

IBM HACK CHALLENGE 2022

GROUP MEMBERS:

PRANAV SINGH

VEDANT TADLA

KUNAL THAKKAR

TANISHK JAIN

PROJECT NAME:

Nidan

TABLE CONTENTS

TITLE	PAGE NO.
INTRODUCTION	1
LITERATURE SURVEY	2
THEORITICAL ANALYSIS	3
EXPERIMENTAL INVESTIGATIONS	4
FLOWCHART	6
ADVANTAGES & DISADVANTAGES	7
APPLICATIONS	8
CONCLUSION	9
FUTURE SCOPE	11
BIBILOGRAPHY	12
APPENDIX	13

1.INTRODUCTION

1.1 OVERVIEW:

‘NIDAN’ , a Sanskrit word for diagnosis or cure, is a platform for people to diagnose chronic diseases and mental health . Our aim is to make it user friendly so that common users can take advantage since we will be displaying tips related to aging and related to improving mental health.

NIDAN’s aim is to be a catalyst that makes the world work better and improve common people’s health.

1.2 PURPOSE:

The main purpose of this project is to:

- Cater professional healthcare sector
- Cross checking scientific data
- Cross checking of diagnosis
- Making user data for research purpose
- User monitoring of their lifestyle
- Improving lifestyle by giving some important tips

LITERATURE SURVEY:

2.1 EXISTING PROBLEM:

Currently the diagnostic process proceeds as follows:

First, a patient experiences a health problem. The patient is likely the first person to consider his or her symptoms and may choose at this point to engage with the health care system. Once a patient seeks health care, there is an iterative process of information gathering, information integration and interpretation, and determining a working diagnosis. Performing a clinical history and interview, conducting a physical exam, performing diagnostic testing, and referring or consulting with other clinicians are all ways of accumulating information that may be relevant to understanding a patient's health problem. The information-gathering approaches can be employed at different times, and diagnostic information can be obtained in different orders. The continuous process of information gathering, integration, and interpretation involves hypothesis generation and updating prior probabilities as more information is learned. Communication among health care professionals, the patient, and the patient's family members is critical in this cycle of information gathering, integration, and interpretation.`

