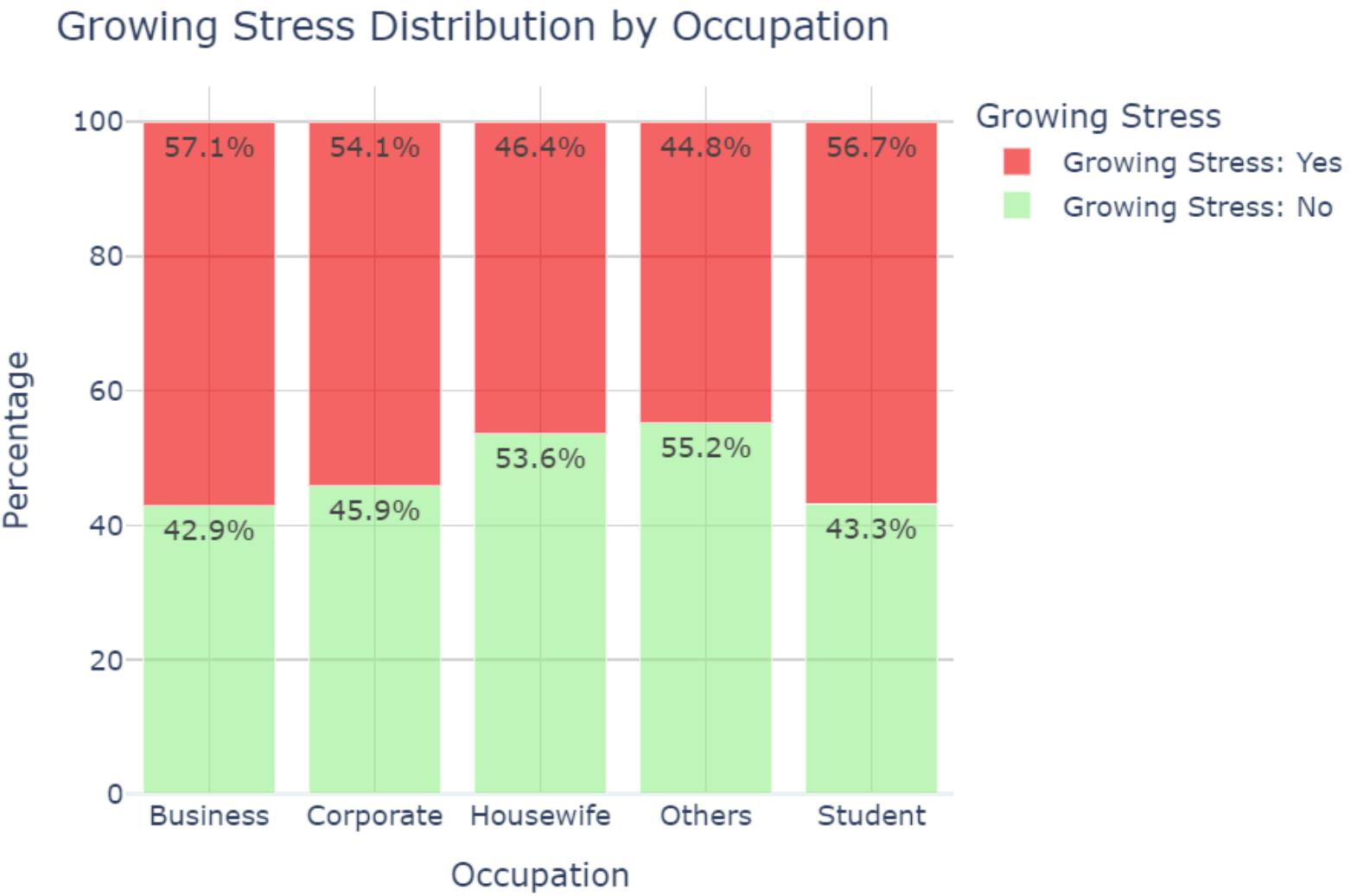
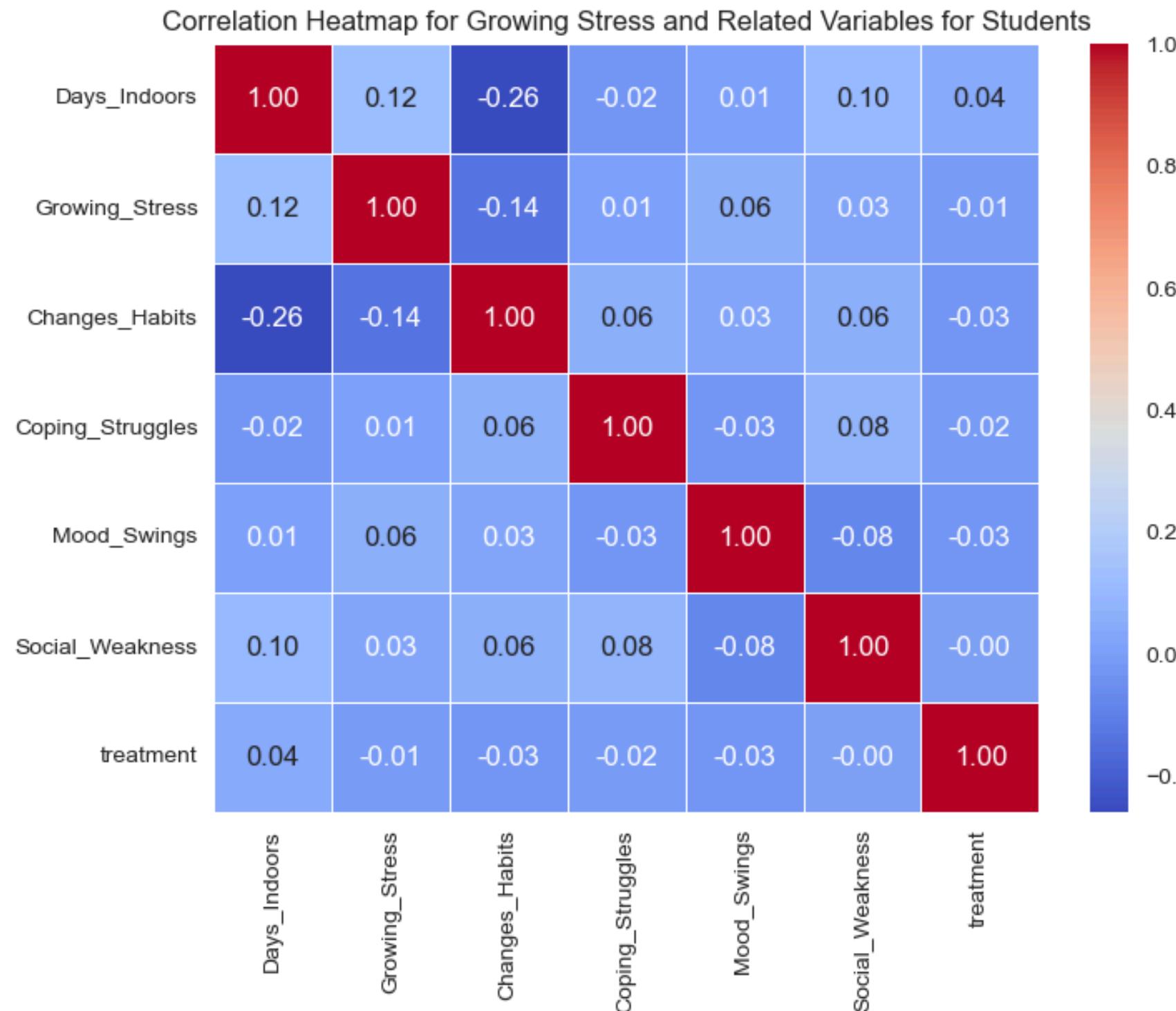
- a. Filter data for students
- b. Drop unwanted columns/predictors
- c. Drop “Maybe” value in our target variable

