

# **MINOR PROJECT REPORT**

ON

## **KS FACE-e-FIER**



**PRESENTED BY :**

**KAUSHIKI TARU (18103053)**

**KUMARI SONI (18103054)**

**DR. B.R. AMBEDKAR NATIONAL INSTITUTE OF TECHNOLOGY, JALANDHAR**

**COMPUTER SCIENCE AND ENGINEERING DEPARTMENT**

## **ACKNOWLEDGEMENT**

It has indeed been a great privilege to be a part of the Department of Computer Science and Engineering, National Institute of Technology, Jalandhar.

We are extremely thankful to Professor A L Sangal, Head, Department of Computer Science and Engineering who has provided all his support.

We are highly grateful to have constant encouragement and guidance from all the faculties of the Department of Computer Science and Engineering who gave us their time and suggestions which helped us in completing our project successfully.

Thank you.

Kaushiki Taru (18103053)

Kumari Soni (18103054)

## **DECLARATION**

We, Kaushiki Taru and Kumari Soni, hereby declare that our Project titled- **KS FACE-e-FIER** being submitted by us in the Department of Computer Science and Engineering is a project work carried by us. The project has not been copied from anywhere and has been made solely by us. We will be solely responsible if any plagiarism is found.

Thank you.

Date: 27 May 2021

Kaushiki Taru (18103053)

Kumari Soni (18103054)

## **TABLE OF CONTENTS**

<b>Sr. no.</b>	<b>Topic</b>	<b>Page no.</b>
<b>0.1</b>	<b>Acknowledgement</b>	<b>1</b>
<b>0.2</b>	<b>Declaration</b>	<b>2</b>
<b>1</b>	<b>Introduction</b>	<b>4-6</b>
1.1	Problem Statement and Necessity	4
1.2	Technical and Non-Technical Feasibility	5
1.3	Future Scope	6
<b>2</b>	<b>Proposed Solution- Ice Breaking</b>	<b>6-10</b>
2.1	Identifying Stakeholders	6
2.2	Detailed Solution	7
<b>3</b>	<b>Technical Analysis</b>	<b>10-11</b>
3.2	UML Diagram	10
3.3	Tech Stack Analysis	11
<b>4</b>	<b>Result and Discussion</b>	<b>11-12</b>
4.1	Application Usage Instruction	11
4.2	Deployment and Testing Status	12
<b>5</b>	<b>Conclusion</b>	<b>12</b>
<b>6</b>	<b>References</b>	<b>14</b>

# 1 Introduction

The project- “ **KS FACE-e-FIER** ” is an Application that takes the input as the dataset link and the target variable, and provides the detailed results of various Classification Algorithms based on their accuracy, precision, recall and other parameters. Thus, is the name of the project:

FACE-e-FIER = Classifier Interface

And,

KS= Team Members: **Kaushiki** and **Soni**

We have implemented six Classification Algorithms namely - ***Information Gain, Gini Index, Naive Bayes, K-Nearest Neighbour, Random Forest and Gradient Boost***. Also, we can get the ***Best Classifier*** for any given valid dataset, based on the *Accuracy* of all the Classifiers. By valid dataset, we mean, the dataset over which Classification can be applied.

Classification is a method of categorization of data into different categories/classes. In layman language, we can say that it is a method of separation of objects or placing of objects into classes. It is a supervised learning method, i.e., the output label / container is already defined. It can be performed on both structured and unstructured data.

Classification is widely used nowadays to classify whether an Email is spam or genuine, in Sentiment Analysis, Disease Detection, and in many other areas of study.

## 1.1 Problem Statement and Necessity

In the growing era of technology and human dependency over it, we have to hunt for methods which are easy to understand and save time. With the increase in the amount of data available, the field of Data Science and Analytics, solely keeps an eye on organisation of data and extract important information out of it.