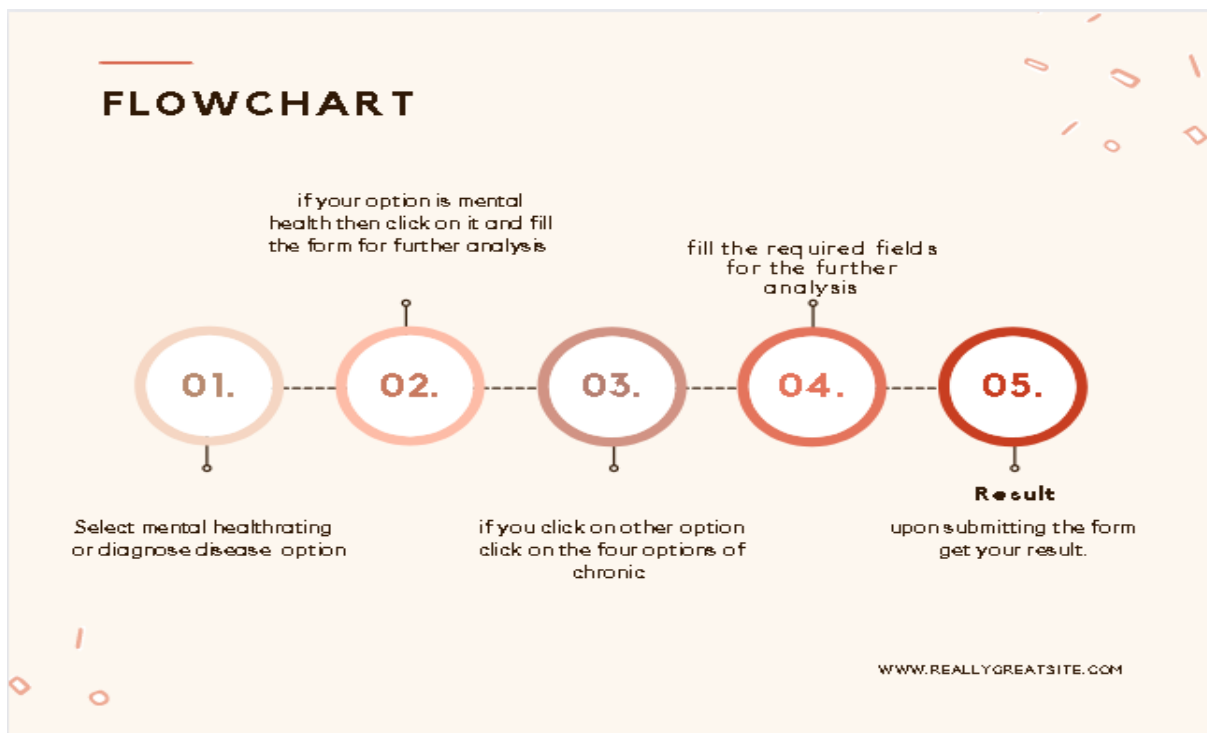
2.2 Proposed solution

As mentioned in introduction we will help the doctors along with the common civilians to perform the diagnosis test provided with an simple user interface which is an multiclass classification model that will predict weather you are prone to a

particular disease or not and if you are affected by a disease then to what extent you are affected and how you can get cured from that as soon as possible by providing some common tips.

3.THEORETICAL ANALYSIS

3.1 FLOWCHART:



3.2 Hardware / Software designing

Software Requirements:

- Html , CSS , JavaScript
- Python and its Libraries [Pandas , Numpy ,Seaborn , Matplotlib]

- iIBM Watson
- Github repository
- Figma
- Google colab
- Flask
- VS Code

4.EXPERIMENTAL INVESTIGATION

1.DIABETES DATASET:

LIBRARIES USED: Pandas , Numpy , Matplotlib , Seaborn ,Sklearn,joblib

DATASET DIMENSIONS :[253680 rows , 22 columns]

RESULTS OF MODELS USED FOR PREDICTION:

1.XGBOOST: ACCURACY SCORE: 0.8466966256701356

CONFUSION MATRIX:

42422,	0,	319
892,	0,	34

6533	0	536
------	---	-----

2.RANDOM FOREST:

ACCURACY SCORE:0.848076316619363

CONFUSION MATRIX:

42422,	0,	319
892,	0,	34
6533	0	536

3.CNN MODEL:

ACTIVATION FUNCTIONS:_relu and softmax

NUMBER OF EPOCH: 100

CONFUSION MATRIX:

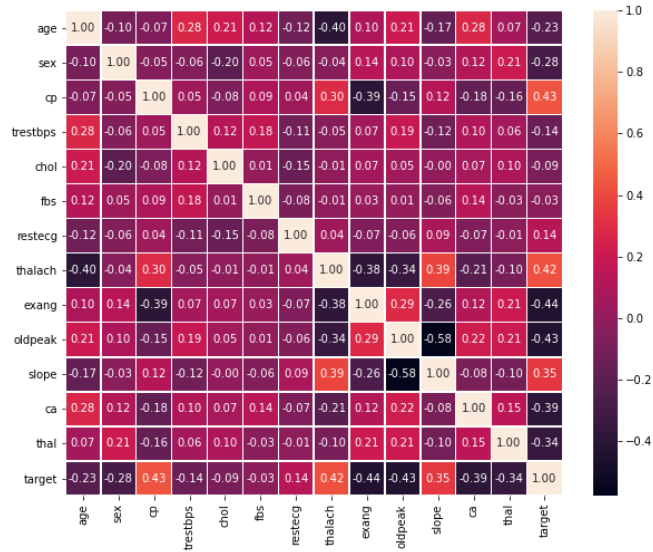
42741	0	0
926	0	0
7069	0	0

HEART DISEASE PREDICTION:

LIBRARIES USED: Pandas , Numpy , Matplotlib , Seaborn ,Sklearn

DATASET DIMENSIONS :[303 rows , 14 columns]

