**VOIS INTERNSHIP 2022**

**GROUP MEMBERS:**

**PRANAV SINGH**

**VEDANT TADLA**

**KUNAL THAKKAR**

**Mentored By-**

MAYANK SHRIVASTAVA

**PROJECT NAME:**

**Nidan**

**TABLE CONTENTS**

| **TITLE** | **PAGE NO.** |
| --- | --- |
| **INTRODUCTION** | **1** |
| **ABOUT MACHINE LEARNING** | **2** |
| **TYPES OF MACHINE LEARNING** | **3** |
| **ML MODEL IN OUR PROJECT** | **5** |
| **ABOUT FRONTEND** | **16** |
| **CONCLUSION** |  |

**INTRODUCTION:**

### **What is machine learning?**

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

Recommendation engines are a common use case for machine learning. Other popular uses include fraud detection, spam filtering, malware threat detection, business process automation (BPA) and Predictive maintenance.

### Why is machine learning important?

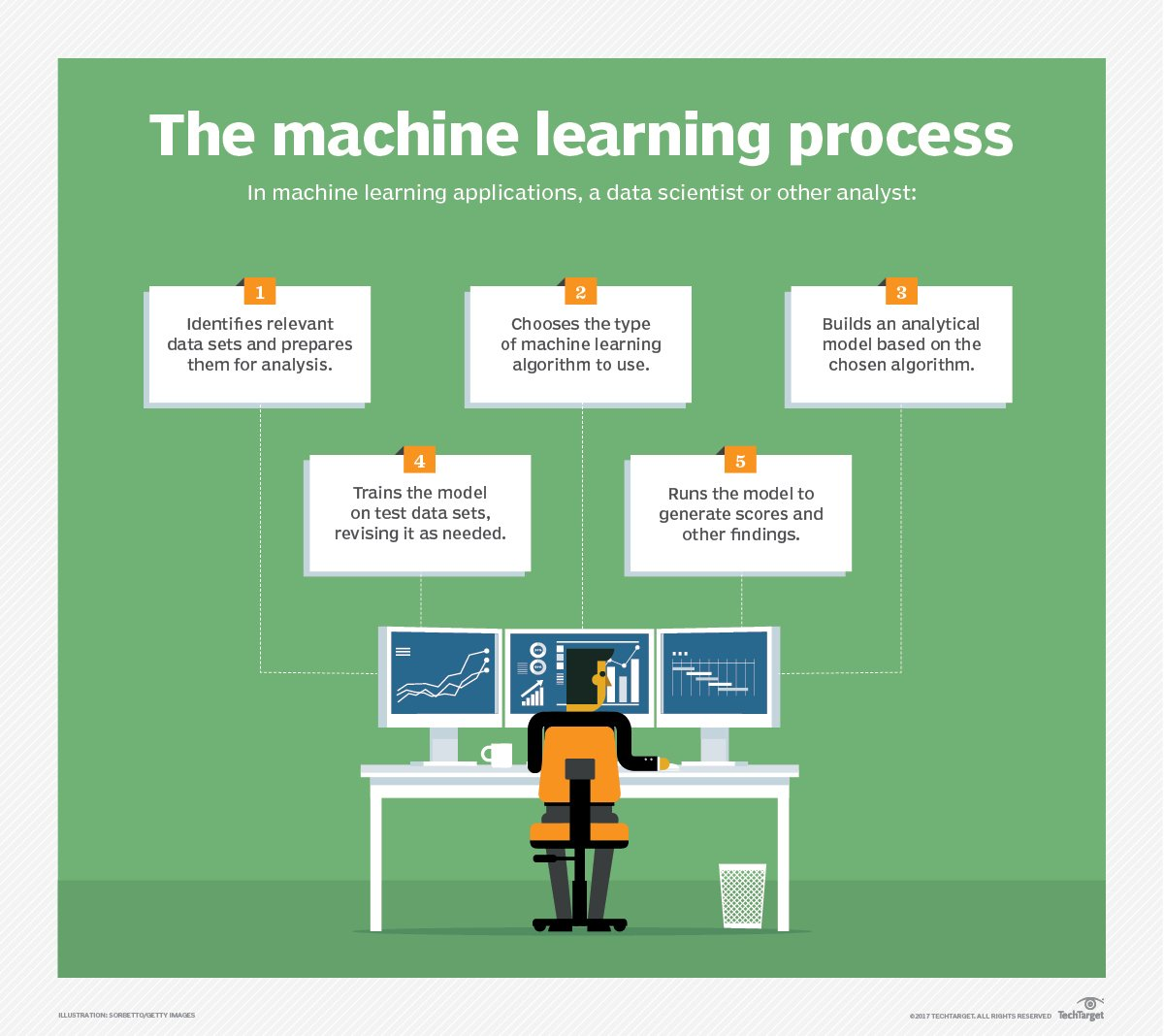
Machine learning is important because it gives enterprises a view of trends in customer behaviour and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