- ❖ As a beginner, the first doubt that comes to the mind is how to choose the dataset. This hurdle consumes a lot of time to resolve and sometimes even act like a demotivator.

To solve this, the “KS FACE-e-FIER” not only provides the results of a particular classification technique but also suggests Classifier which gives the best result on any dataset.

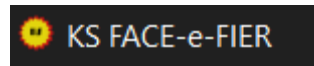
- ❖ Many times, an Analyst has a lot of information to handle at a time. In such situations, instead of trying with all the types of Classifier separately, one can save a lot of time and reduce the clumsy task by simply using this Application.

## 1.2 Technical and Non-Technical Feasibility

### ❖ Technical Feasibility

The project “KS FACE-e-FIER” is a System Application. The only technology you require is **Python**.

We have created the **.exe file (Application)**, thus, you only need to run an application named “KS Classifier Interface” to get the work done.



### ❖ Social

No such application is present which finds the best classifier result and helps users choose between the classifiers. Thus, this application would really be useful for developers at an initial stage of choosing the type of classifier.

### ❖ Economic feasibility

This project only requires Python, which most of the developers already have. Thus, this project doesn't require any cost and is economically feasible.



### 1.3 Future Scope

The project is a stepping stone towards an advanced technology in the field of Data Science and Analytics.

There are many ideas that can combine with the existing technologies to make the project “**KS FACE-e-FIER**” more useful and handier.

- ❖ We can add other machine learning techniques like Clustering, Regression, Neural Networks, Rule-based mining, etc. Further collaboration for Datasets which require feature scaling can also be included.
- ❖ Along with the Application, we can integrate it with the web and deploy like a Web Application for easy and quick accessibility.
- ❖ We can improve flexibility with adding choices for sample size. The User can also choose to include preprocessing steps and see their effects on the result directly.
- ❖ Graphical display for Data Visualization and better Result Analysis can also become an additional feature.
- ❖ The Application can also be integrated with Databases and a History for previous results can be stored. For data confidentiality, this facility will be available for only Authorised Users.

## 2 Proposed Solution- Ice Breaking

### 2.1 Identifying Stakeholders

This project would be of concern for Data Analysts, Researchers, Students, Developers, Testers and Project Managers.

## 2.2 Detailed Solution

We have tried to solve the above stated problem and the results can be seen as follows:

- ❖ The Application opens with a main window which allows us to choose a method for Classification out of Information Gain, Gini Index, Naive Bayes, K-Nearest Neighbours, Random Forest and Gradient Boost.

Also, if it allows us to find the best classifier for the input dataset using the “Know the Best Classifier” option. This main screen also displays the clock time of the system in which it is running.