### **2. Models used**

Logistic Regression

# Q3. What are the factors that impact the growing stress for students?

## Background



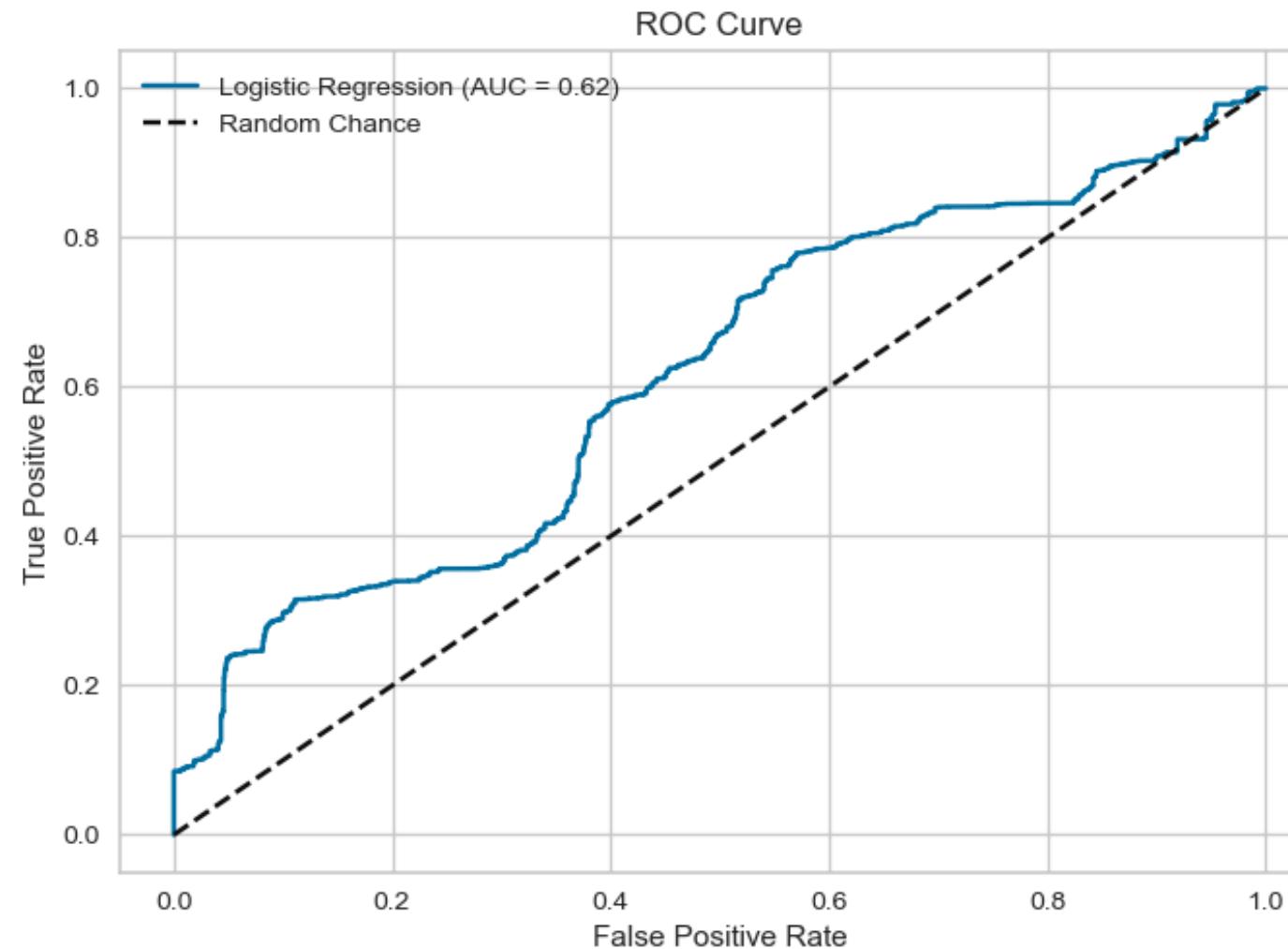
$H_0$  : {Occupation} and Growing Stress are independent.  
 $H_1$  : {Occupation} and Growing Stress are not independent.

