



**An  
Internship Report  
On**

**“DIABETES PREDICTION USING MACHINE LEARNING”**

An Internship work submitted to Khaja BandaNawaz University, Kalaburagi  
In partial fulfillment of the requirement for the award of the Degree of

**BACHELOR OF ENGINEERING  
IN  
COMPUTER SCIENCE AND ENGINEERING**

**Submitted by:**

**TANMAY D MORE    UIN:19KB02BS[082]**

**Internship carried out at:  
“EDUTECH”**



**UNDER THE GUIDANCE OF**

**PROF. NIKHAT FATIMA  
M.Tech(Ph.D)**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
FACULTY OF ENGINEERING AND TECHNOLOGY  
KHAJA BANDANAWAZ UNIVERSITY**

Recognized by Government of Karnataka

KALABURAGI-585104

2022-2023



**FACULTY OF ENGINEERING AND TECHNOLOGY**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**Certificate**

This is to certify that the Internship on **“DIABETES PREDICTION USING MACHINE LEARNING”** is carried out by **TANMAY DATTU MORE UIN:19KB02BS[082]**, is a bonafide student of **Faculty of Engineering and Technology** in partial fulfillment for the award of Degree of **BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING** of the Khaja Banda Nawaz University, Kalaburagi during the academic year 2022-2023. It is certified that all the corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the department library.

The Internship report has been approved as it satisfies the academic requirement in respect of internship work prescribed for the said Degree.

Guide

**PROF. NIKHAT FATIMA**  
**MTech, (Ph.D)**

Head of the Department

**Dr. SAMEENA BANU**  
**MTech, (Ph.D)**

Examiner

- 1.
- 2.



## **ACKNOWLEDGEMENT**

The internship opportunity I had with “**EDUTECH**” was a great chance for learning and professional development. Therefore, I consider myself as a very lucky individual as I was provided with an opportunity to be a part of it. I am also grateful for having a chance to meet so many wonderful people and professionals who led me through this internship period.

I express my deepest thanks to “**PROF.NIKHAT FATIMA**”, for taking part in useful decision and giving necessary advices. I choose this moment to acknowledge her contribution gracefully.

I am very much thankful and obliged to my Head of the Department **Dr. SAMEENA BANU** for her encouragement and insightful comments throughout my internship.

I express my deep sense of gratitude to my Dean **Dr. S K Azam** who has provided me an opportunity to fulfill the most cherished desire to reach my goal.

**TANMAY D MORE**

**UIN: -19KB02BS[082]**



# **CONTENTS**

<b>SL.NO</b>	<b>PARTICULARS</b>	<b>PAGE NO.</b>
1.	INTRODUCTION 1.1 Start-up Brief Understanding 1.2 History of Company 1.3 Company Structure 1.4 Services offered By Company	01
2.	NARRATION 2.1 Basics of Python 2.2 Anaconda 2.3 Spyder 2.4 Introduction to Artificial Intelligence 2.5 Introduction to Machine Learning 2.6 Introduction to Deep Learning	06
3.	TASK PERFROMED Diabetes prediction using ML 3.1 Introduction 3.2 Methodology 3.3 Results and Discussions	11
4.	SELF EVALUATION	16
	CONCLUSION	17
	REFERENCES	18

<b>Figure No.</b>	<b>Figure Name</b>	<b>Page No.</b>
Figure 1.1	Company structure	03
Figure 2.1	Anaconda Navigator	07
Figure 2.2	Spyder	08
Figure 2.3	Uses of Artificial Intelligence	09
Figure 2.4	Evolution of AI	09
Figure 2.5	Machine Learning Techniques with Python	10
Figure 2.6	AI, ML, & Deep learning	10
Figure 3.1	USER INPUT	14
Figure 3.2	Person Don't Have Diabetes	15
Figure 3.3	Person Has Diabetes	15

---

---

## CHAPTER-1

### INTRODUCTION

#### 1.1 STARTUP- A BRIEF UNDERSTANDING

A Startup (Startup or start-up) is an entrepreneurial venture which is typically a newly emerged, fast-growing business that aims to meet a marketplace need by developing a viable business model around an idea. A Startup is usually a company designed to effectively develop and validate a scalable business model. Start-ups have high rates of failure, but the minority of successes includes companies that have become large and influential.

