DIABETES PROGNOSTICATION USING MACHINE LEARNING

A DESIGN PROJECT REPORT

Submitted by

GANESH KUMAR .G (730920104030)

HEMANTH KUMAR .V (730920104035)

NIKHIL KUMAR .K (730920104048)

in partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

EXCEL ENGINEERING COLLEGE (AUTONOMOUS)

KOMARAPALAYAM-637303

ANNA UNIVERSITY:: CHENNAI 600 025

NOV 2023

EXCEL ENGINEERING COLLEGE (AUTONOMOUS)

KOMARAPALAYAM-637 303

ANNA UNIVERSITY: CHENNAI-600 303

BONAFIDE CERTIFICATE

Certified that this project report "**DIABETES PROGNOSTICATION USING MACHINE LEARNING**" is the bonafide work of "**GANESH KUMAR .G (730920104030), HEMANTH KUMAR .V (730920104035), NIKHIL KUMAR .K (730920104048)**" who carried out the project under my supervision.

SIGNATURE	**SIGNATURE**
Dr. P.C. SENTHIL MAHESH M.E., Ph.D.,	**Mrs.S.L.SWARNA M.Tech.,**
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Associate Professor	Assistant Professor
Department of CSE	Department of CSE
Excel Engineering College	Excel Engineering College
Komarapalayam-637 303	Komarapalayam-637 303

Submitted for the University Examination held on___________

INTERNAL EXAMINAR	**EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to those who actuated it, without them it would never have into existence. To them, we lay word of gratitude imprinted within ourselves.

We wish our heartfelt thanks to our respected Founder and Chairman of Excel Group Institutions Prof. **Dr. A. K. NATESAN, M.Com., M.B.A., M.Phil., PhD.,** FTA and Vice Chairman **Dr. N. MATHAN KARTHICK M.B.B.S., M.H.Sc (Diabetology)** for allowing us to have the extensive use of the college facilities to do our project effectively.

We express our sincere gratitude and heartfelt thanks to the respected Principal **Dr. K. BOMMANNA RAJA Ph.D.,** for his encouragement and support to complete the project.

We would like to express our profound interest and sincere gratitude to **Dr. P. C. SENTHIL MAHESH M.E., Ph.D.,** Head of the Department, Department of Computer Science and Engineering for his encouragement and support to complete the project.

We are privileged to express our deep sense of gratitude to Project Supervisor **Mrs. S.L. SWARNA M.Tech.,** Assistant Professor, Department Computer Science and Engineering who gave guidance and support throughout our work and made this as a successful project.

We would like to give our sincere gratitude and heartfelt thanks to our Project Coordinator **Mrs. J. OBURADHA M.E.,** Assistant Professor, Department of Computer Science and Engineering, who gave guidance and support throughout my work and made this as a successful project.

Finally, we thank the Almighty,

Parents, Friends and well Wishers for the moral support throughout the project

ABSTRACT

This research presents a machine learning-based predictive model for diabetes, utilizing a comprehensive dataset encompassing demographic and clinical parameters. Employing advanced analytics and algorithmic learning, the model discerns intricate patterns indicative of diabetes onset.

Through rigorous training and validation processes, the model achieves notable accuracy, demonstrating its potential for early detection. Key risk factors, intricately woven into the model, contribute to its robust predictive capabilities, enhancing reliability. The findings underscore the model's promise as a valuable tool for proactive diabetes management, facilitating timely interventions and personalized care strategies.

This approach not only holds the potential to improve patient outcomes but also alleviates the strain on healthcare systems by promoting preventative measures. The study contributes to the evolving landscape of predictive healthcare analytics, emphasizing the importance of data-driven insights for optimizing diabetes care through personalized and preemptive approaches.

In this project, we have used such a dataset of 769 instances. Random Forest Algorithm is applied on dataset to predict early stage of diabetics. Also we have analyzed the dataset with Decision Tree, Native Baye's andSVM algorithms after applying on the dataset. Among these classification algorithms the Random forest has been found having best

accuracy to project the early stage of Diabetes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVATIONS

1. DM or T2DM: Diabetes Mellitus or Type 2 Diabetes Mellitus
2. HbA1c: Hemoglobin A1c
3. FPG: Fasting Plasma Glucose
4. OGTT: Oral Glucose Tolerance Test
5. BMI: Body Mass Index
6. BP: Blood Pressure
7. HDL-C: High-Density Lipoprotein Cholesterol
8. LDL-C: Low-Density Lipoprotein Cholesterol
9. TG: Triglycerides
10. ADA: American Diabetes Association
11. NIDDK: National Institute of Diabetes and Digestive and Kidney Diseases
12. AI: Artificial Intelligence
13. ML: Machine Learning
14. DL: Deep Learning
15. NN: Neural Network
16. SVM: Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1. Introduction