p-value = 0.00

# Q3. What are the factors that impact the growing stress for students?

## Logistic Regression

Number of Features Selected : 13

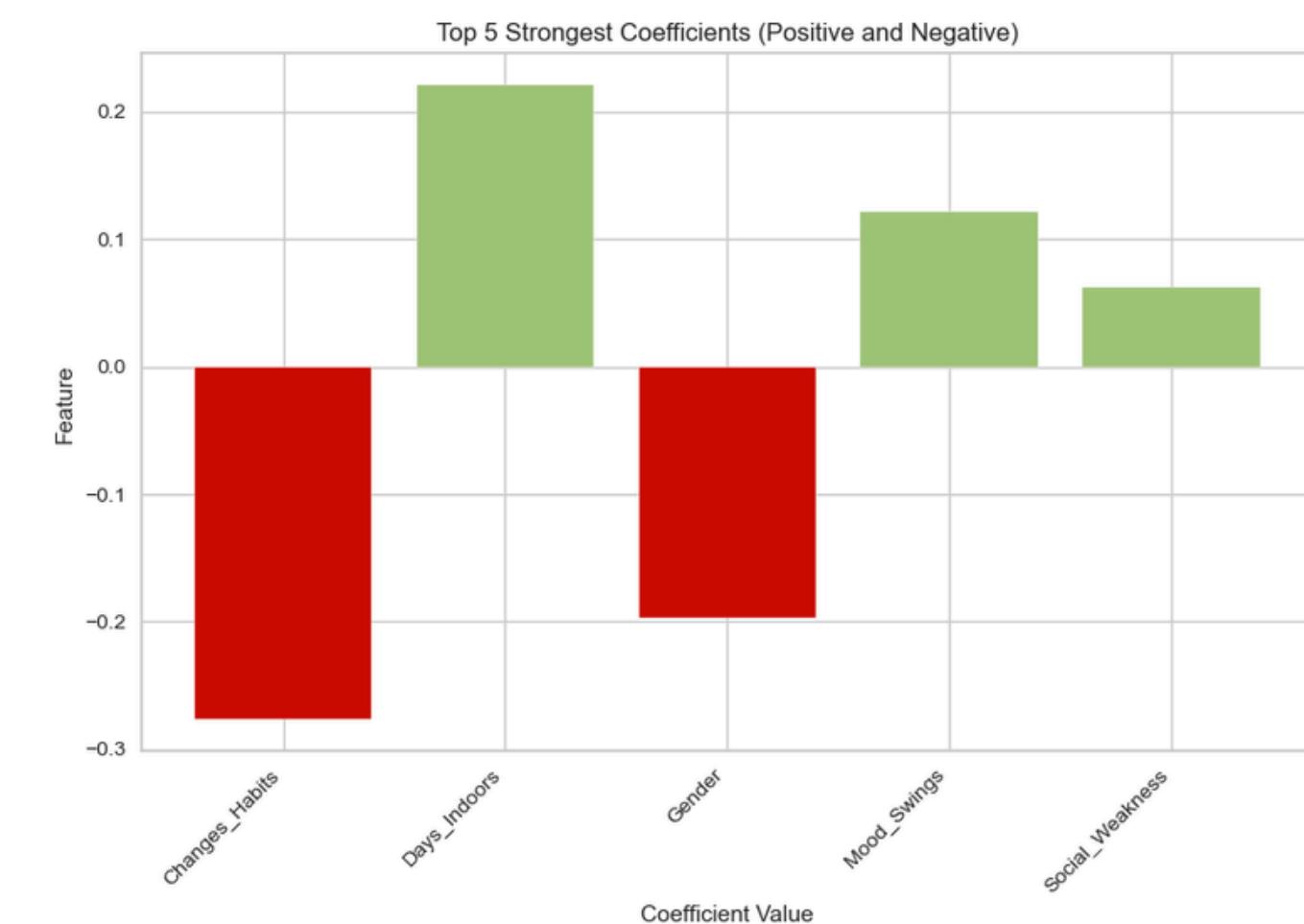


Accuracy: 61%

ROC: 0.62

Changes\_Habits: -0.27  
Days\_Indoors: 0.22  
Gender: -0.20  
Mood\_Swings: 0.12  
Social\_Weakness: 0.06

