Obsessive-Compulsive Disorder Patient's current illness stage Prediction using Machine Learning Algorithms

By
Fabiha Fatin Borsha
ID:163-15- 8265
AND

Afia Anjum
ID: 171-15-8882

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. Riazur Rahman
Sr. Lecturer, Department of CSE
Daffodil International University

Co-Supervised By

Zerin Nasrin Tumpa
Lecturer,Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
DECEMBER 2019

APPROVAL

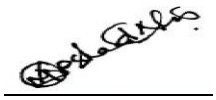
This Project titled "**Obsessive-Compulsive Disorder Patient's current illness stage Prediction using Machine Learning Algorithms** ", submitted by ***Fabiha Fatin Borsha*** and ***Afia Anjum*** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 28.01.2021.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Tarek Habib — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Saiful Islam — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

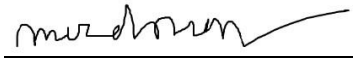


Dr. Md Arshad Ali **External Examiner**
Associate Professor
Department of Computer Science and Engineering
Hajee Mohammod Danesh Scinece and Technology University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Riazur Rahman, Sr. Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Mr. Riazur Rahman
Sr. Lecturer Department of CSE
Daffodil International University

Co-Supervised by:

Zerin Nasrin Tumpa
Lecturer, Department of CSE
Daffodil International University

Submitted by:

Fabiha Fatin Borsha
ID:163-15-8265
Department Of CSE
Daffodil International University

Afia Anjum
ID:171-15-8882
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to **Mr. Riazur Rahman Sr. Lecturer,**Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Obsessive-Compulsive Disorder Patient's current illness stage Prediction using Machine Learning Algorithms* " to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to -----, -----, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Nowadays mental health problems are increasing day by day. Among many mental health issues obsessive compulsive disorder is very common in this modern era. According to research out of 100 people 3 people have obsessive compulsive disorder. Though medical science has become so advanced but still today we are not that conscious about our mental health issues. People think that obsessive compulsive disorder is only about cleaning habits and it will be ok after a certain time. Obsessive compulsive disorder is not only a cleaning habit and it's about a lot of mental issues that a normal person who doesn't have obsessive compulsive disorder cant even relate to. People having obsessive compulsive disorder also don't know which level of their disease they are currently. As like other physical health diseases mental health disease also has some state of condition like (primary, mid, higher or extreme). "Prediction of current status of obsessive compulsive disorder " system based on predictive modeling predicts the disease of the user on the basis of the symptoms that the user provides as an input to the system. The system analyzes the symptoms provided by the user as input and gives the probability of the current status of obsessive compulsive disorder as an output. "prediction of current status of obsessive compulsive disorder " is done by implementing many techniques such as Naïve Bayes, KNN, Decision Tree, Linear Regression and Random Forest Algorithms etc . These techniques calculate the probability current status of the disease. Therefore, average prediction accuracy probability 53% is obtained.

TABLE OF CONTENTS

CHAPTER 1

Introduction

1.1 Introduction

Obsessive Compulsive Disorder is a mental health problem where negative thoughts push the patient do different rituals again and again. We often diagnosis either the patient has Obsessive-compulsive disorder (OCD) or not. We don't give much importance and time in diagnosis the current status of patient's Obsessive-compulsive disorder (OCD). As a result we go for wrong procedure of treatment. We have created a system called "prediction of current status of obsessive compulsive disorder " . We have use Machine learning algorithms and created a model for our system. User will give input their own symptoms and our system will tell them in which state of Obsessive-compulsive disorder (OCD) they are currently suffering now. So that they can know whether they should take only professional counselling or medication or both.

1.2 Motivation

In mental health treatment medication is not only treatment. Professional counseling can also recover the disease. Again sometimes people have to go under medication. Some patients need to take both counseling and medication. But the problem is when it comes to mental health issues people don't have enough knowledge which treatment they should go for. To know that it is very important to know which state of illness you are now. Like if they go to a doctor only to prescribe medicines and if they go to a psychologist they only give counseling. Nobody tells you correctly that you are now in this condition and you have to take medicine now or only counseling. If a patient is in his or her primary level only counseling can recover that. When they are in mid or higher level then they have to be under medication and also counseling. If the patient is at an extreme level then it's too serious. Then he or she needs medication, counseling and also huge family support as they become suicidal also at that time. So our motivation for this project is to make aware patients and their families that in which condition of Obsessive-compulsive disorder (OCD) they're suffering. So that they can understand properly what they should go for.

1.3 Objective:

1.To predict the current illness stage of an obsessive compulsive disorder patient .

2.To predict on the basis of previous patient’s data by using machine learning algorithms.

1.4 Literature Review

yaan,, ,have proposed a method that uses least squares support vector machine (LS SVM) to predict the suicidal rate of obsessive compulsive disorder patients. Jack, et al have done research work involving five hundred and thirty-three patients who had suffered from obsessive compulsive disorder and they were integrated in the analysis of probabilities. They performed classical statistical analysis and data mining analysis using mostly Bayesian networks. Parthan ,et al. have proposed an AI model for obsessive compulsive disorder patients' situation in this pandemic by using data mining algorithms.

1.5 What is Obsessive compulsive disorder ?

It’s a mental health disorder where a patient's brain is stuck with intrusive thoughts and he or she performs different rituals as compulsion.

1.6 Types of Obsessive Compulsive Disorder (OCD)

1. Contamination
2. ROCD
3. HOCD
4. POCD
5. Checking
6. Ordering
7. Hoarding

CHAPTER 2

RELATED WORKS ON CURRENT OBSESSIVE COMPULSIVE DISORDER STATE USING MACHINE LEARNING ALGORITHMS

2.1 About obsessive compulsive disorder:

This is a mental health condition where a patient's brain is stuck to do the same thing multiple times and it relates some negative thinking with that multiple task.

