**Assignment**

# **Q. For some given Dataset learn to apply various classification algorithms on your own and give a regress comparison of results obtained and conclude.**

Solution-

**Abstract**-

The dataset is about Analysis on depression condition, we will apply several classification algorithms like KNN, Logistic regression, Naive bayes on the data set and we will compare the results of certain algorithms and conclude by identifying which algorithm is giving best results and minimum error.

**Introduction-**

This data set consists of condition group for depression. It contains the following columns; number (patient identifier), days (number of days of measurements), gender (1 or 2 for female or male), age (age in age groups), afftype (1: bipolar II, 2: unipolar depressive, 3: bipolar I), melanch (1: melancholia, 2: no melancholia), inpatient (1: inpatient, 2: outpatient), edu (education grouped in years), marriage (1: married or cohabiting, 2: single), work (1: working or studying, 2: unemployed/sick leave/pension) , madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped).

We will apply certain Classification algorithms and conclude the results.

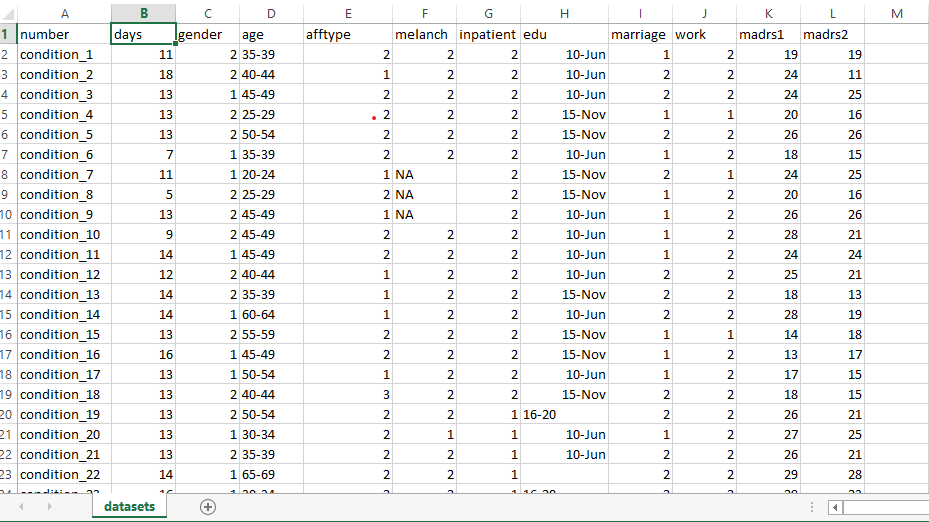
**Experiment** –

In this experiment we have trained all three algorithms on a single dataset and compared its results.

There are different types of Classification Algorithms-

1. Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable.
2. A naive Bayes classifier is an algorithm that uses Bayes' theorem to classify objects. Naive Bayes classifiers assume strong, or naive, independence between attributes of data points. Popular uses of naive Bayes classifiers include spam filters, text analysis and medical diagnosis. These classifiers are widely used for machine learning because they are simple to implement.
3. A k-nearest-neighbor is a data classification algorithm that attempts to determine what group a data point is in by looking at the data points around it. An algorithm, looking at one point on a grid, trying to determine if a point is in group A or B, looks at the states of the points that are near it.

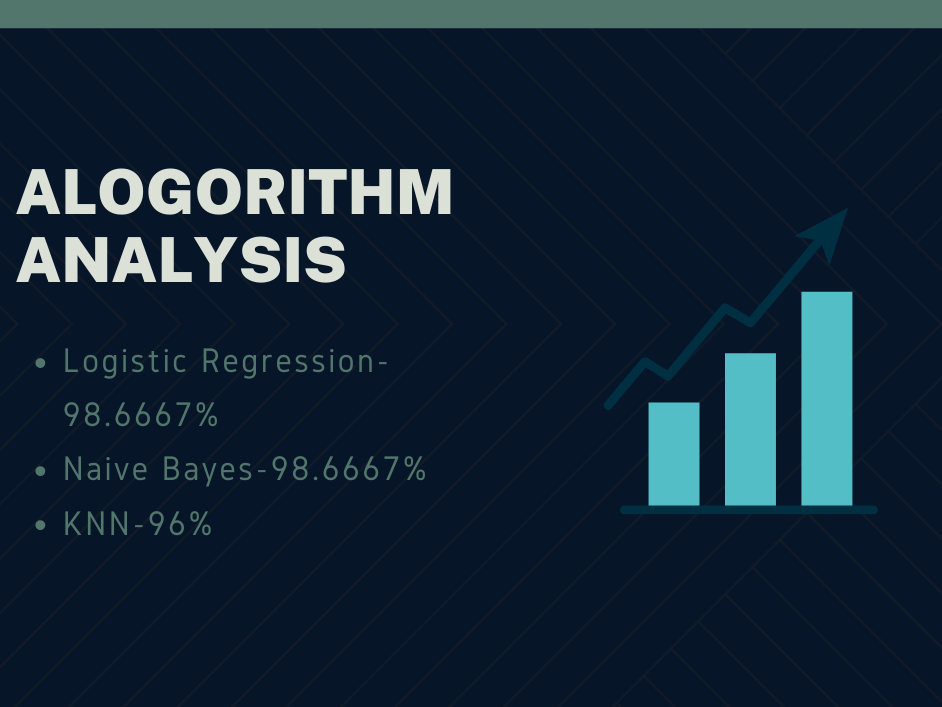
**Dataset-** It consist of depression data of students.



**Dataset Link**- <https://www.kaggle.com/arashnic/the-depression-dataset>

**Analysis Algorithm with google Colab Link**- <https://colab.research.google.com/github/jthakur29/ML-lab/blob/main/Classification_algorithms_on_depression_dataset_.ipynb>

**Results** - Accuracy of each model-



**Conclusion** – Here we can see that the logistic regression and naïve bayes both has given the highest accuracy.Therefore we can conclude that for this dataset using logistic regression classication and naïve bayes algorithm are ideal.