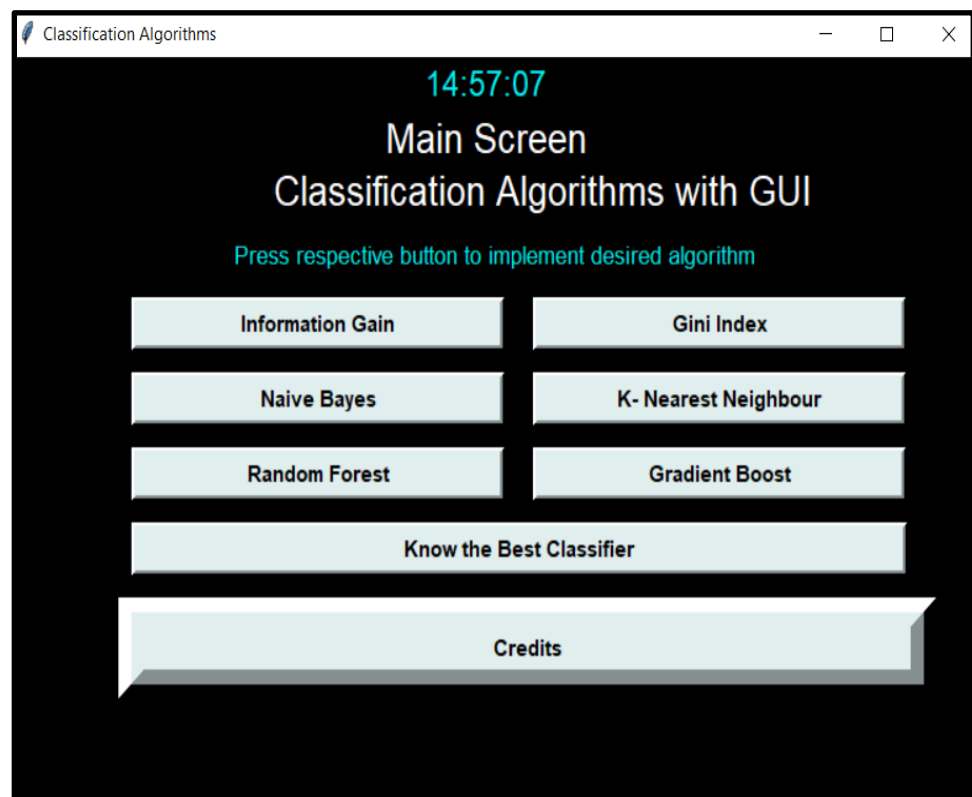


Fig 1: Main Window



- ❖ After we choose one, the application directs us to another window where a link for the dataset and the target column number (starting with 0) is given as input.

Note:

1. The link for the dataset should be in .csv format.
2. Both the inputs should be filled.

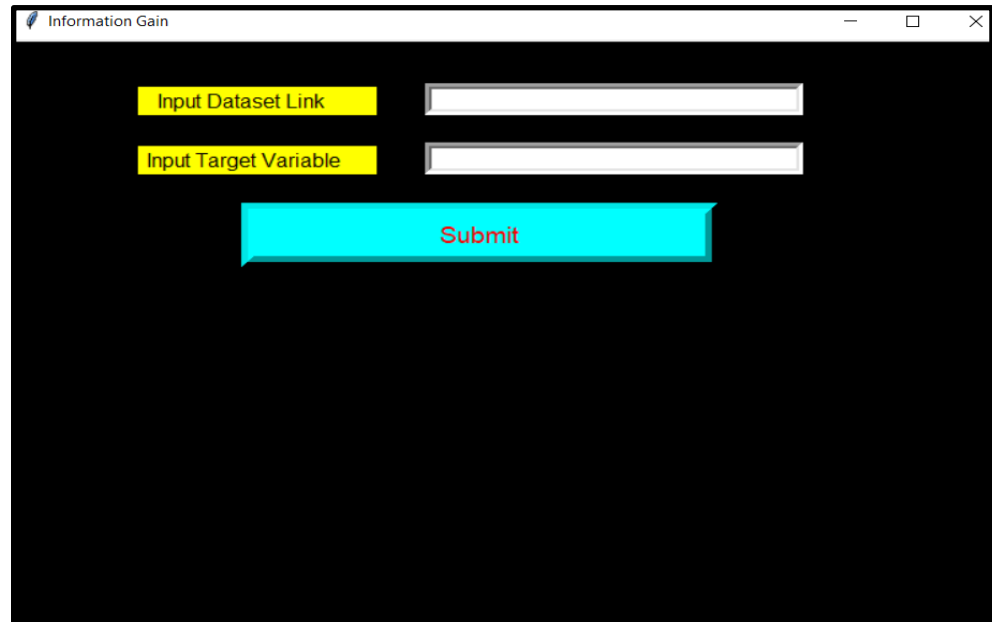
The image shows a web browser window titled "Information Gain". The background is black. There are two input fields with yellow labels: "Input Dataset Link" and "Input Target Variable". Below these is a red button with the text "Submit".