Startup companies can come in all forms and sizes. Some of the critical tasks are to build a co-founder team to secure key skills, know-how, financial resources, and other elements to conduct research on the target market. Typically, a startup will begin by building a first Minimum viable Product (MVP), a prototype to validate, assess and develop the new ideas or business concepts. In addition, startups founders do research to deepen their understanding of the ideas, technologies or business concepts and their commercial potential. A Share Holders' Agreement (SHA) is often agreed early on to confirm the commitment, ownership and contributions of the founders and investors and to deal with the intellectual properties and assets that may be generated by the startup. Business models for startups are generally found via a "bottom-up" or "top-down" approach.

A company may cease to be a startup as it passes various milestones, such as becoming publicly traded on the stock market in an Initial Public Offering (IPO), or ceasing to exist as an independent entity via a merger or acquisition. Companies may also fail and cease to operate altogether, an outcome that is very likely for startups, given that they are developing disruptive innovations which may not function as expected and for which there may not be market demand, even when the product or service is finally developed. Given that startups operate in high-risk sectors, it can also be hard to attract investors to support the product/service development or attract buyers.

The size and maturity of the startup ecosystem where the startup is launched and where it grows have an effect on the volume and success of the startups. The startup ecosystem consists of the individuals (entrepreneurs, venture capitalists, Angel investors, mentors); institutions and organizations (top research universities and institutes, business schools and entrepreneurship programs operated by universities and colleges, non-profit entrepreneurship

support organizations, government entrepreneurship programs and services, Chambers of commerce. A region with all of these elements is considered to be a "strong" startup ecosystem.

Investors are generally most attracted to those new companies distinguished by their strong co-founding team, a balanced "risk/reward" profile (in which high risk due to the untested, disruptive innovations is balanced out by high potential returns) and "scalability" (the likelihood that a start-up can expand its operations by serving more markets or more customers). Attractive start-ups generally have lower "bootstrapping" (self-funding of start-ups the founders) costs, higher risk, and higher potential return on investment. Successful start-ups are typically more scalable than an established business, in the sense that the start-up has the potential to grow rapidly with a limited investment of capital, labour or land. Timing has often been the single most important factor for biggest start-up successes, while at the same time it's identified to be one of the hardest things to master by many serial entrepreneurs and investors.

### **1.2 HISTORY OF THE COMPANY**

EduTech established in April 2015, is a Start-up based in Gulbarga, which provide Software related services including Web-Design and Development and Software Development. Of late EduTech has started working on its own ideas which aim to solve community problems. To provide a strong foothold for budding Engineers and to create a strong and able workforce, EduTech has initiated Industrial Internship Program.

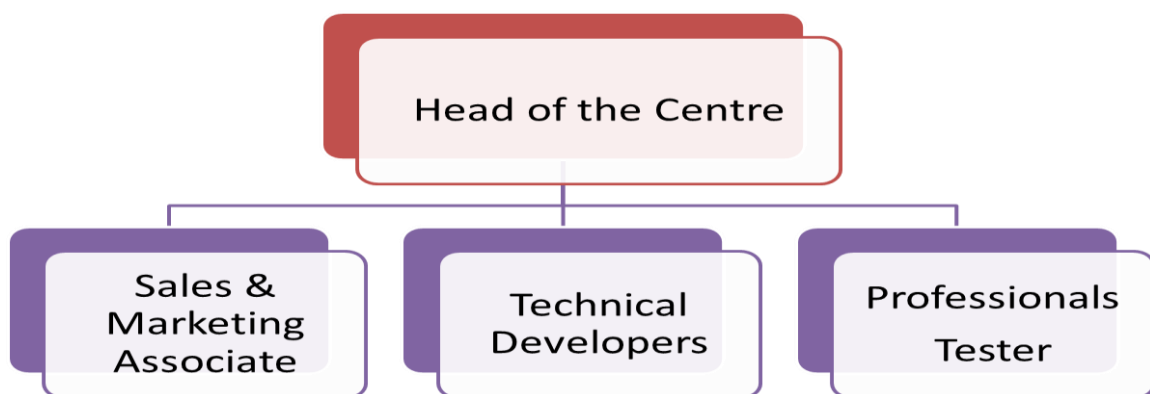
Internship and Learning bootcamp@EduTech Industrial Internship Program at EduTech aims to provide hands-on experience to the intern including the process of Market Analysis, understanding the various technologies present in the market that can be used towards the development of the Product and select the right Technology stack suitable for the Product, design of scalable, modular and elastic architecture and product development including the knowledge of writing legible code following best practices. The basic aim of the Internship is to provide the intern a strong base which will expose the intern to various work fields present in the company, this will help the intern select and decide his/her carrier objective.