The introduction serves as the gateway to the research, providing a context and rationale for the study. It outlines the background of the research problem, highlights its significance, and establishes the gap in current knowledge that the study aims to address. Typically, the introduction includes a brief review of relevant literature, emphasizing key studies and findings in the field. It then introduces the research question or objective, articulating the purpose of the study and its potential contributions. The introduction sets the stage for the reader, offering a roadmap for what follows in the subsequent sections. In essence, it serves to engage the audience, establish the research's relevance, and generate interest in the study's outcomes.

1.2. Project Overview

The project overview provides a succinct yet comprehensive glimpse into the research, elucidating its scope, goals, and methodology. This section is instrumental in guiding the reader through the research's landscape, outlining the specific objectives and questions that the study endeavors to address. It acts as a roadmap for the ensuing sections, offering a structured preview of the research journey.

The project overview also touches on the broader significance of the research, hinting at its potential implications in the field or its practical applications. By encapsulating key elements of the study, the project overview acts as a crucial bridge, linking the introduction to the more detailed sections that follow, and inviting the reader into the heart of the research endeavor.

1.3. Problem Statement

The problem statement addresses the rising challenge of diabetes prevalence and the imperative for early detection. Recognizing the limitations in current diagnostic approaches, this research aims to develop a precise predictive model. The urgency stems from the need to enhance timely interventions and mitigate the escalating impact of diabetes. By pinpointing this specific problem, the study seeks to contribute to improved healthcare strategies and outcomes in the realm of diabetes management.

1.4. Motivation of Diabetes Prediction

The motivation section of a research project elucidates the driving force behind undertaking the study. It articulates the reasons, significance, and potential impact of the research, aiming to inspire interest and convey its broader relevance. Motivation often stems from identifying gaps in current knowledge, addressing practical challenges, or fulfilling a broader societal need. This section provides context for why the research is timely, essential, or innovative, fostering a deeper understanding of its purpose and implications. In essence, it serves to answer the question: why is this research important and worth pursuing.

1.1.1. Diversity Challenge

The diversity challenge addressed in this research centers on the underrepresentation of diverse perspectives in the existing body of knowledge. Recognizing the importance of inclusivity, the study aims to investigate and address this gap, emphasizing the need to incorporate a broader range of voices and experiences. By highlighting the diversity challenge, the research seeks to contribute to a more comprehensive and representative understanding of the subject matter, fostering inclusivity in both research methodologies and outcomes. This focus on diversity is crucial for enriching scholarly discourse and ensuring that research findings are applicable and beneficial across diverse contexts.

1.1.2. Inclusivity and Audience Diversity

The focus on inclusivity and audience diversity in this research acknowledges the imperative of accommodating a broad range of perspectives and engaging a diverse audience. Recognizing that inclusivity extends beyond research participants to readership, the study aims to adopt methodologies and communication strategies that resonate across various demographic and cultural contexts. By embracing diversity in both content and dissemination, the research seeks to ensure its accessibility and relevance to a wide audience. This commitment to inclusivity aligns with the broader goal of fostering a more equitable and representative academic discourse, enhancing the impact and applicability of the research findings across diverse communities.

1.1.3. Democratizing Diabetes Knowledge

The objective of democratizing diabetes knowledge in this research is to break down barriers to access and understanding. Recognizing disparities in healthcare information, the study aims to create a more inclusive dissemination strategy. By employing accessible language and diverse communication channels, the research seeks to empower individuals across various educational backgrounds and socio-economic levels with comprehensive diabetes insights. This democratization of knowledge aligns with the broader goal of promoting health literacy and proactive management.

1.1.4. Technological Innovation

The emphasis on technological innovation in this research underscores a commitment to advancing methodologies and tools in the study of diabetes. Recognizing the transformative potential of technology, the research aims to integrate cutting-edge approaches such as machine learning and data analytics. By leveraging these innovations, the study seeks to enhance the accuracy and efficiency of diabetes prediction and management. The incorporation of technology not only signifies a commitment to staying at the forefront of scientific inquiry but also anticipates practical applications that may revolutionize healthcare practices.

1.1.5. Clarity and Formal Accuracy

The emphasis on clarity and formal accuracy in this research underscores a commitment to precision and transparency in communication. Recognizing the importance of effective scholarly discourse, the study prioritizes clear and unambiguous language, ensuring that concepts are communicated with accuracy and coherence. By adhering to formal standards, the research aims to enhance the credibility and reliability of its findings. This commitment to clarity and formal accuracy aligns with the broader goal of contributing to a rigorous academic environment, where ideas are conveyed transparently, fostering a deeper understanding and trust in the research outcomes.