Fig 2: Input Window

- ❖ The required output for the dataset provided to each Classifier is displayed on their respective screens. It includes performance metrics like
  - accuracy
  - precision
  - recall
  - F1-score
  - confusion matrix.

One can study the result and utilize that information for further analysis.

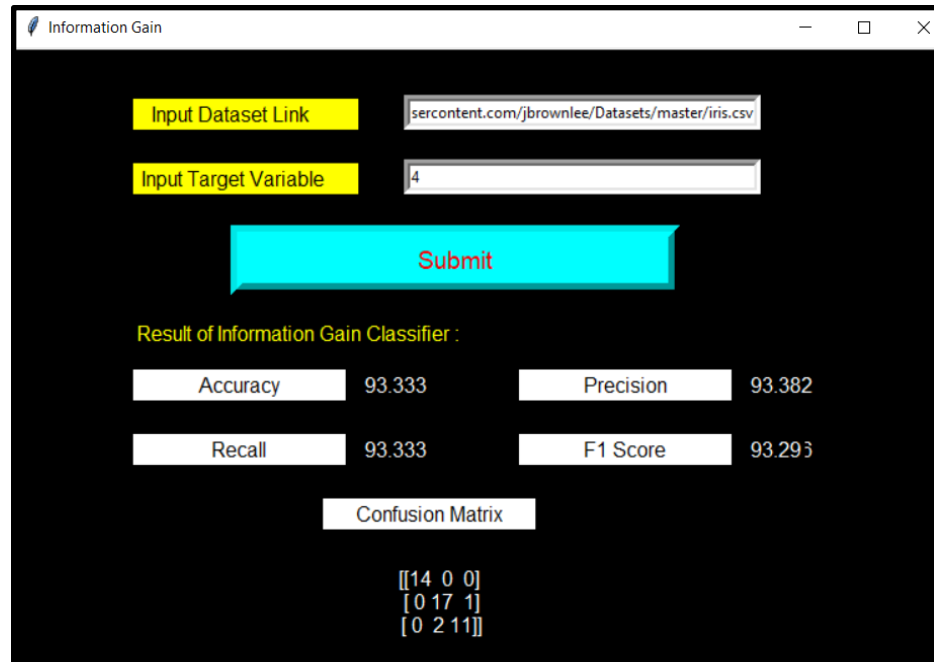


Fig 3: Result of Information Gain for [iris](#) dataset

The output of the Best Classifier includes the accuracy of all the six classifiers, and gives the name of the best classifier.