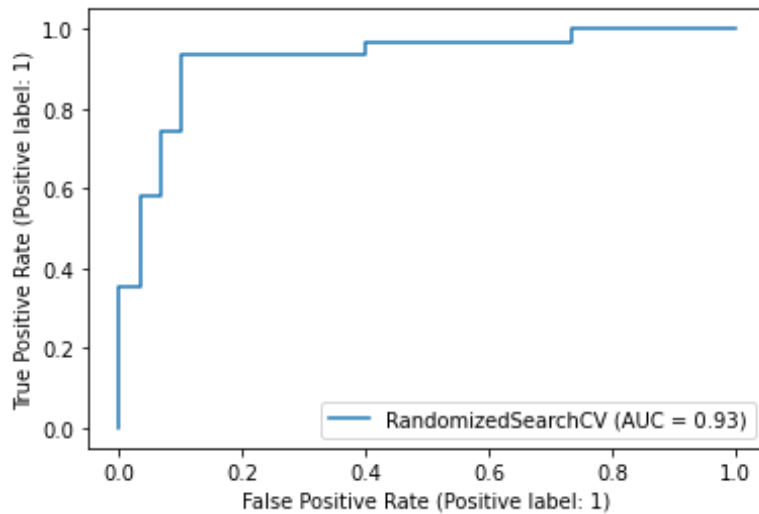
CORRELATION MATRIX:



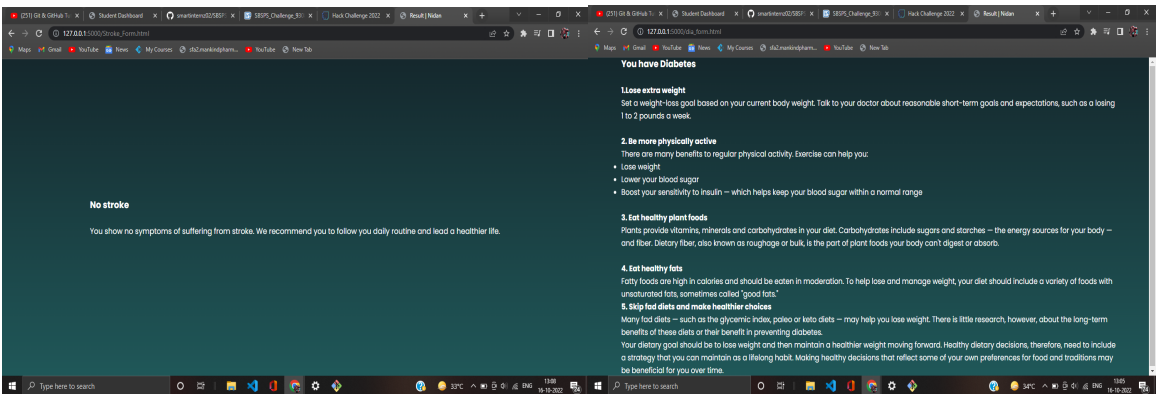
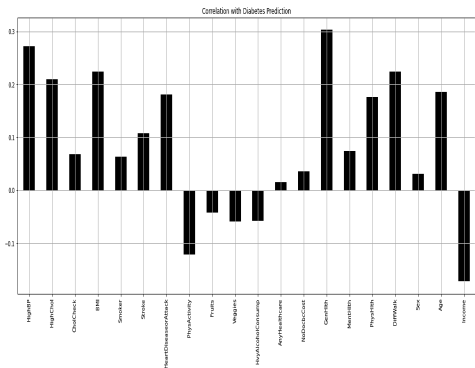
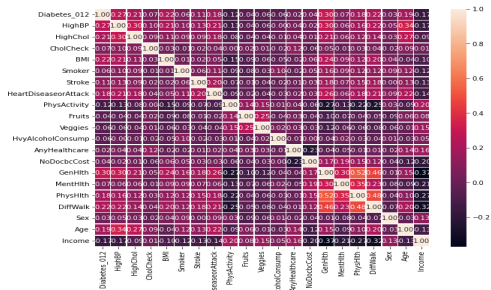
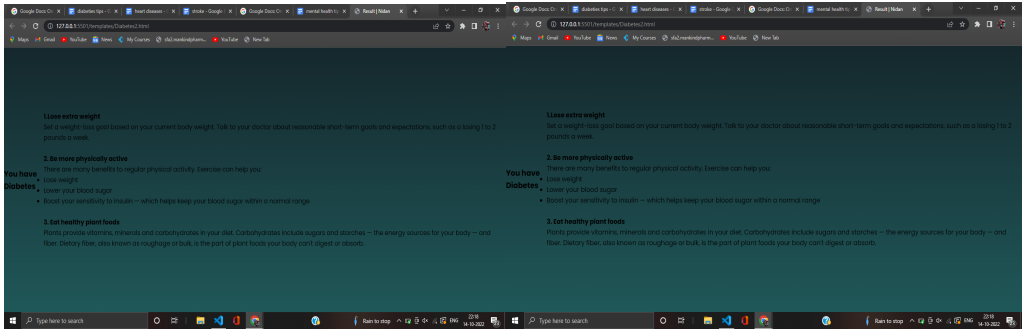
LOGISTIC REGRESSION : ACCURACY