### **What are the different types of machine learning?**

Classical machine learning is often categorised by how an algorithm learns to become more accurate in its predictions. There are four basic approaches:supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. The type of algorithm data scientists choose to use depends on what type of data they want to predict.

**TYPES OF MACHINE LEARNING**

* **Supervised learning:** In this type of machine learning, data scientists supply algorithms with labelled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.
* **Unsupervised learning:** This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through datasets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.
* **Semi-supervised learning:** This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labelled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.
* **Reinforcement learning:** Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

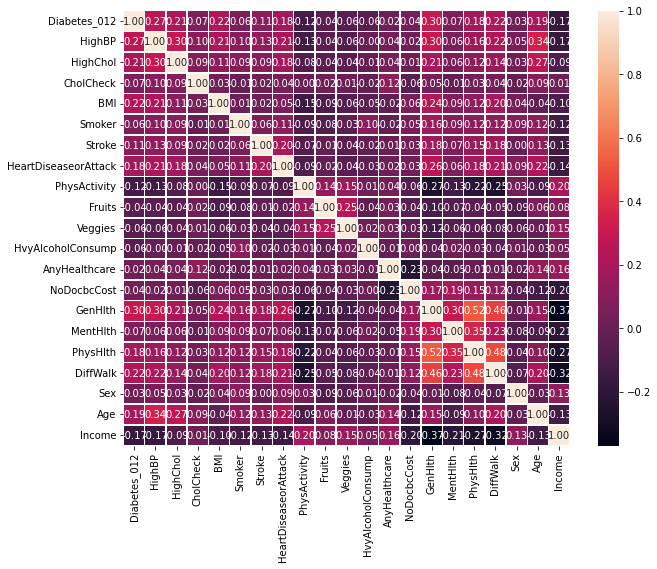


**ML MODEL IN OUR PROJECT:**

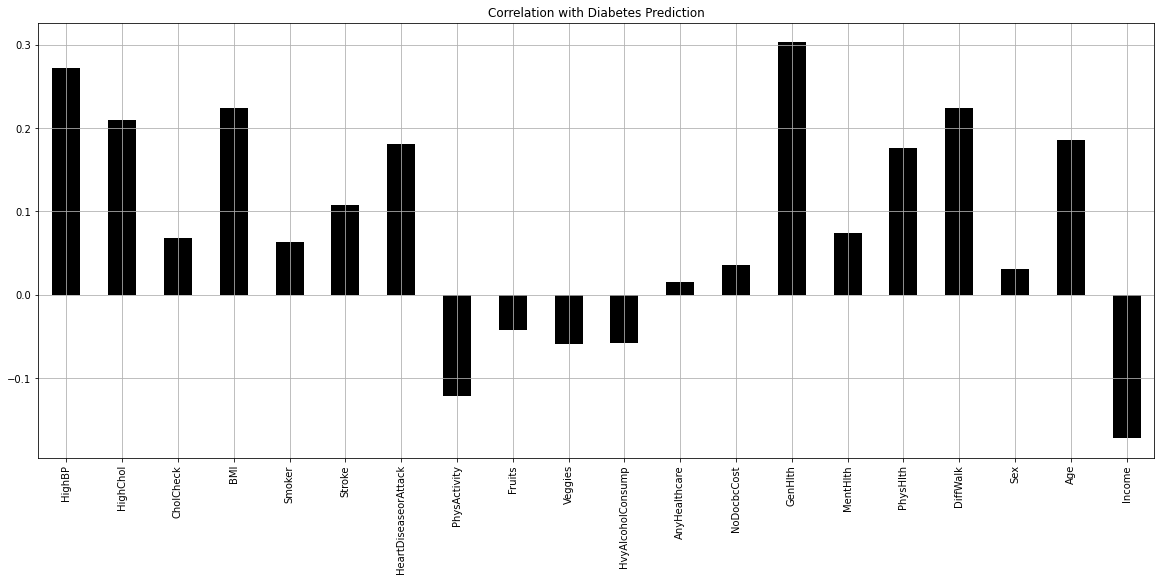
**1.DIABETES DATASET:**

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

**DATASET DIMENSIONS :[**253680 rows , 22 columns]

**CORRELATION MATRIX:**

**CORRELATION WITH DIABETES PREDICTION:**



**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:**

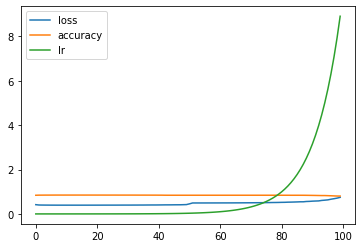
**ACTIVATION FUNCTIONS:** relu and softmax

**NUMBER OF EPOCH:** 100

**CONFUSION MATRIX:**

| 42741 | 0 | 0 |
| --- | --- | --- |
| 926 | 0 | 0 |
| 7069 | 0 | 0 |

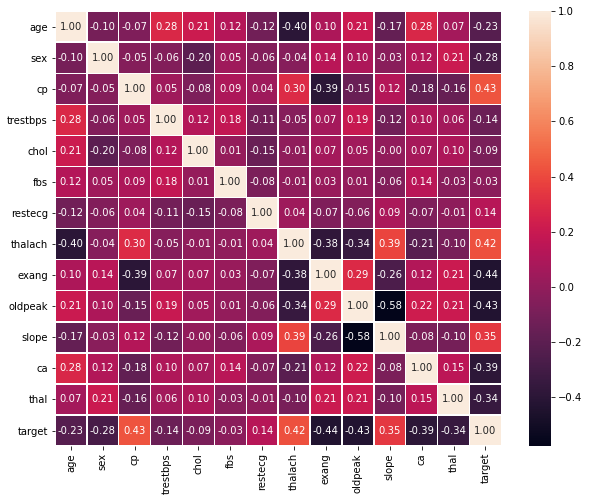
**ACCURACY SCORE:**0.8424

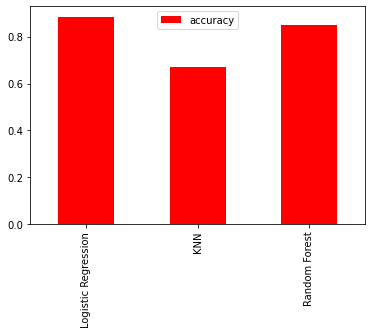


**HEART DISEASE PREDICTION:**

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

**DATASET DIMENSIONS :[**303 rows , 14 columns]

**CORRELATION MATRIX:**

**MODEL ACCURACY SCORES:**

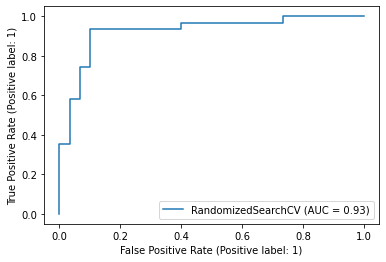
**‘Logistic Regression':** 0.8852459016393442,