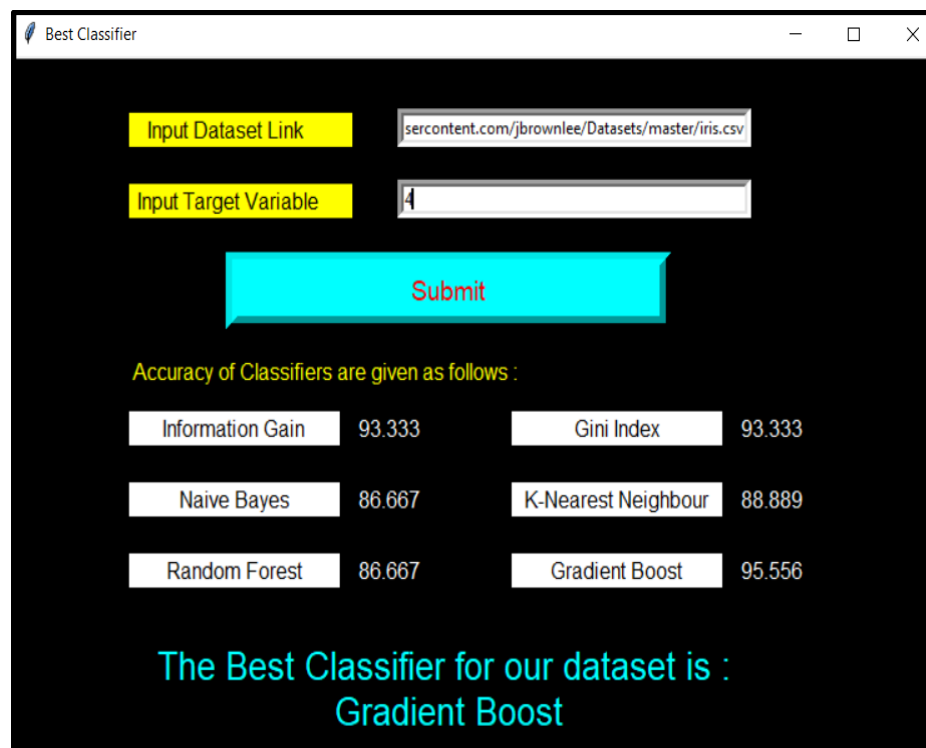


Fig 4: Result of the Best Classifier for [iris](#) dataset

- ❖ One of the buttons is dedicated for displaying the **Credit** information which includes the team members details.