---

### 1.3 COMPANY STRUCTURE

Business needs to be organised in a specific form of shape that is generally referred to as organisational structure. Organisations are structured in a variety of ways, dependent on their objectives and culture. The structure of the company often dictates the way it operates and performs (Waterman et al., 1980). Traditionally, the businesses have been structured in a hierarchical way with several divisions and departments, each responsible for a specific task such as human resources management, production or marketing. Many layers of management controlled the operations, with each answerable to the upper layer of management. Although this is still the most widely used organisational structure, the recent trend is increasingly towards a flat structure where the work is done in teams of specialists rather than fixed departments.

The importance of human resources has thus got the central position in the strategy of the organisation, away from the traditional model of capital and land. It is also important for the organisation to instil confidence among the employees about their future in the organisation and future career growth as an incentive for hard work.



**Figure 1.1 Company Structure**

### 1.4 SERVICES OFFERED BY THE COMPANY

It is in the field of Web Development and Client-Server Application by providing amazing services at affordable prices. With more than 5 years of experience in kitty, it is expertise to accomplish the job in hand, right from requirement gathering to delivery, in a professional way. To build beautiful Websites and amazing Applications which showcases the business in the most professional way possible. EduTech takes pride in providing new talent an opportunity to showcase their skills.



## **1.5 NUMBER OF EMPLOYEES WORKING IN THE COMPANY**

IT software solution professionals always work closely with clients to modify existing systems or create new systems. They may modify “off the shelf” software and integrate it into the existing systems. They often work as part of a team of software professionals responsible for the requirement specification, system analysis and design, construction, testing, training and implementation, as well as maintenance of a business software system.

## **1.6 MANAGEMENT TEAM**

The management team is an essential part of every business. This team analyses and identifies the business’ goals and objectives and implements and enforces the strategies the employees need to achieve success. In a business plan, the management team includes the business’ owners, board of directors and managers.

## **1.7 ABOUT DEPARTMENT**

### **1.7.1 FUNCTIONALITY**

Any operating organisation should have its own structure in order to operate efficiently. For an organisation, the organisational structure is a hierarchy of people and its functions.

The organizational structure of an organization tells you the character of an organization and the values it believes in. therefore, when you do business with an organization or getting into a new job in an organization, it is always a great idea to get to know and understand their organizational structure.

Depending on the organizational values and the nature of the business, organizations tend to adopt one of the following structures for management purpose. Although the organization follows a particular structure, there can be department and teams following some other organizational structure in exceptional cases. Sometimes, some organizations may follow a combination of the following organizational structures as well.

### **1.7.2 R & D DEPARTMENT**

Research and Development Department is tasked with developing new products, or staffed with industrial scientists and tasked with applied research in scientific or technological fields which may facilitate future product development.

A company's research and development department play an integral role in the life cycle of a product. While the department usually is separate from sales, production and other divisions, related and often requires collaboration. A thorough understanding of the research and development department allows you to maximize those duties at your small business, even if you don't have a big department.

### **1.7.3 SOFTWARE DEVELOPMENT PROCESS ADOPTED**

A software development process or life cycle is a structure imposed on the development of a software product. There are several models for such processes, each describing approaches to a variety of tasks or activities that take place during the process.

- Process Activities/Steps
- Requirements Analysis
- Specification
- Software architecture
- Implementation
- Testing
- Documentation
- Training and Support

## CHAPTER 2

### NARRATION

#### 2.1 PYTHON

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation.[1]

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library.

Guido van Rossum began working on Python in the late 1980s as a successor to the ABC programming language and first released it in 1991 as Python 0.9.0. Python 2.0 was released in 2000 and introduced new features such as list comprehensions, cycle-detecting garbage collection, reference counting, and Unicode support. Python 3.0, released in 2008, was a major revision that is not completely backward-compatible with earlier versions.

Python is used for:

- Web Development (server-side),
- Software Development,
- Mathematics,
- System scripting.