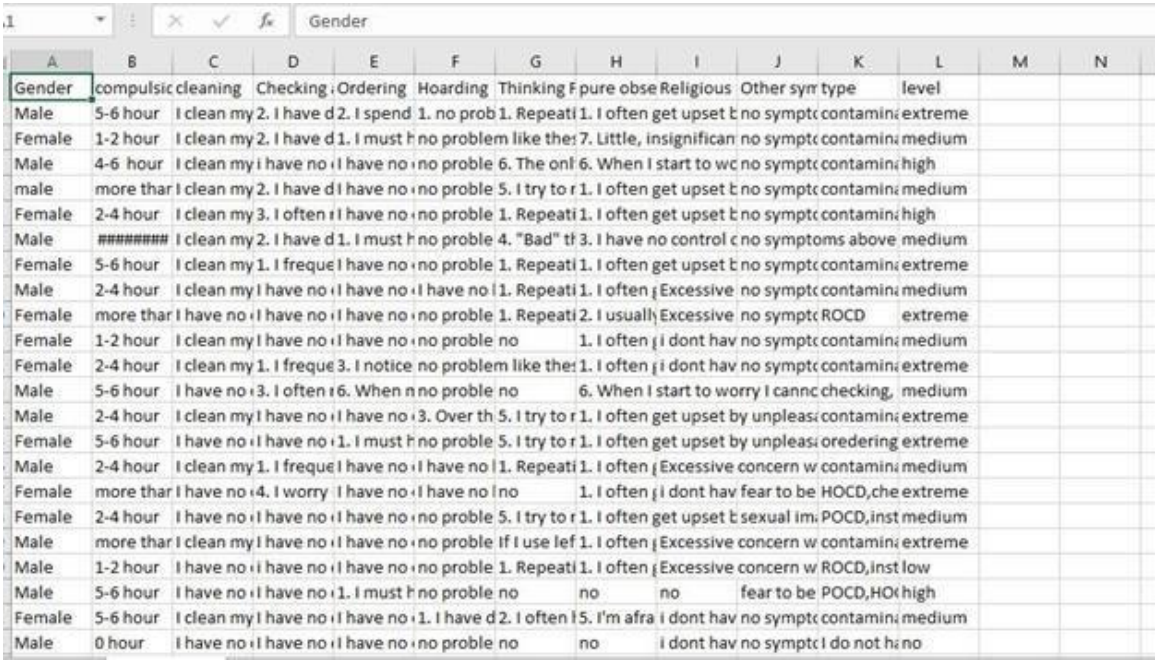
2.2 Project proposal:

Here we will use different Machine Learning Algorithms to predict the current obsessive compulsive disorder (OCD) state (primary, mid, high) of the patient by analyzing previous datasets we have and will create the model by which will give the higher accuracy.

2.3 Data set structure:

We have collected 200 obsessive compulsive disorder (OCD) patients data by using a google form. There we have used different attributes like gender , compulsion_Time, cleaning, checking , hoarding, type , level and so on. To show the output we only use one column of level .

Sample data set:



Gender	compulsic	cleaning	Checking	Ordering	Hoarding	Thinking F	pure obse	Religious	Other sym	type	level
Male	5-6 hour	I clean my	2. I have d	2. I spend	1. no prob	1. Repeati	1. I often	get upset b	no sympto	contamina	extreme
Female	1-2 hour	I clean my	2. I have d	1. I must h	no problem	like thes	7. Little,	insignifican	no sympto	contamina	medium
Male	4-6 hour	I clean my	i have no	I have no	no proble	6. The onl	6. When I	start to wo	no sympto	contamina	high
male	more thar	I clean my	2. I have d	I have no	no proble	5. I try to r	1. I often	get upset b	no sympto	contamina	medium
Female	2-4 hour	I clean my	3. I often	I have no	no proble	1. Repeati	1. I often	get upset b	no sympto	contamina	high
Male	########	I clean my	2. I have d	1. I must h	no proble	4. "Bad" th	3. I have	no control	no symptoms above		medium
Female	5-6 hour	I clean my	1. I freque	I have no	no proble	1. Repeati	1. I often	get upset b	no sympto	contamina	extreme
Male	2-4 hour	I clean my	I have no	I have no	I have no	1. Repeati	1. I often	Excessive	no sympto	contamina	medium
Female	more thar	I have no	I have no	I have no	no proble	1. Repeati	2. I usuall	Excessive	no sympto	ROCD	extreme
Female	1-2 hour	I clean my	I have no	I have no	no proble	no		1. I often	i dont hav	no sympto	contamina	medium
Female	2-4 hour	I clean my	1. I freque	3. I notice	no problem	like thes	1. I often	i dont hav	no sympto	contamina	extreme
Male	5-6 hour	I have no	3. I often	6. When n	no proble	no		6. When I	start to worry	I cannc	checking,	medium
Male	2-4 hour	I clean my	I have no	I have no	3. Over th	5. I try to r	1. I often	get upset by	unpleas	contamina	extreme
Female	5-6 hour	I have no	I have no	1. I must h	no proble	5. I try to r	1. I often	get upset by	unpleas	oredering	extreme
Male	2-4 hour	I clean my	1. I freque	I have no	I have no	1. Repeati	1. I often	Excessive	concern w	contamina	medium
Female	more thar	I have no	4. I worry	I have no	I have no	no		1. I often	i dont hav	fear to be	HOCD,che	extreme
Female	2-4 hour	I have no	I have no	I have no	no proble	5. I try to r	1. I often	get upset b	sexual im	POCD,inst	medium
Male	more thar	I clean my	I have no	I have no	no proble	If I use lef	1. I often	Excessive	concern w	contamina	extreme
Male	1-2 hour	I have no	i have no	I have no	no proble	1. Repeati	1. I often	Excessive	concern w	ROCD,inst	low
Male	5-6 hour	I have no	I have no	1. I must h	no proble	no		no	no	fear to be	POCD,HO	high
Female	5-6 hour	I clean my	I have no	1. I have d	2. I often	5. I'm afra	i dont hav	no sympto	contamina	medium	
Male	0 hour	I have no	I have no	I have no	no proble	no		no	i dont hav	no sympto	I do not h	no