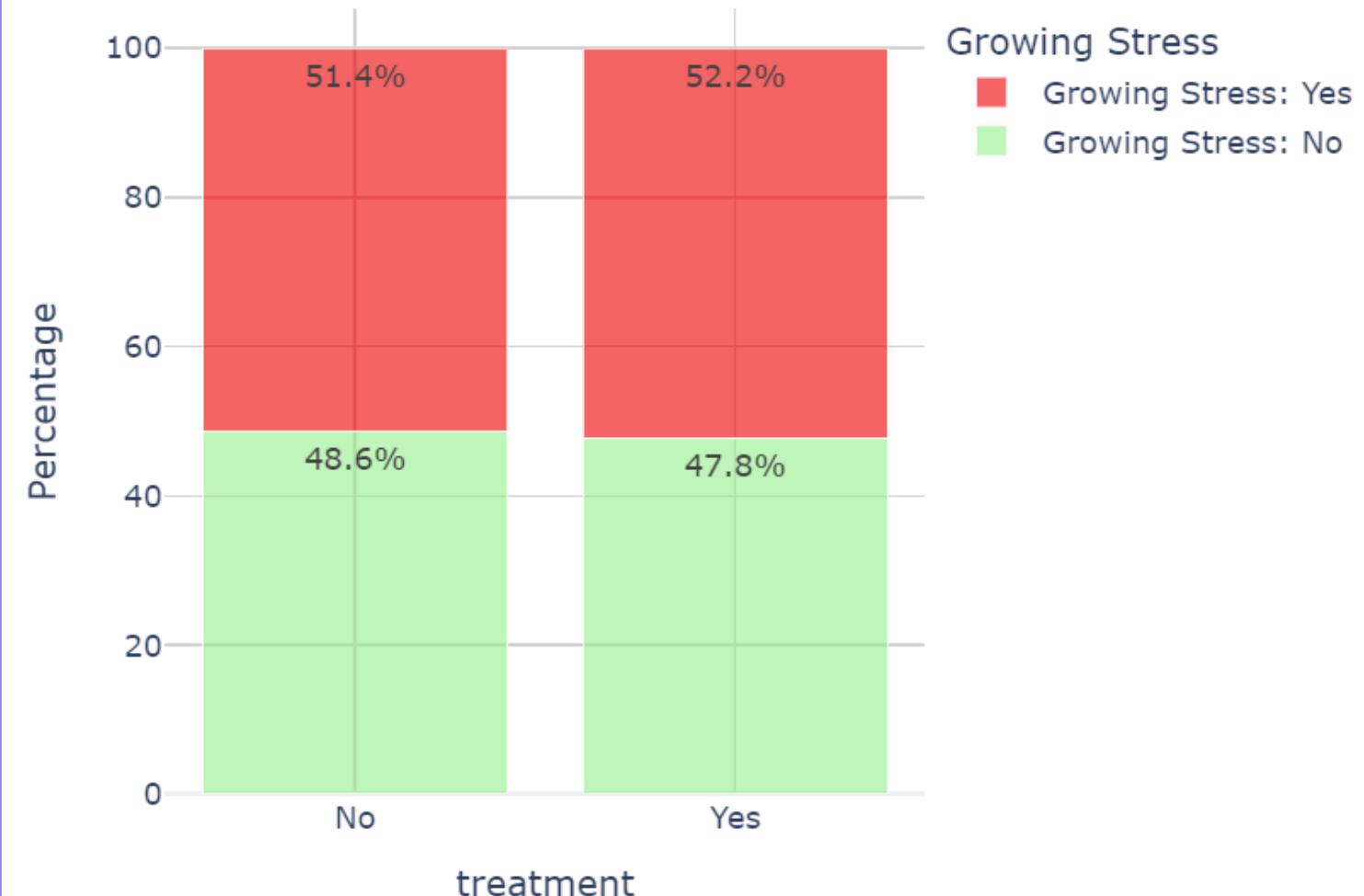
Actual	Predicted	
	0	1
0	1164	2063
1	839	3469



# Conclusion

- The top 5 factors affecting growing stress across all occupations are **Mental Health History, Days Indoors, Work Interest, Mood Swings, and Changes in Habits.**
- For students, the top 5 factors are **Changes in Habits, Days Indoors, Gender, Mood Swings, and Social Weakness.**
- Individuals with a **family history of mental health issues** might be more likely to **acknowledge their symptoms or seek treatment.**
- Across all sectors, it is better to:
  - Spend more time outdoors.
  - Acknowledge stress if there are changes in habits or frequent mood swings.
  - Growing stress is negatively related to treatment, highlighting the importance of seeking treatment when experiencing such symptoms.

Growing Stress Distribution by treatment



# Thank You

# Appendix. Dataset Structure

Timestamp: ['8/27/2014 11:37' , '8/27/2014 11:43' , '8/27/2014 11:49' , ... , '2/1/2016 23:04']

Gender: ['Female', 'Male']

Country: ['United States' , 'Poland' , 'Australia' , 'Canada' , 'United Kingdom' , 'South Africa' , ... , 'Philippines']

Occupation: ['Corporate' , 'Student' , 'Business' , 'Housewife' , 'Others']

self\_employed: ['No' , 'Yes']

family\_history: ['Yes' , 'No']

treatment: ['Yes' , 'No']

Days\_Indoors: ['1-14 days' , 'Go out Every day' , 'More than 2 months' , '15-30 days' , '31-60 days']

Growing\_Stress: ['Yes' , 'No' , 'Maybe']

Changes\_Habits: ['No' , 'Yes' , 'Maybe']

Mental\_Health\_History: ['Yes' , 'No' , 'Maybe']

Mood\_Swings: ['Medium' , 'Low' , 'High']

Coping\_Struggles: ['No' , 'Yes']

Work\_Interest: ['No' , 'Maybe' , 'Yes']

Social\_Weakness: ['Yes' , 'No' , 'Maybe']

mental\_health\_interview: ['Maybe' , 'No' , 'Yes']

care\_options: ['Yes' , 'Not sure' , 'No']