1.1.6. Expanding Reach and Impact

The objective of expanding reach and impact in this research highlights a commitment to extending the influence of the study beyond academic confines. Acknowledging the broader societal relevance, the research aims to employ strategies for widespread dissemination and application of its findings. This includes outreach to diverse communities, policymakers, and healthcare practitioners. By maximizing accessibility and relevance, the research seeks to amplify its impact on public awareness, healthcare practices, and policy decisions related to diabetes. The focus on expanding reach aligns with the overarching goal of ensuring that the research contributes meaningfully to positive change and improvement in diabetes management on a broader scale.

1.5. Project Scope

The project scope delineates the boundaries and parameters within which the research will be conducted. It defines the extent of the study in terms of subject matter, time frame, and geographical or conceptual limits. This section provides a clear outline of what the research will and will not cover, managing expectations for both researchers and readers. By establishing a well-defined scope, the research aims to maintain focus and relevance, ensuring that the study remains manageable and feasible within the given constraints.

1.5.1. Diabetes Prediction

The project's primary focus is on diabetes prediction, involving the development, evaluation, and refinement of predictive models. This encompasses the exploration of diverse machine learning algorithms, thorough analysis of datasets, and rigorous processes for model training and validation. The scope extends to identifying and integrating key risk factors to enhance the precision of predictions. It's important to clarify that the study is specifically centered on predictive aspects and doesn't delve into comprehensive diabetes management or treatment strategies. The temporal scope emphasizes contemporary methodologies, incorporating recent advancements in predictive analytics for diabetes. By maintaining a welldefined focus on diabetes prediction.

1.5.2. User-Friendly Interface

In this segment, the focus lies on creating a user-friendly interface and ensuring the seamless integration of tools. The research aims to optimize the accessibility and usability of interfaces while effectively incorporating diverse tools into the workflow. This encompasses a detailed exploration of design principles, user experience considerations, and the integration of various tools to enhance overall efficiency. The study emphasizes the importance of a harmonious interaction between users and tools, fostering a more intuitive and streamlined content management process.

1.5.3. Ensuring Quality Standards

This section outlines the project's commitment to maintaining and assuring quality standards throughout the research process. It encompasses rigorous methodologies, validation procedures, and adherence to established criteria. The scope emphasizes a comprehensive approach to quality assurance, encompassing data integrity, analysis accuracy, and overall research robustness.

1.6. Project Objectives

The project objectives articulate the specific, measurable, achievable, relevant, and time-bound (SMART) goals that the research aims to achieve. These objectives serve as a roadmap, outlining the key milestones and accomplishments the study seeks to attain. They provide a clear and focused framework for the research process, guiding the implementation of methodologies and the interpretation of results. By establishing concrete objectives

1.6.1. Enhance Accessibility

The project objective is to optimize the accessibility of materials related to diabetic prediction by developing a user-friendly interface. This involves creating clear and comprehensible documentation, resources, and tools tailored for diverse audiences. The primary goal is to facilitate ease of understanding and utilization, especially for individuals involved in or seeking information about diabetic prediction. This objective aligns with the broader aim of democratizing diabetes knowledge, making project materials more inclusive, and ensuring that stakeholders can engage effectively with predictive tools for proactive diabetes management.

1.6.2. Help People Understand

The project objective is to develop resources and tools that simplify complex concepts related to diabetes prediction, aiding individuals in understanding the intricacies of the field. The primary goal is to create accessible materials that bridge knowledge gaps and promote comprehension, ensuring that information about diabetes prediction is readily understandable for a diverse audience. This objective aligns with the overarching mission of enhancing health literacy and contributing to proactive diabetes management by empowering individuals with clear and digestible information.

1.6.3. Various Document Types

The project objective is to develop a system capable of efficiently handling various types of documents related to diabetes prediction. This includes but is not limited to research papers, medical reports, and data sheets. The goal is to create a versatile platform that accommodates diverse document formats, streamlining the process of information extraction and analysis for more comprehensive insights into diabetes prediction. This objective aligns with the project's aim of enhancing content handling and underscores the commitment to accommodating the varied nature of materials relevant to predictive healthcare analytics.

1.6.4. Capture Problem

The project objective is to capture and articulate the problem with utmost precision. This involves a detailed analysis and documentation to clearly define the parameters, challenges, and nuances of the issue at hand. The primary goal is to create a comprehensive understanding of the problem, providing a solid foundation for the subsequent phases of research and development.