2.4 packages

```
# import necessary libraries
import pandas as pd
import numpy as np

# import viusualization libraries
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# import the dataset
df = pd.read_csv('thesiswork1-1.csv',encoding='cp1252')
```

2.5 Attributes:

```
[3]  # check the names of the column
     df.columns
```

```
Index(['Gender', 'compulsion_time', 'cleaning', 'Checking and Repeating',
       'Ordering', 'Hoarding', 'Thinking Rituals',
       'pure obsession and thinking                    ',
       'Religious issues', 'Other symptoms', 'type', 'level'],
      dtype='object')
```

Data Preprocessing and Visualization

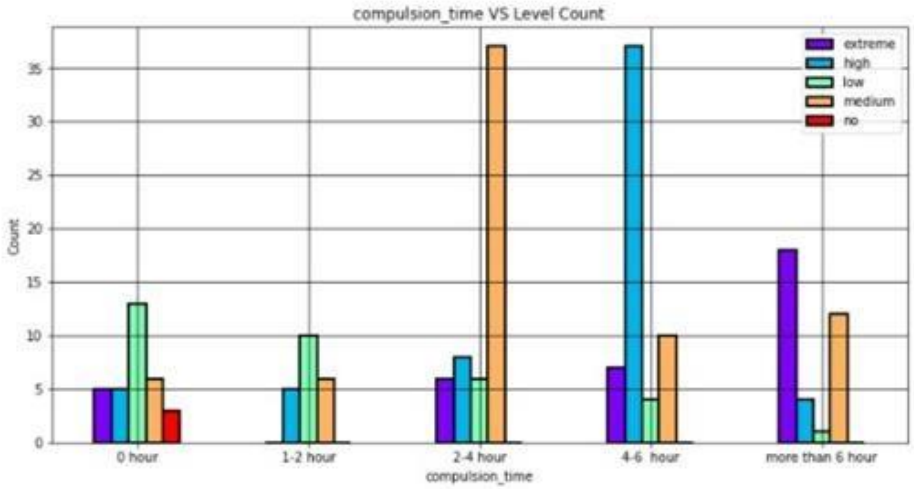
```
[4]  df.ffill(axis = 0, inplace = True)
     df.bfill(axis = 0, inplace = True)
```

```
[5]  df["Gender"].replace({"male": "Male", "Prefer not to say":"Male"}, inplace=True
```

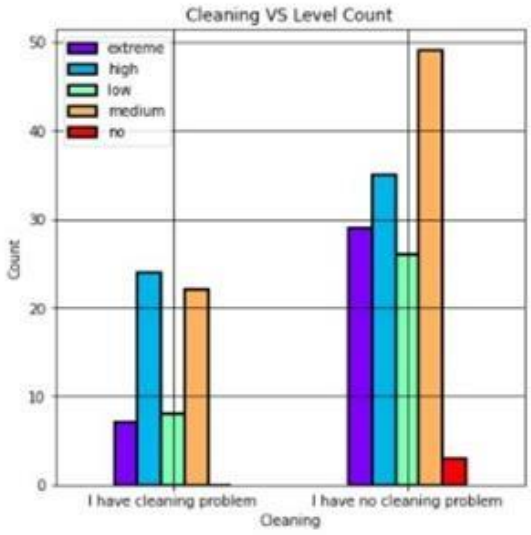
```
[6]  gender = pd.crosstab(df['Gender'], df['level'])
     gender
```

level	extreme	high	low	medium	no
Gender					
Female	17	23	16	25	1
Male	19	36	18	46	2

```
compulsion_time.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                    figsize = (12,6), linewidth = 2)
plt.ylabel('Count')
plt.xlabel("compulsion_time")
plt.xticks(rotation = 0)
plt.title('compulsion_time VS Level Count')
plt.legend(loc='best')
plt.grid(color = "black")
plt.show()
```

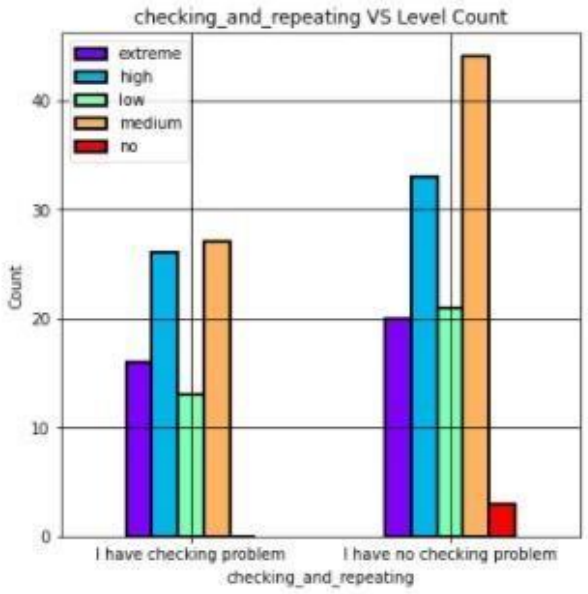



```
Cleaning.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                  figsize = (6,6), linewidth = 2)
plt.ylabel('Count')
plt.xlabel("Cleaning")
plt.xticks(rotation = 0)
plt.title('Cleaning VS Level Count')
plt.legend(loc='best')
plt.grid(color = "black")
plt.show()
```