#### 2.2 Anaconda (Python distribution)

Anaconda is a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment[1]. The distribution includes data-science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012. As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition and Anaconda Enterprise Edition, both of which are not free.

Package versions in Anaconda are managed by the package management system conda. This package manager was spun out as a separate open-source package as it ended up being useful on its own and for things other than Python. There is also a small, bootstrap version of

Anaconda called Miniconda, which includes only conda, Python, the packages they depend on, and a small number of other packages.

### 2.2.1 ANACONDA NAVIGATOR

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows users to launch applications and manage conda packages, environments and channels without using command-line commands[3.]Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository, install them in an environment, run the packages and update them. It is available for Windows, macOS and Linux.



**Figure 2.1 Anaconda Navigator**

The following applications are available by default in Navigator:

- JupyterLab
- Jupyter Notebook
- QtConsole
- Spyder
- Glue
- Orange
- RStudio
- Visual Studio Code

### 2.3 SPYDER

Spyder is an open-source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates with a number of prominent packages in the scientific Python stack, including NumPy, SciPy, Matplotlib, pandas, IPython, SymPy and Python, as well as other open-source software. It is released under the MIT license.

Initially created and developed by Pierre Raybaut in 2009, since 2012 Spyder has been maintained and continuously improved by a team of scientific Python developers and the community.

Spyder is extensible with first-party and third-party plugins, includes support for interactive tools for data inspection and embeds Python-specific code quality assurance and introspection instruments, such as Pyflakes, Pylint and Rope. It is available cross-platform through Anaconda, on Windows, on macOS through MacPorts, and on major Linux distributions such as Arch Linux, Debian, Fedora, Gentoo Linux, open-use and Ubuntu.

Spyder uses Qt for its GUI and is designed to use either of the PyQt or PySide Python bindings. QtPy, a thin abstraction layer developed by the Spyder project and later adopted by multiple other packages, provides the flexibility to use either backend.



**Figure 2.2 Spyder**

## 2.4 ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is intelligence - perceiving, synthesizing, and inferring information - demonstrated by machines, as opposed to intelligence displayed by animals and humans. Example tasks in which this is done include speech recognition, computer vision, translation between (natural) languages, as well as other mappings of inputs. OED (OUP) defines artificial intelligence as:

The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

## ✧ Uses of Artificial Intelligence



Figure 2.3 Uses of Artificial Intelligence

## ✧ Evolution of Artificial Intelligence

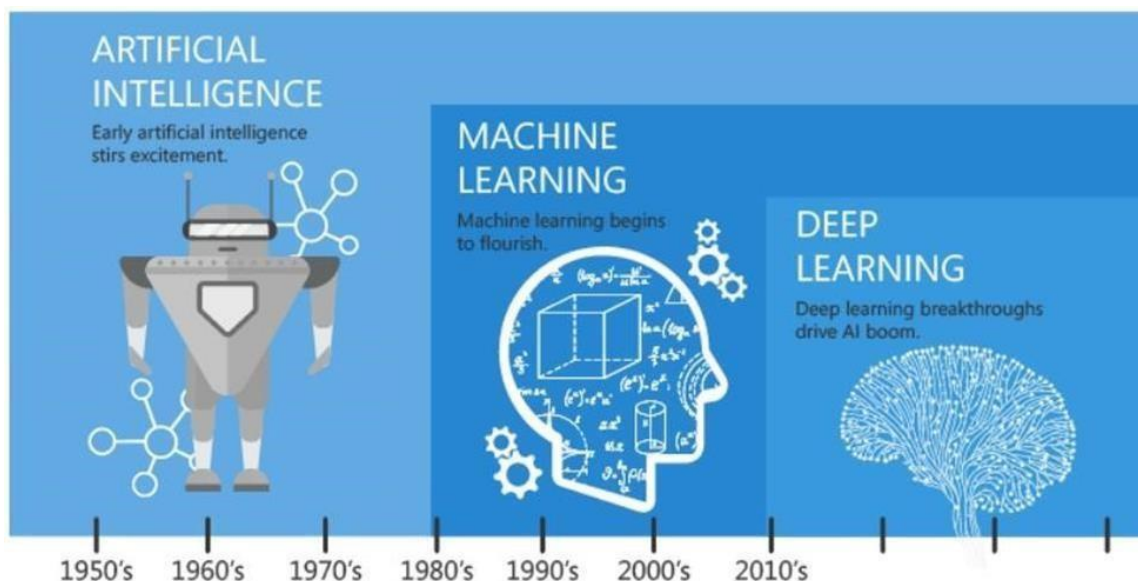


Figure 2.4 Evolution of AI

## 2.5 MACHINE LEARNING

Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being

explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, agriculture, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

Some implementations of machine learning use data and neural networks in a way that mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as predictive analytics.