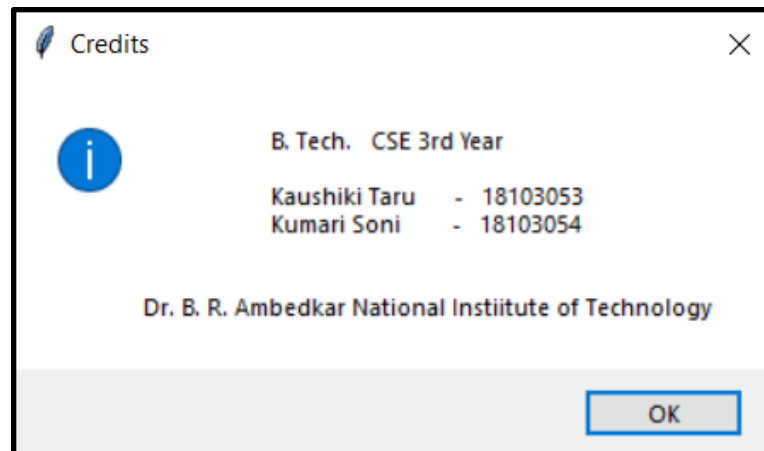


Fig 5: Credits window

### 3 Technical Analysis

#### 3.1 UML Diagram

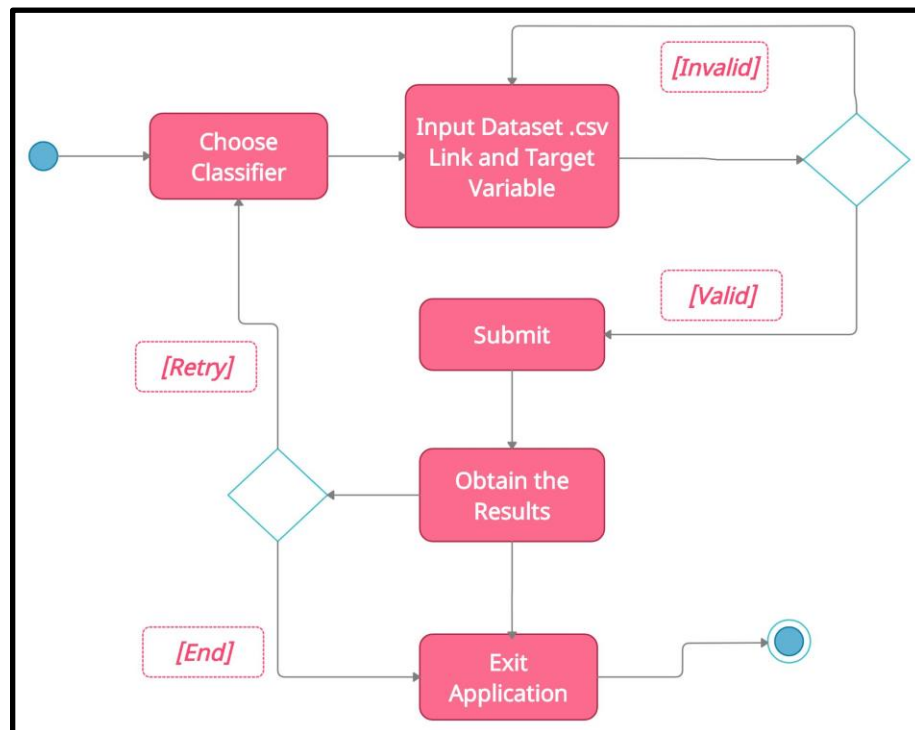


Fig 6: UML Diagram

## 3.2 Tech Stack Analysis

For implementing our idea into a working model, we have used the following technologies:

- ❖ Python: Platform for coding.
- ❖ Tkinter: Integrate python codes with GUI.
- ❖ PyInstaller: Deploy the model as an Application.

## 4 Result and Discussion

### 4.1 Application Usage Instruction

The System Application “ KS FACE-e-FIER ” is very easy to use as we have made it user-friendly.

Follow the following steps and you will get your required results in just a few seconds:

**Step 1-** Open the application named “ KS FACE-e-FIER ”.

**Step 2 -** When the application opens, you will see a main window (see Fig 1) which contains a list of classifiers and an option “Know the Best Classifier”. Click on any of those buttons according to the need.

**Step 3-** After choosing any of your choices, you will get to see an Input Screen (Fig 2) where you need to input the Dataset Link and the Target Column Number.

**Step 4-** When you are done with inputting the required fields, click on the “Submit” button and you will get the required results (similar to Fig 3,4).

## 4.2 Deployment and Testing Status

At present, the Application has already been tested with few datasets and their respective results have been recorded and verified. The first version of the Application is deployed and is ready to be used.

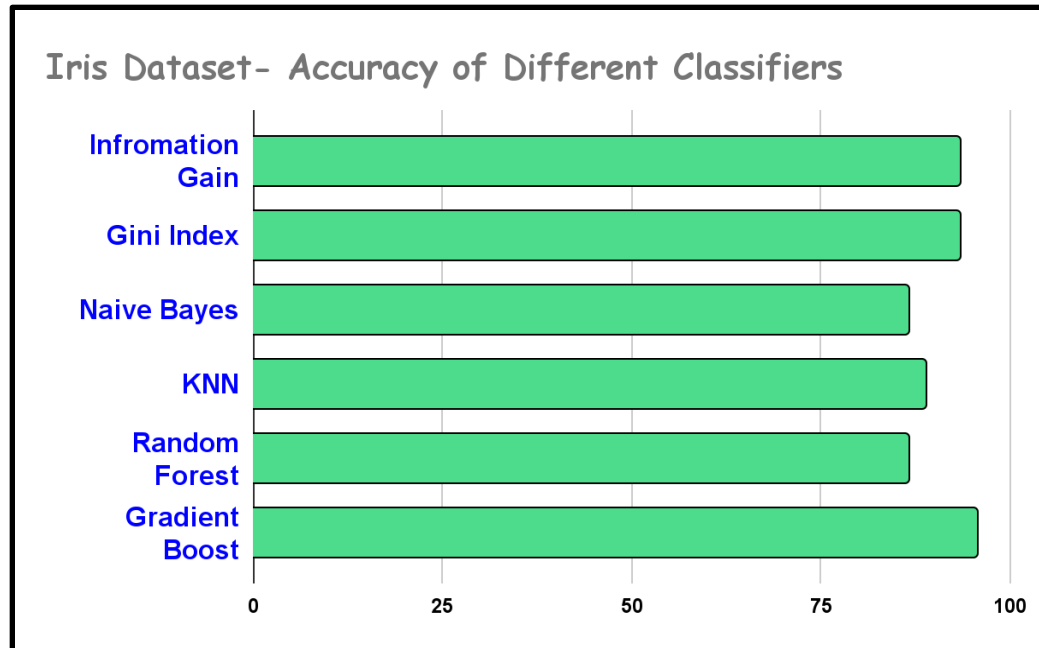
Further developments like improving flexibility with variables like split ratio, choices for the fields and graphical visualisation are planned to be introduced in the updated version of Application.