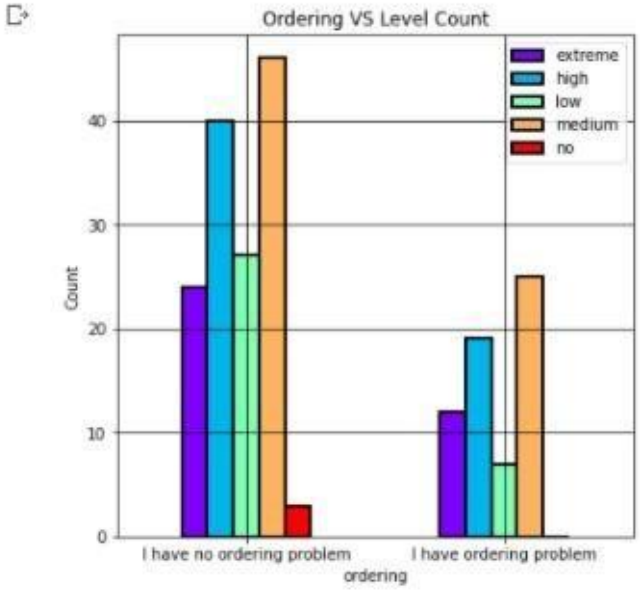
```
checking_and_repeating.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                        figsize = (6,6), linewidth = 2)

plt.ylabel('Count')
plt.xlabel("checking_and_repeating")
plt.xticks(rotation = 0)
plt.title('checking_and_repeating VS Level Count')
plt.legend(loc='best')
plt.grid(color = "black")
plt.show()
```

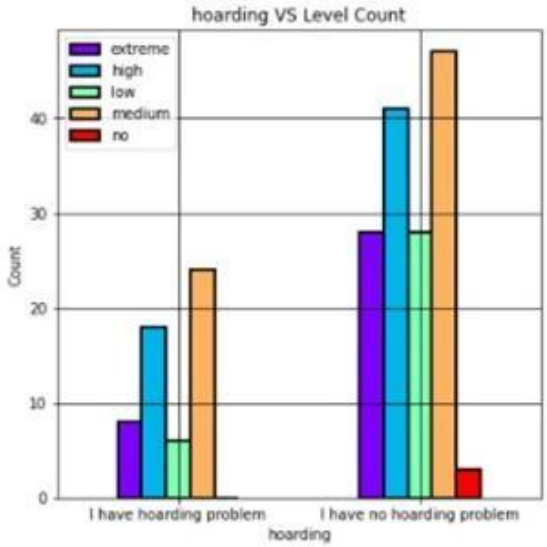


+ Code + Text

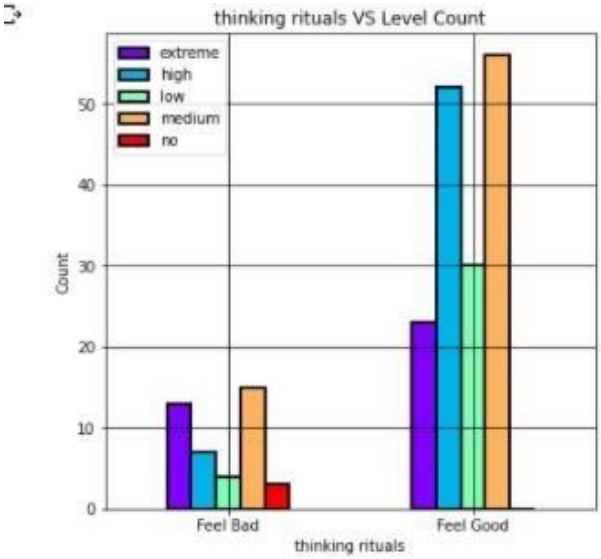
```
ordering.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                    figsize = (6,6), linewidth = 2)
plt.ylabel('Count')
plt.xlabel("ordering")
plt.xticks(rotation = 0)
plt.title('Ordering VS Level Count')
plt.legend(loc='best')
plt.grid(color = "black")
plt.show()
```



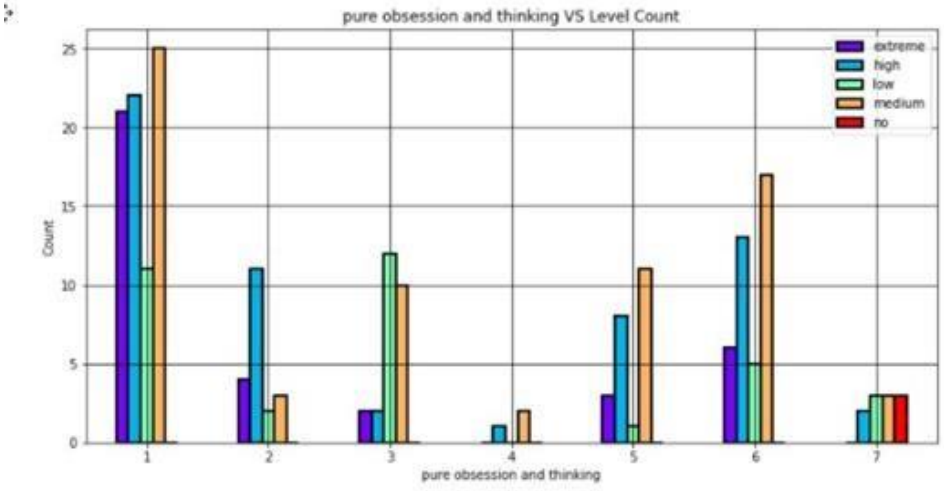
```
[33] hoarding.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                    figsize = (6,6), linewidth = 2)
     plt.ylabel('Count')
     plt.xlabel("hoarding")
     plt.xticks(rotation = 0)
     plt.title('hoarding VS Level Count')
     plt.legend(loc='best')
     plt.grid(color = "black")
     plt.show()
```