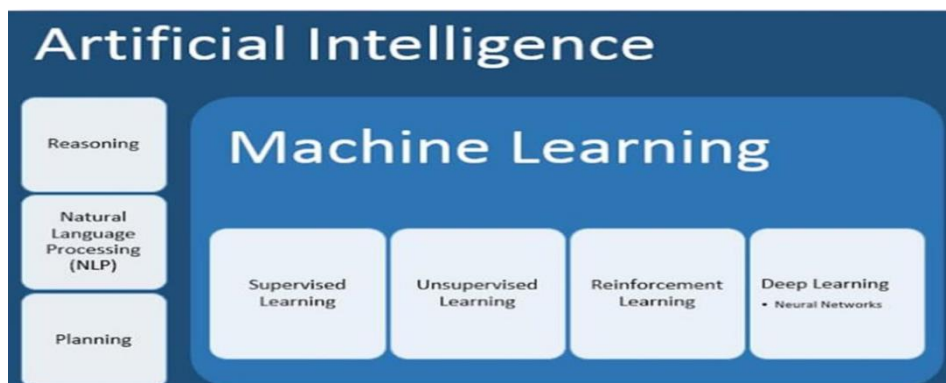


**Figure 2.5 Machine Learning Techniques with Python**

## 2.6 DEEP LEARNING

Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.

Deep-learning architectures such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks, conventional neural networks and Transformers have been applied to fields including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, climate science, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.



**Figure 2.6 AI, ML & Deep learning**

## CHAPTER-3

### TASK PERFORMED

## DIABETES PREDICTION USING ML

### 3.1 INTRODUCTION

All around there are numerous ceaseless infections that are boundless in evolved and developing nations. One of such sickness is diabetes. Diabetes is a metabolic issue that causes blood sugar by creating a significant measure of insulin in the human body or by producing a little measure of insulin. Diabetes is perhaps the deadliest sickness on the planet. It is not just a malady yet, also a maker of different sorts of sicknesses like a coronary failure, visual deficiency, kidney ailments and nerve harm, and so on Subsequently the identification of such chronic metabolic ailment at a beginning period could help specialists around the globe in forestalling loss of human life. Presently, with the ascent of machine learning, AI, and neural systems, and their application in various domains we may have the option to find an answer for this issue.

ML strategies and neural systems help scientists to find new realities from existing well-being-related informational indexes, which may help in ailment supervision and detection. The current work is completed utilizing the Pima Indians Diabetes Database. The point of this frame work is to make an ML model, which can anticipate with precision the likelihood or the odds of a patient being diabetic.

The ordinary distinguishing process for the location of diabetes is that the patient needs to visit a symptomatic focus. One of the key issues of bio-informatics examination is to achieve precise outcomes from the information. Human mistakes or various laboratory tests can entangle the procedure of identification of the disease. This model can foresee whether the patient has diabetes or not, aiding specialists to ensure that the patient in need of clinical consideration can get it on schedule and also help anticipate the loss of human lives DNA makes neural networks the apparent choice.



## **3.2 METHODOLOGY**

In this section we shall learn about the various classifier used in machine learning to predict diabetes, we shall also explain our proposed methodology to improve the accuracy few different methods of the machine learning model then ,the model can be used in prediction.<sup>3</sup>

### **3.2.1 Decision Tree**

Decision tree is a basic classification and regression methods Decision tree model has a tree structure, which can describe the process of classification instances based on features. It can be considered as a set of if then rule, which also can thought of as conditional probability distribution

### **3.2.2 Random Forest**

Random forest is an ensemble learning method for classification. This algorithm consists of trees and the number of tree structures present in the data is used to predict the accuracy.

Where leaves are corresponds to the class labels and attributes are corresponds to internal node of the tree. Here number of trees in forest used is 100 in number and Gini index is used for splitting the nodes.