## 5 Conclusion

For the [iris](#) dataset, we got the Performance Metrics Comparison

Sr. no	Classifier	Accuracy
1	Information Gain	93.333
2	Gini Index	93.333
3	Naive Bayes	86.664
4	KNN	88.889
5	Random Forest	86.667
6	Gradient Boost	95.556

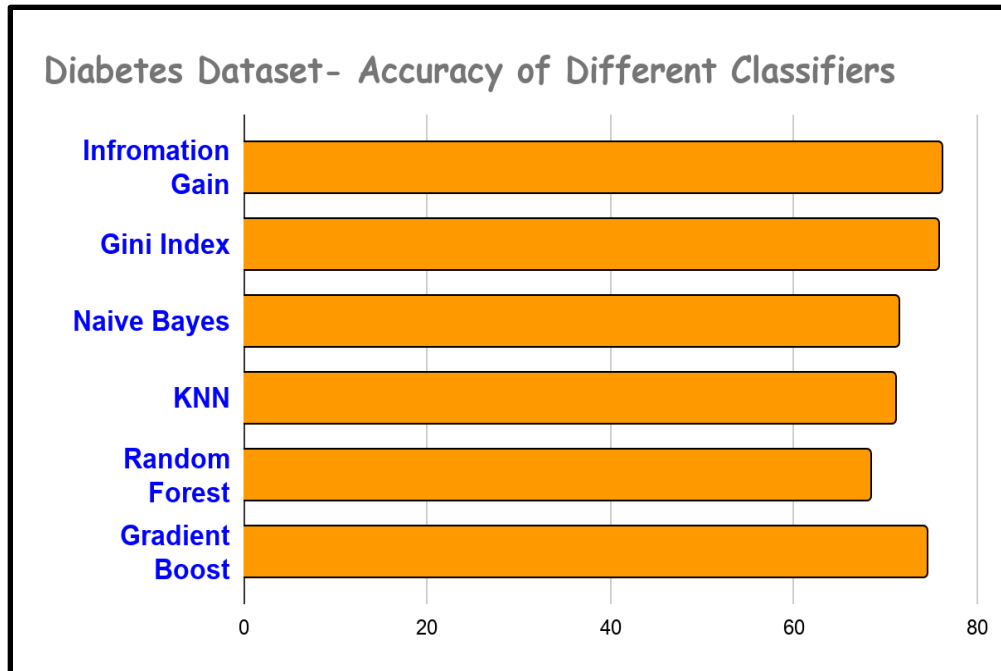
Output of Best Classifier is: **Gradient Boost.**



For the [diabetes](#) dataset, the result was as following:

Sr. no	Classifier	Accuracy
1	Information Gain	76.19
2	Gini Index	75.758
3	Naive Bayes	71.429
4	KNN	70.996
5	Random Forest	68.398
6	Gradient Boost	74.459

Output of Best Classifier is: **Information Gain**



## 6 References

- <https://en.wikipedia.org/wiki/Classification>
- <https://www.analytixlabs.co.in/blog/classification-in-machine-learning/>
- [https://www.youtube.com/playlist?list=PLu0W\\_9lII9ajLcqRcj4PoEihkukF\\_OTzA](https://www.youtube.com/playlist?list=PLu0W_9lII9ajLcqRcj4PoEihkukF_OTzA)
- <https://www.youtube.com/watch?v=QWqxRchawZY&list=LL&index=1&t=334s>
- [Iris dataset](#)
- [Diabetes dataset](#)