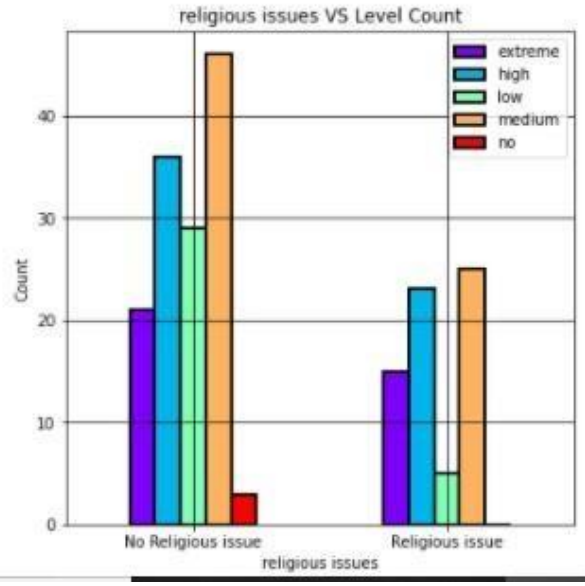
```
thinking_rituals.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                    figsize = (6,6), linewidth = 2)
plt.ylabel('Count')
plt.xlabel("thinking rituals")
plt.xticks(rotation = 0)
plt.title('thinking rituals VS Level Count')
plt.legend(loc='best')
plt.grid(color = "black")
plt.show()
```



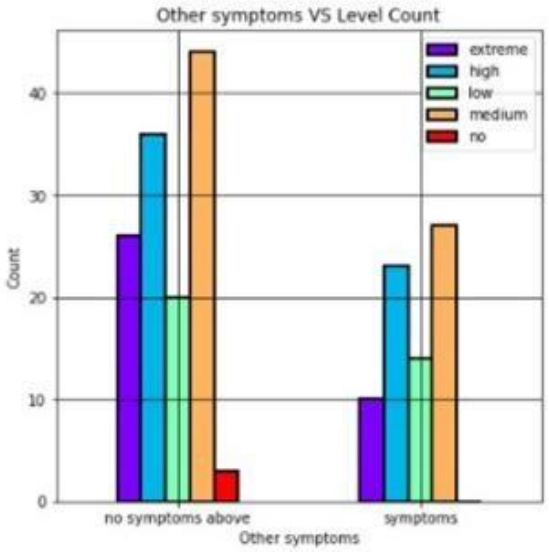

```
POAT.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                    figsize = (12,6), linewidth = 2)
plt.ylabel('Count')
plt.xlabel("pure obsession and thinking")
plt.xticks(rotation = 0)
plt.title('pure obsession and thinking VS Level Count')
plt.legend(loc='best')
plt.grid(color = "black")
plt.show()
```



```
[48] RI.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                        figsize = (6,6), linewidth = 2)
     plt.ylabel('Count')
     plt.xlabel("religious issues")
     plt.xticks(rotation = 0)
     plt.title('religious issues VS Level Count')
     plt.legend(loc='best')
     plt.grid(color = "black")
     plt.show()
```



```
[51] Other_symptoms.plot(kind='bar', cmap='rainbow',  edgecolor = "black",
                                                          figsize = (6,6), linewidth = 2)
     plt.ylabel('Count')
     plt.xlabel("Other symptoms")
     plt.xticks(rotation = 0)
     plt.title('Other symptoms VS Level Count')
     plt.legend(loc='best')
     plt.grid(color = "black")
     plt.show()
```




```
df.head()
```

	Gender	compulsion_time	Other symptoms	level	Cleaning	checking and repeating	ordering	hoarding	thinking rituals	pure obsession and thinking	religious issues
0	Male	4-6 hour	no symptoms above	extreme	I have no cleaning problem	1. I frequently have to check things over and ...	I have ordering problem	I have no hoarding problem	Feel Good	1	Religious issue
1	Female	1-2 hour	no symptoms above	medium	I have no cleaning problem	1. I frequently have to check things over and ...	I have ordering problem	I have no hoarding problem	Feel Good	7	Religious issue
2	Male	4-6 hour	no symptoms above	high	I have no cleaning problem	I have no checking problem	I have no ordering problem	I have no hoarding problem	Feel Good	6	Religious issue
3	Male	more than 6 hour	no symptoms above	medium	I have no cleaning problem	1. I frequently have to check things over and ...	I have no ordering problem	I have no hoarding problem	Feel Bad	1	Religious issue
4	Female	2-4 hour	no symptoms above	high	I have no cleaning problem	1. I frequently have to check things over and ...	I have no ordering problem	I have no hoarding problem	Feel Good	1	Religious issue

```
[54] df=df.iloc[10:]

[55] test=df.iloc[0:10]

[56] df.shape

     (193, 11)

[57] test.shape

     (10, 11)
```

▾ Machine Learning

Modelling and predicting with Machine Learning

Here we our main achievement will to predict an OCD patient's current illness stage. In this project at first we have collected data . we used that data as train data and test data. After that we applied some AI algorithms. After applying many algorithms we got highest accuracy rate by applying logistic regression. So that we have used logistic regression for our model. By
applying logistic regression we have predict the current stage (primary, mid, high, extreme) of OCD patient.