### **3.2.3 AdaBoost Classifier:**

Boosting is an ensemble modelling technique that attempts to build a strong classifier from the number of weak classifiers. It is done by building a model by using weak models in series. Firstly, a model is built from the training data. Then the second model is built which tries to correct the errors present in the first model.[5]. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models are added. AdaBoost was the first really successful boosting algorithm developed for the purpose of binary classification. AdaBoost is short for Adaptive Boosting and is a very popular boosting technique that combines multiple “weak classifiers” into a single “strong classifier”. It was formulated by Yoav Freund and Robert Schapire. They also won the 2003 Gödel Prize for their work.

### **3.3 Design and implementation of classification model:**

In this task, comprehensive studies are done by applying different ML classification techniques like DT, KNN, RF, NB, LR, SVM.

#### **3.3.1 Machine learning classifier:**

We have developed a model using Machine learning Technique. Used different classifier and ensemble techniques to predict diabetes dataset.[4] We have applied SVM, LR, DT and RF Machine learning classifier to analyses the performance by finding accuracy of each classifier All the classifiers are implemented using scikit learn libraries in python. The implemented classification algorithms are described in next section.

#### **3.3.2 SCOPE:**

Healthcare profession found it hard to find healthcare data and perform analysis on them due to lack of tools, resources. But using ML, we can overcome this and can perform analysis on Real time data leading to better modelling predictions. This enhances and improves overall healthcare services. Now, IOTs being integrated with ML in order to make smart healthcare devices that sense if there is any change in person body, health data when he uses the device and this will notify the person regarding this through an app. this help in easy monitoring, advanced prediction and analysis there by reducing error, saving time and life of people

#### **3.3.3 PROJECT CONSTRAINTS**

Diabetes prediction is a classification technique with two mutually exclusive possible outcomes, either the person is diabetic or not diabetic. After extensive research, we came to conclusion that although numerous classification techniques can be used for the purpose of prediction, the observed accuracy varied. On careful examination of the performance of techniques used in prevalent works, logistic regression, KNN and logistic regression techniques were able to achieve 80% accuracy. The primary factor which influenced our algorithm selection was its adaptability and compatibility with future applications. Human mistakes or various laboratory test scan entangle the procedure of identification of the disease. This model can foresee whether the patient has diabetes or not, aiding specialists to ensure that the patient in need of clinical consideration can get it on schedule and also help anticipate the loss of human lives

### 3.4 RESULTS AND SCREENSHOT

Diabetes Prediction

Pregnancies

Number of Pregnancies

Glucose

Glucose

Blood Pressure

(mm Hg)

Skin Thickness

(mm)

Insulin

(mu U/ml)