**'KNN':** 0.6721311475409836,

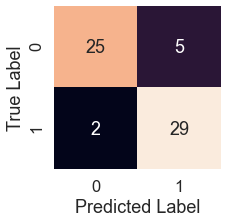
**'Random Forest':** 0.8524590163934426

**LOGISTIC REGRESSION ACCURACY SCORE:**0.8524590163934426

**RANDOMISED SEARCH ACCURACY SCORE:0.8852459016393442**

****

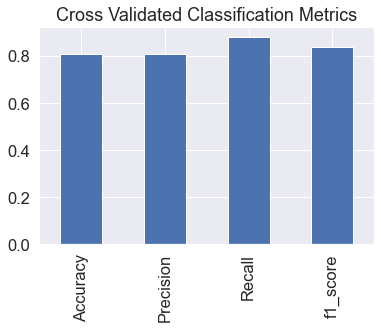
**CONFUSION MATRIX:**

****

**EVALUATION OF MATRIX USING CROSS VALIDATION METHOD**

**CROSS VALIDATION SCORE(RECALL METHOD):0.8803418803418802**

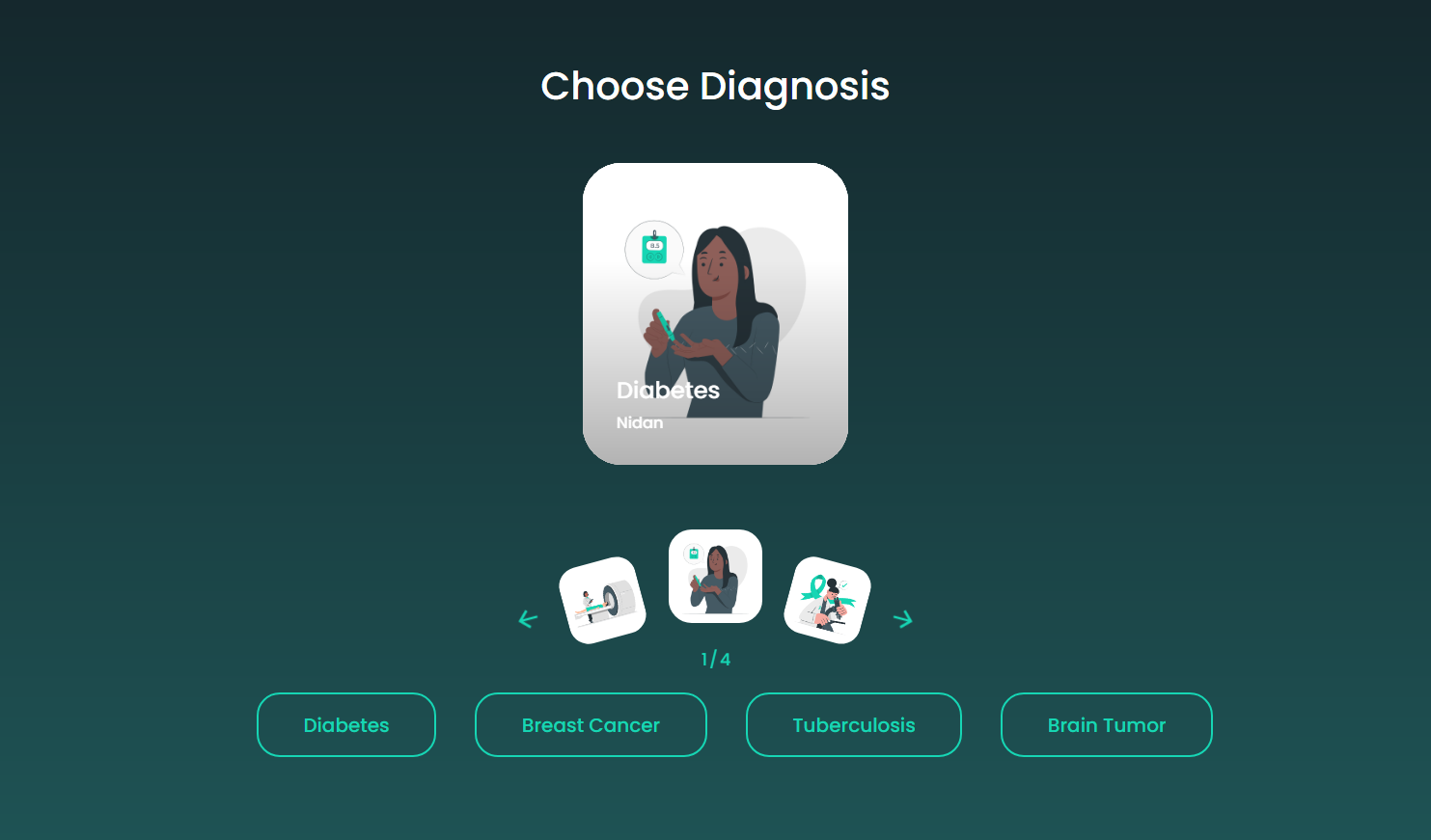
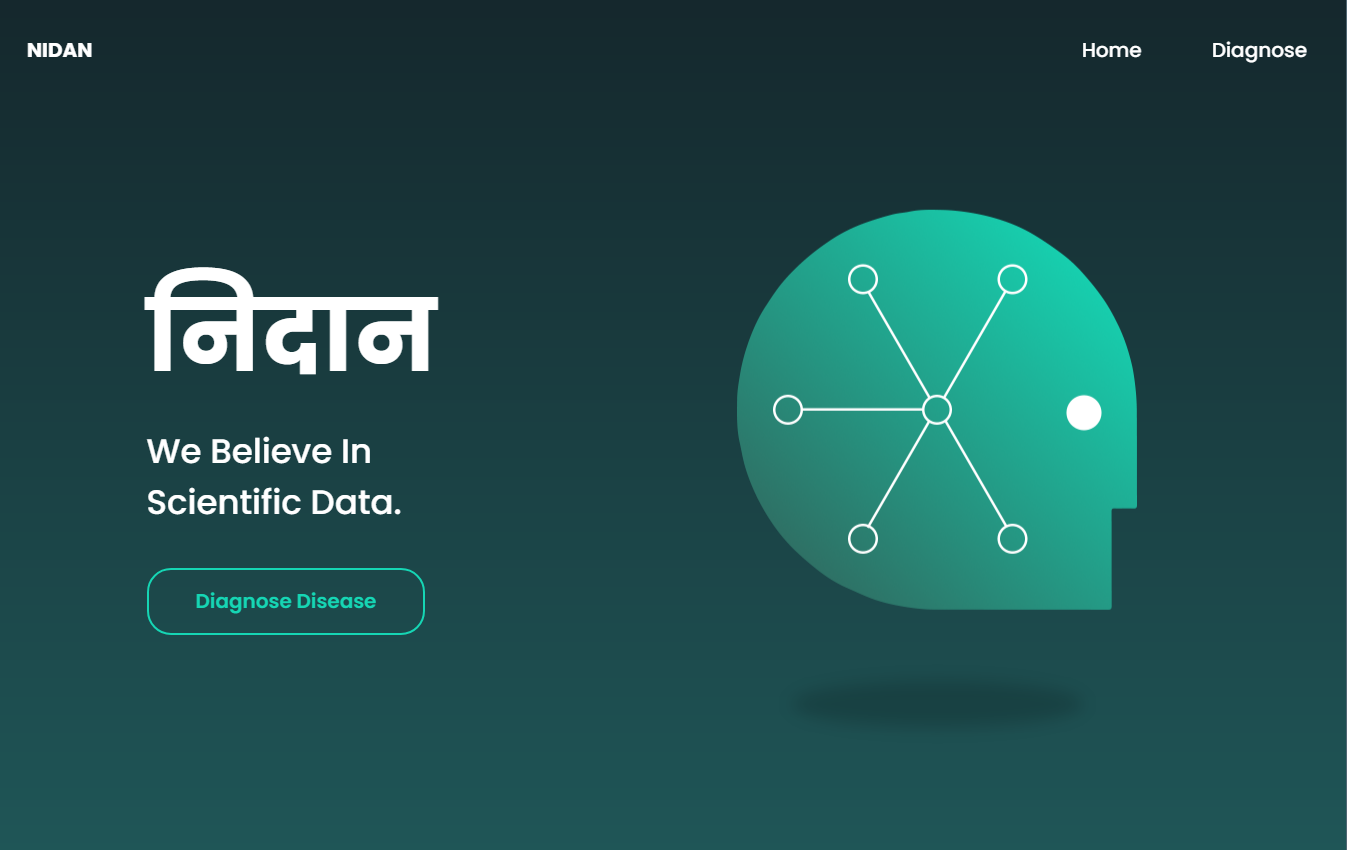
**CROSS VALIDATION SCORE(RECALL METHOD):0.8090795641140469**