Model performance checking : All top model performance of data

```
[60] best_model = compare_models(fold=5)
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lr	Logistic Regression	0.5185	0.2763	0.4765	0.5394	0.5045	0.3515	0.3620	0.604
lda	Linear Discriminant Analysis	0.5111	0.2870	0.5012	0.5251	0.5045	0.3502	0.3587	0.026
ridge	Ridge Classifier	0.5037	0.0000	0.4262	0.5032	0.4909	0.3387	0.3470	0.022
rf	Random Forest Classifier	0.4815	0.2982	0.5063	0.5300	0.4751	0.2999	0.3091	0.544
knn	K Neighbors Classifier	0.4741	0.2754	0.4329	0.5101	0.4708	0.2997	0.3094	0.138
gbc	Gradient Boosting Classifier	0.4667	0.2861	0.5085	0.4835	0.4510	0.2869	0.2974	0.484
nb	Naive Bayes	0.4593	0.2772	0.4944	0.4854	0.4521	0.2781	0.2871	0.020
lightgbm	Light Gradient Boosting Machine	0.4444	0.2738	0.4837	0.4596	0.4359	0.2558	0.2609	0.086
xgboost	Extreme Gradient Boosting	0.4370	0.2942	0.4683	0.4352	0.4118	0.2423	0.2505	0.486
et	Extra Trees Classifier	0.4296	0.2852	0.4674	0.4489	0.4166	0.2322	0.2408	0.564
dt	Decision Tree Classifier	0.4148	0.2622	0.4477	0.4841	0.4045	0.2252	0.2334	0.020
catboost	CatBoost Classifier	0.4148	0.2808	0.4465	0.4569	0.4093	0.2105	0.2164	7.204
svm	SVM - Linear Kernel	0.3778	0.0000	0.4278	0.4309	0.3644	0.1730	0.1805	0.076
ada	Ada Boost Classifier	0.3778	0.2411	0.4148	0.3787	0.3304	0.1706	0.1965	0.106
qda	Quadratic Discriminant Analysis	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.022

Classification Algorithms

We have used different classifications algorithm in our project. These algorithms help to predict data into different class.

a) Decision Tree In decision tree algorithm there are some nodes and branch. These nodes works as input and those branch works as output.

b) Random Forest we can consider random forest algorithm as a larger version of decision tree. In this algorithm the average result of decision tree are calculated so that the result is better. .

c) Support Vector Machines We come to know a word that's hyperplane in this algorithm. This algorithm use hyperplane to classify different data.

d) Linear regression It uses consistent variables. Among all regression this is the simpliest to implement . With the help of this we can only do work in linear relationships.

e) Logistic Regression in logistic regression we work with many variables and predict our result in a categorical format.

f) K-Means Clustering this algorithm mainly works with data clustering. It combine same type of data and make cluster.

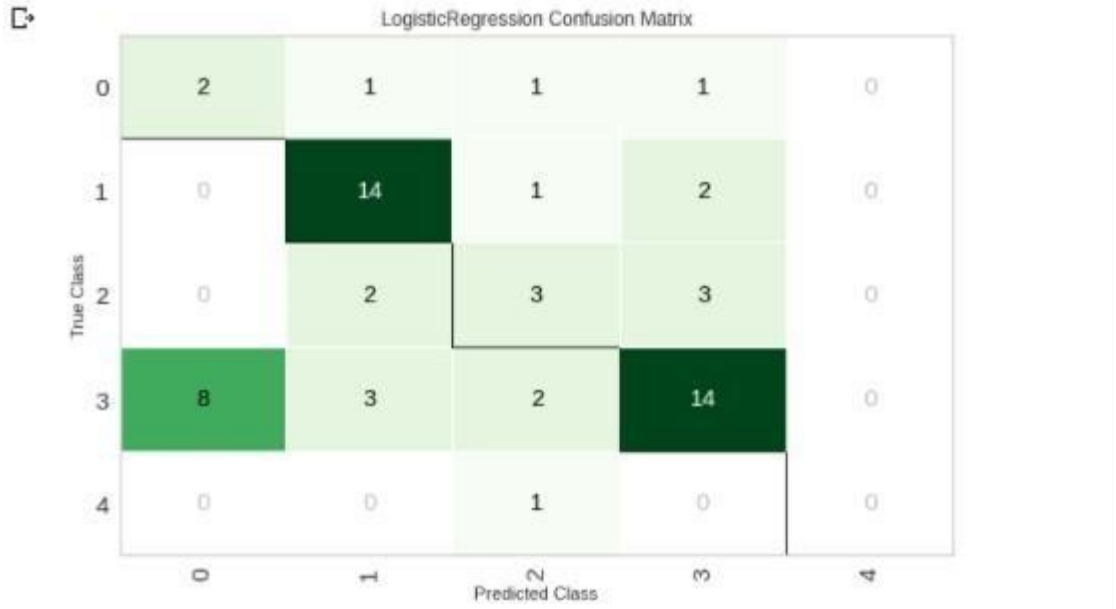
2.7 models:

We have choose (LR : Logistic regression) and train it with 15 folds again because it perform better in all : where mean accuracy was : 0.77