Body Mass Index

weight in kg/(height in m)<sup>2</sup>

Diabetes Pedigree Function

DPF

Age

In Years

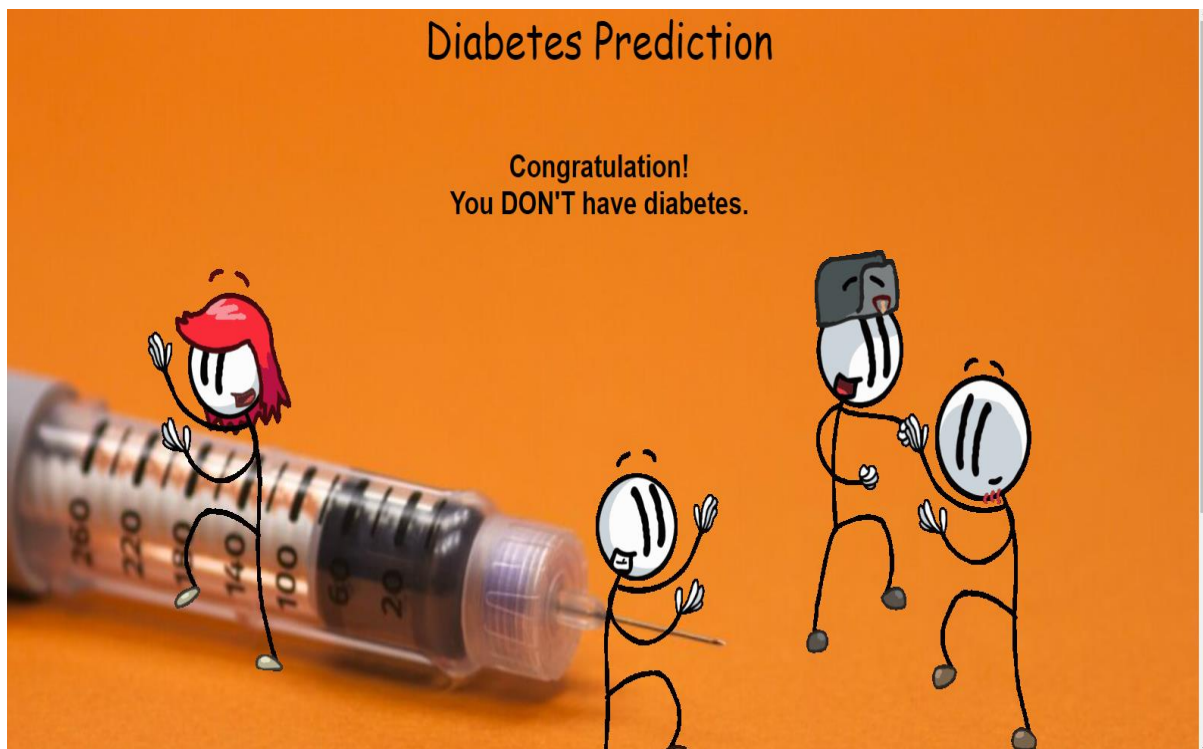
Submit

©2022 KBNU

**Fig 3.1. USER INPUT**

---

**OUTPUT:**



**Fig 3.2 PERSON DON'T HAVE DIABETES**



**Fig 3.3 PERSON HAS DIABETES**

---

---

## CHAPTER 4

### SELF-EVALUATION

An internship is a monitored work experience that has intentional learning outcomes and goals for me. Internships focus on personal career development and can be a semester, summer, or even year-long program. I put a thought on how to start my internship. Thus, I need to initially check for my interest in field of software. This would help me to choose my technology on which I want to work. The technology selection does not only depend on interest but also the how accurate, efficient, modular, scalable etc is the technology. The technology must be such it should always it should be in present use that will attract as well as make the user convenient to use.

Communication skills play a major role when it comes to express your ideas and thoughts and even to have a basic conversation with your colleagues. My stay has contributed to my communication skills, but I would like to pay more attention to it in the future. I can come across as reserved and uncertain. To contribute more to projects and to progress faster, I want to learn to make a more confident impression and to express my ideas and opinions more certain .

In course of internship, I have been asked to work in team. Different fun, non-technical and technical activities play a major role because they contribute to a better understanding among the team members. The joy of working and the thrill involved while tackling the various problems and challenges gave me a feel of developer's industry. It was due to this internship I came to know how professional software's are designed. I enjoyed each and every bit of work I had put into this internship.

Through this internship, I have seen what elements of my career I like and I got enthusiastic to Work hardly and heartily on it. I have found out that I should continue to do work on my technical and non-technical programs even after the completion of internship. By doing this, I think, I can and will have a better and satisfactory outcome of my life. This report concludes with my overall impressions of my work experience as well as my opinion of the Industrial Internship Program in general.

## CONCLUSION

In a nutshell, this internship has been an excellent and rewarding experience. I can conclude that there have been a lot I've learnt from my work at the research centre. Needless to say, the technical aspects of the work I've done are not flawless and could be improved provided enough time.

This internship was very interesting to me as I got an opportunity to learn in a different and more importantly in a corporate environment. During the course of internship, I had worked on project idea. I have skilled with how to take on project development. The overall internship is so useful, it improved my personality and I will carry this attitude throughout my professional and personal life.

This project took to the various phases of project development and gave me real insight into the world of software engineering. I enjoyed each and every bit of work I had put into this project. This project is further extendable. I believe my time spent in research and discovering new scripting languages was well worth it. Two main things that I've learned the importance of our time-management skills and self-motivation.

## REFERENCES

1. <https://python.org.com/>
2. <https://google.co.in>
3. <https://www.anaconda.com/>
4. <https://www.kaggle.com/datasets/mathchi/diabetes-data>
5. <https://www.analyticsvidhya.com/blog/2022/01/diabetes-prediction-using-machinelearning/>