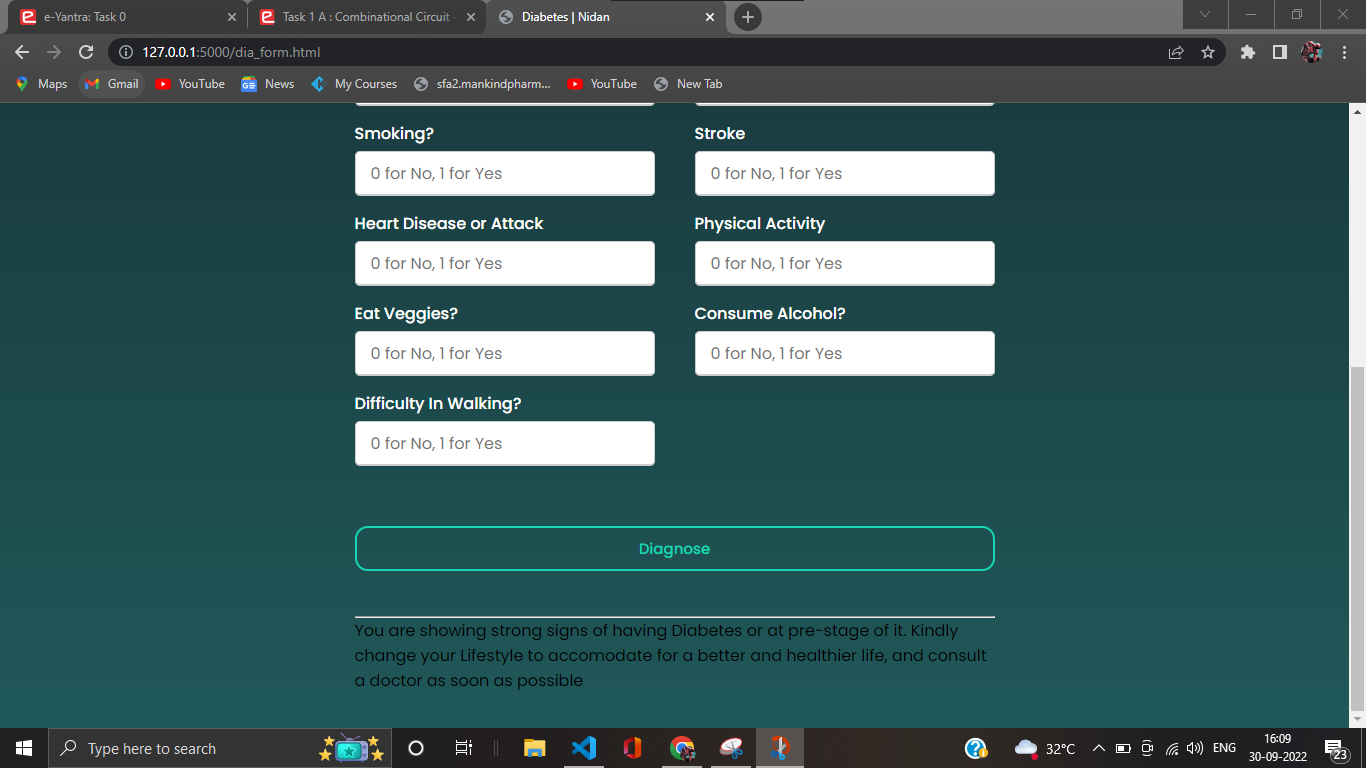
****

**ABOUT FRONTEND:**

We have developed the front in a way that even domestic users will find it easy to use NIDAN. We have tried making it aesthetic and user friendly.

For the frontend we used HTML, CSS and JavaScript. And our website runs on Flask.



****

**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.

**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.