```
lr = create_model('lr', fold = 15)
```

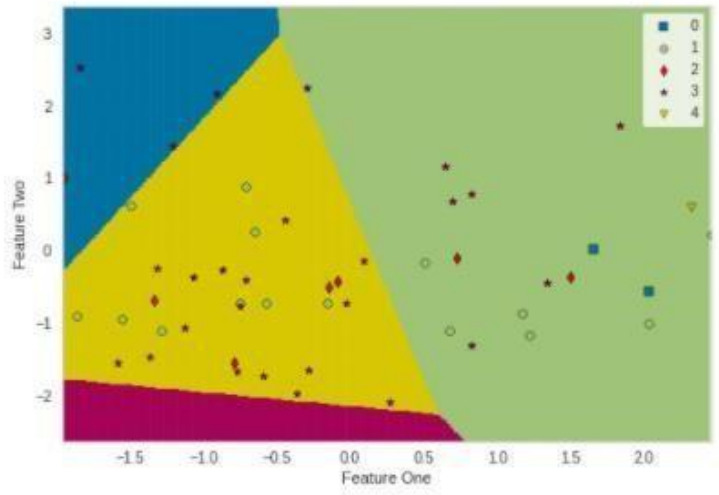
	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.5556	0.0000	0.6250	0.6111	0.5608	0.3684	0.3752
1	0.6667	0.0000	0.7083	0.7407	0.6889	0.5424	0.5517
2	0.6667	0.0000	0.6250	0.7556	0.6574	0.5345	0.5660
3	0.3333	0.0000	0.2917	0.2111	0.2540	0.0847	0.1021
4	0.2222	0.0000	0.2083	0.2222	0.2222	-0.0500	-0.0509
5	0.7778	0.0000	0.7500	0.8889	0.7778	0.7000	0.7379
6	0.4444	0.0000	0.4583	0.4630	0.4444	0.2623	0.2667
7	0.6667	0.0000	0.7083	0.7222	0.6519	0.5574	0.5866
8	0.4444	0.0000	0.4167	0.5000	0.4667	0.2500	0.2543
9	0.4444	0.0000	0.5417	0.5370	0.4481	0.2623	0.2759
10	0.5556	0.0000	0.5417	0.6296	0.5778	0.4098	0.4167
11	0.5556	0.0000	0.5417	0.4815	0.5111	0.4000	0.4068
12	0.4444	0.0000	0.4583	0.3704	0.4000	0.2500	0.2543
13	0.4444	0.7989	0.5000	0.3981	0.3704	0.3182	0.3629
14	0.4444	0.7897	0.4667	0.6574	0.4481	0.3182	0.3629
Mean	0.5111	0.1059	0.5228	0.5459	0.4986	0.3472	0.3646
SD	0.1394	0.2700	0.1456	0.1892	0.1516	0.1845	0.1919

```
plot_model(lr, plot = 'confusion_matrix')
```



This is the figure of our confusion matrix

```
[63] plot_model(lr, plot = 'boundary')
```



2.8 Here is the boundary plot of our model

```
save_model(lr, model_name='best-model')
```

```
Transformation Pipeline and Model Succesfully Saved
(Pipeline(memory=None,
          steps=[('dtypes',
                  DataTypes_Auto_infer(categorical_features=[],
                                       display_types=True, features_todrop=[],
                                       id_columns=[],
                                       ml_usecase='classification',
                                       numerical_features=[], target='level',
                                       time_features=[])),
                 ('imputer',
                  Simple_Imputer(categorical_strategy='not_available',
                                 fill_value_categorical=None,
                                 fill_value_numerical=None,
                                 numeric_strate...
                                                        random_state=123,
                                                        target='level',
                                                        variance_retained_or_number_of_components=None)),
                 ['trained_model',
                  LogisticRegression(C=1.0, class_weight=None, dual=False,
                                     fit_intercept=True, intercept_scaling=1,
                                     l1_ratio=None, max_iter=1000,
                                     multi_class='auto', n_jobs=None,
                                     penalty='l2', random_state=123,
                                     solver='lbfgs', tol=0.0001, verbose=0,
                                     warm_start=False)]],
          verbose=False), 'best-model.pkl')
```

```
# loading Model
loaded_bestmodel = load_model('best-model')
print(loaded_bestmodel)
```

```
Transformation Pipeline and Model Successfully Loaded
Pipeline(memory=None,
         steps=[('dtypes',
                 DataTypes_Auto_infer(categorical_features=[],
                                      display_types=True, features_todrop=[],
                                      id_columns=[],
                                      ml_usecase='classification',
                                      numerical_features=[], target='level',
                                      time_features=[])),
                ('imputer',
                 Simple_Imputer(categorical_strategy='not_available',
                                fill_value_categorical=None,
                                fill_value_numerical=None,
                                numeric_strate...
                                                       random_state=123,
                                                       target='level',
                                                       variance_retained_or_number_of_components=None)),
                ['trained_model',
                 LogisticRegression(C=1.0, class_weight=None, dual=False,
                                    fit_intercept=True, intercept_scaling=1,
                                    l1_ratio=None, max_iter=1000,
                                    multi_class='auto', n_jobs=None,
                                    penalty='l2', random_state=123,
                                    solver='lbfgs', tol=0.0001, verbose=0,
                                    warm_start=False)]],
         verbose=False)
```

2.9 This is the test part of prediction