1.6.5. Diabetes

The project objective is to contribute to a deeper understanding of diabetes. This involves researching and developing resources to clarify the complexities of diabetes, ensuring that individuals gain a comprehensive knowledge of the condition. The primary goal is to create educational materials that promote awareness, prevention, and effective management strategies. This objective aligns with the overarching mission of improving health literacy and fostering proactive measures in diabetes care, ultimately contributing to better-informed individuals and communities in the realm of diabetes.

1.6.6. Keep it Simple

The project objective is to streamline processes and communication, emphasizing simplicity in all aspects. This involves developing userfriendly interfaces, clear documentation, and straightforward communication strategies. The primary goal is to enhance accessibility and understanding by minimizing complexity.

1.6.7. Prevent Co-Occurring Diseases

The project objective is to develop strategies that mitigate the risk of additional health issues alongside diabetes. This involves researching and implementing preventive measures to avoid the onset of other

CHAPTER 2

LITERATURE SURVEY

2.1. A survey on diabetes risk prediction

Authors: J Family Med Prim Care. 2022 Nov;

Diabetes mellitus (DM) is a chronic condition that can lead to a variety of consequences. Diabetes is a condition that is caused by factors such as age, lack of exercise, sedentary lifestyle, family history of diabetes, high blood pressure, depression and stress, poor food, and so on. Diabetics are at a higher risk of developing diseases such as heart disease, nerve damage (diabetic neuropathy), eye problems (diabetic retinopathy), kidney disease (diabetic nephropathy), stroke, and so on. According to the International Diabetes Federation, 382 million people worldwide suffer from diabetes. By 2035, this number will have risen to 592 million. Every day, a large number of people become victims, and many are ignorant whether they have it or not..

2.2. Machine learning for predicting diabetes risk in western China adults

Authors :Lin Li, Yinlin Cheng, Weidong Ji, Mimi Liu,
Zhensheng ,Yining Yang

Diabetes mellitus (DM) is a metabolic disease characterized by hyperglycemia. Hyperglycemia can cause chronic damage to tissues over time [1]. Diabetes has become a major health problem worldwide with a significant increase in DM patients. According to the International Diabetes Federation (IDF), approximately 537 million adults worldwide had diabetes in 2021 (with a prevalence of 10.5%), and it is estimated that by 2045, approximately 783 million people worldwide are likely to have diabetes

(with a prevalence of approximately 12.2%) [2, 3]. According to a survey, because individuals with type-2 diabetes mellitus (T2DM) usually lack the relevant knowledge, or they are asymptomatic, some individuals with T2DM patients can not be detected in time (approximately 50% of individuals with T2DM are undiagnosed) [3, 5]. It is necessary to identify individuals with diabetes in the population in an efficient and accurate manner.

2.3. A survey on prediction of diabetes using classification algorithms

Authors: A. Khanwalkar, R. Soni.

Diabetes is a chronic disease that pays for a large proportion of the nation's healthcare expenses when people with diabetes want medical care continuously. Several complications will occur if the polymer disorder is not treated and unrecognizable. The prescribed condition leads to a diagnostic center and a doctor's intention. One of the real-world subjects essential is to find the first phase of the polytechnic. In this work, basically a survey that has been analyzed in several parameters within the polyinfected disorder diagnosis.

2.4. Machine learning for prediction of diabetes risk in middle-aged Swedish people

Authors: Lara Lama a, Oskar Wilhelmsson a, Erik Norlander a, Lars Gustafsson

The health care sector needs better opportunities for individualized support for both the patient and the healthcare staff. Program 4D developed during 2012–2017 as a project focusing on type 2 diabetes (T2D) as a collaboration between Karolinska Institutet and Stockholm County Council that is in charge of most health care within the county. It included a process of screening for T2D, a standardized care process to support the healthcare staff, and a specific digital support for patients and healthcare professionals were developed. The functions specified in this project are now standard routine in e-health solutions and are implemented in commercially available solutions. In this and similar ehealth solutions, healthcare professionals and the patient jointly set up a personalized interactive healthcare plan with individually tailored activities, where the patient's measurements and activities are reported.

2.5. A Comprehensive Review of Various Diabetic Prediction Models

Author: Roshi Saxena , Sanjay Kumar Sharma, Manali Gupta.

Diabetes is a chronic disease characterized by a high amount of glucose in the blood and can cause too many complications also in the body, such as internal organ failure, retinopathy, and neuropathy. According to the predictions made by WHO, the figure may reach approximately 642 million by 2040, which means one in a ten may suffer from diabetes due to unhealthy lifestyle and lack of exercise. Many authors in the past have researched extensively on diabetes prediction through machine learning algorithms.