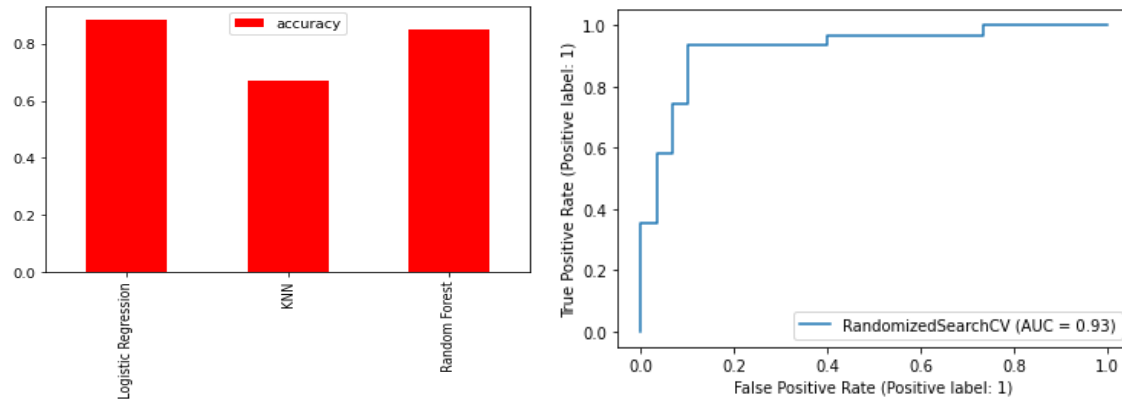
SCORE:0.8524590163934426

RANDOMISED SEARCH ACCURACY SCORE:**0.8852459016393442**



6.OUTPUT / RESULT :





7.ADVANTAGES & DISADVANTAGES

ADVANTAGES:

- Cost efficient
- Easy to use both for Doctors and common civilians
- Project helps hospitals for monitoring and cross checking patients data along with gathering data for further analysis and study.
- Helps in saving both in terms of time and money
- Tips will be provided so that person can get alert for his health condition.
- We have aimed for higher recall in our project.

DISADVANTAGES:

The above model is not 100 percent accurate and gives only predicted outputs.

8.APPLICATIONS

The above project can be used to:

- Cater professional healthcare sector
- Cross checking scientific data
- Cross checking of diagnosis
- Making user data for research purpose
- User monitoring of their lifestyle
- Improving lifestyle by giving some important tips

9.CONCLUSION:

The main aim of this project is to predict the disease in accordance with symptoms put down by the patients with proper implementation of Machine Learning algorithm. In this project we have used four Machine Learning algorithms for prediction and achieved the mean accuracy of more than 85% which shows remarkable rectification and higher accuracy than previous work and also makes this system more reliable than the existing one for this job and hence provides better satisfaction to the user in comparison with the other one. We have also created a GUI for better interaction with the system by users which is very easy to operate .This project shows that Machine Learning algorithms can be used to predict diseases easily with different parameters and models. In the end we can say that our system has no threshold of the users because

everyone can use this system.

10.FUTURE SCOPE:

There are many possible improvements that could be explored to diversify the research by discovering and considering extra features. Due to time boundation , the following work is required to be performed in future. There is a plan to use more classification techniques \ methods, different discretization techniques, multiple classifier voting methods. Would like to use different rules such as association rules and various algorithms. In the future, we are willing to make use of filter based feature selection methods in order to achieve more appropriate as well as functional results.

Furthermore database can be added to store patients health.

11.BIBLIOGRAPHY:

Referred books :

Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd
Edition