```
# Model Prediction on TEST Data
predict_unseen = predict_model(loaded_bestmodel, test)
predict_unseen
```

	Gender	compulsion_time	Other symptoms	level	Cleaning	checking and repeating	ordering	hoarding	thinking rituals	pure obsession and thinking	religious issues	Label	Score
10	Female	2-4 hour	no symptoms above	extreme	I have no cleaning problem	1. I frequently have to check things over and ...	I have ordering problem	I have no hoarding problem	Feel Bad	1	No Religious issue	medium	0.4338
11	Male	4-6 hour	no symptoms above	medium	I have no cleaning problem	1. I frequently have to check things over and ...	I have ordering problem	I have no hoarding problem	Feel Bad	6	No Religious issue	high	0.4186
12	Male	2-4 hour	no symptoms above	extreme	I have no cleaning problem	I have no checking problem	I have no ordering problem	I have hoarding problem	Feel Bad	1	No Religious issue	extreme	0.4376
13	Female	4-6 hour	no symptoms above	extreme	I have no cleaning problem	I have no checking problem	I have ordering problem	I have no hoarding problem	Feel Bad	1	No Religious issue	extreme	0.4725
14	Male	2-4 hour	no symptoms above	medium	I have no cleaning problem	1. I frequently have to check things over and ...	I have no ordering problem	I have no hoarding problem	Feel Good	1	Religious issue	medium	0.5962
15	Female	more than 6 hour	symptoms	extreme	I have no cleaning problem	1. I frequently have to check things over and ...	I have no ordering problem	I have no hoarding problem	Feel Bad	1	No Religious issue	extreme	0.7569
16	Female	2-4 hour	symptoms	medium	I have no cleaning problem	I have no checking problem	I have no ordering problem	I have no hoarding problem	Feel Bad	1	No Religious issue	extreme	0.4504
17	Male	more than 6 hour	symptoms	extreme	I have no cleaning problem	I have no checking problem	I have no ordering problem	I have no hoarding problem	Feel Bad	1	Religious issue	extreme	0.8267
18	Male	1-2 hour	symptoms	low	I have no cleaning problem	I have no checking problem	I have no ordering problem	I have no hoarding problem	Feel Good	1	Religious issue	high	0.3942
19	Male	4-6 hour	symptoms	high	I have no cleaning problem	I have no checking problem	I have ordering problem	I have no hoarding problem	Feel Bad	7	No Religious issue	high	0.5044

CHAPTER 3

CONCLUSIONS

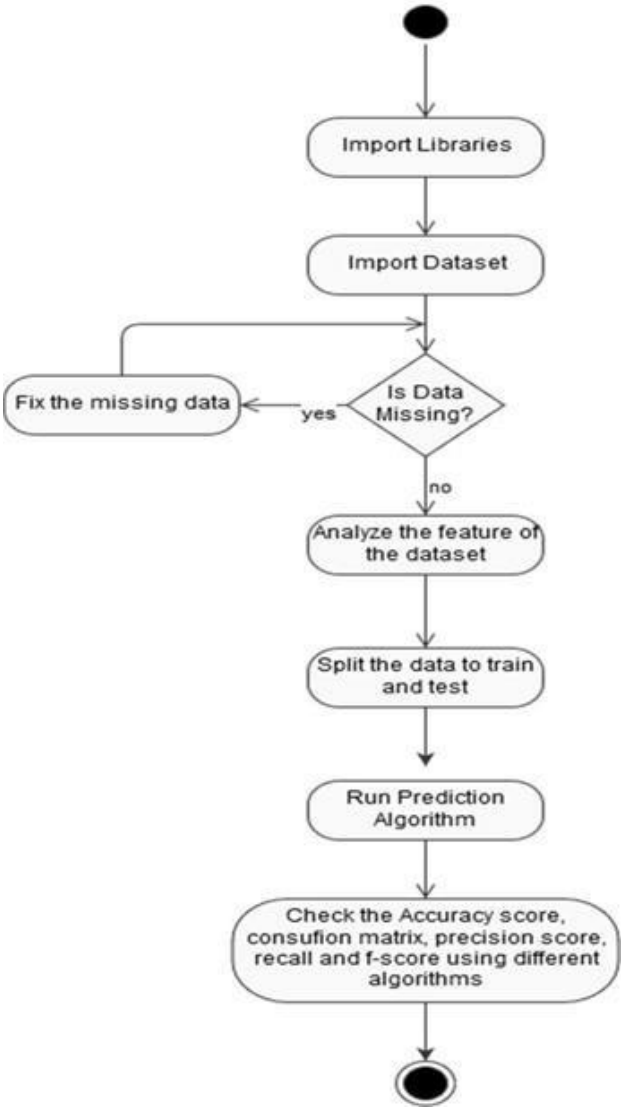
3.1 Drawbacks:

We could not achieve 100 percent accuracy rate. That is why when the system will show the result it will not be fully appropriate. There could be some unappropriate result also. That can affect a user. This project is mainly basis on mental health issues. This is very sensitive . By increasing accuracy rate almost 100 percent we can solve this limitation.

3.2 Future Scope:

We can develop the system and can add some extra features. We can make a community where there will be only ocd patients and doctors. Also there will be a psychologist. We can add the feature after predicting the patient's current level and it will be suggested that she needs counselling, medication or both. And Here will be an online intelligent phycologist and an online doctor to help the patient. They will take input of their regular activity and will analyze those data. After that patients can get good counseling based on their lifestyle and need.

Activity Diagram of our project:



Reference:

[1] Le Duffi, Cristian, “OCD with Data Mining Method,” Stud. Health Technol. Inform., vol. Vol. 105, no. 2, p. No. 2, pp. 1256–1259, 2009.

[2] W. J. F. and G. Piatetsky-Shapiro, “Knowledge Discovery in OCD: An Overview,” AI Mag., vol. Vol. 13, N, no. 3, pp. 57–70, 1996.

[3] K. Y. N. and K. H. R. Heon Lee, “OCD with AI,” Proc. Int. Conf. . Technol. Knowl. Discov. Data Min., p. pp. 56–66, 2008.

[4] B. J. L. and K. PR, Lee, Ho-Sun Shon, “OCD and Life” Signal Process. Pattern., vol. 345, pp. 721–727, 2006.

[5] L. P. and R. Subramanian, “Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm,” Int. J. Biol. Biomed. Med. Sci., vol. Vol. 3, no. No.

Obsessive Compulsive Disorder Patient's current illness stage Prediction using Machine Learning Algorithms

ORIGINALITY REPORT

7%	7%	0%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	7%
2	etd.lib.nsysu.edu.tw Internet Source	1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		