CHAPTER 3

SYSTEM DESIGN

3.1. System Design

In the context of diabetes prediction, the system design phase involves structuring a comprehensive framework for the predictive model. This encompasses defining the architecture, components, and interactions essential for accurate diabetes prediction. The design outlines the data flow, processing mechanisms, and integration of machine learning algorithms. Considerations include the scalability of the model, adaptability to diverse datasets, and user-friendly interfaces for healthcare professionals. Diagrams and flowcharts are employed to illustrate the system's structure and functionality.

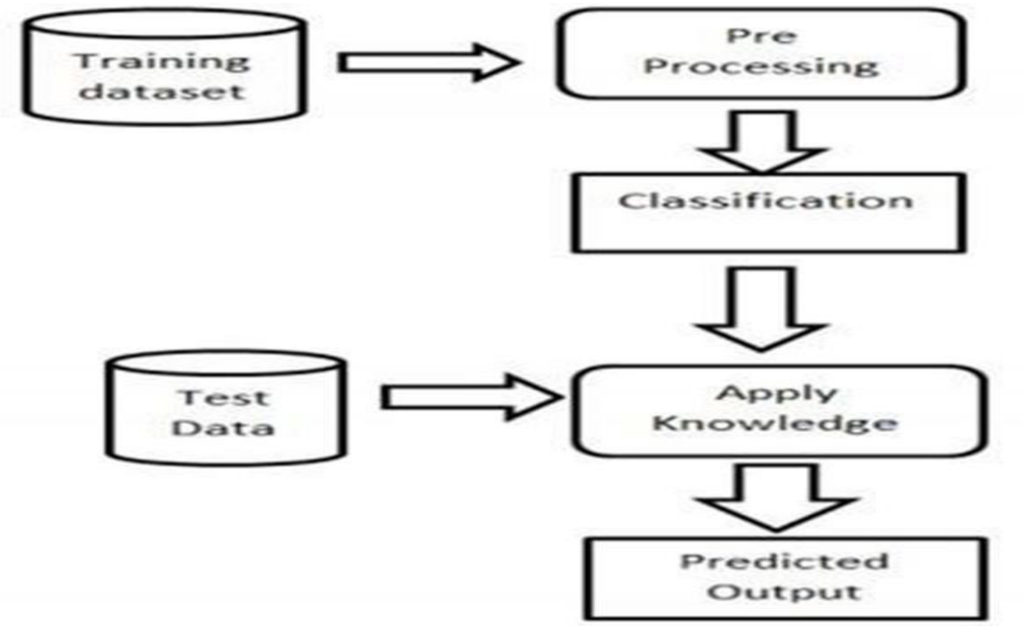


Fig 3.1. System Design

3.1.1. Datasets Collection

We cannot create a huge dataset which can translated document file and huge dictionary of word based on user preferences, so collect or buy a substantial amount of resources materials and texts in English along with their translations in the target Indian regional languages (known as datasets).

Steps in datasets collections are:

a) Define Data Requirements

Clearly outline the objectives of data collection, specifying the information needed to achieve the project's goals, such as patient demographics, medical history, and lifestyle factors. **b) Select Source Language Materials**

Enumerate and prioritize the variables critical to diabetes prediction, including blood glucose levels, BMI, age, family history, and lifestyle habits.

c) Review Existing Datasets:

Explore publicly available healthcare datasets, research repositories, and clinical databases to identify potential sources that align with the project's scope.

d) Consider Ethical and Legal Aspects:

Ensure compliance with ethical standards and data protection regulations. Obtain necessary approvals and permissions before accessing or using any datasets.

e) Define Inclusion and Exclusion Criteria:

Clearly define the criteria for including or excluding data points, ensuring relevance and consistency across the dataset.

f) Select Data Collection Methods:

Choose appropriate methods for data collection, which may include electronic health records, surveys, wearable devices, or a combination of sources.

g) Ensure Data Quality:

Implement measures to maintain data accuracy and completeness, addressing issues such as missing values, outliers, and inconsistencies. **h) Secure Data Access:**

Establish protocols for data access, ensuring that only authorized individuals or systems can retrieve, process, and analyze the collected datasets.

i) Consider Longitudinal Data:

If relevant to the project, consider collecting longitudinal data to observe changes and trends over time, providing a more comprehensive view of health parameters.

j) Document Data Sources:

Maintain a detailed record of the datasets, including information on the source, collection methods, date of acquisition, and any preprocessing steps applied, to facilitate transparency and reproducibility.

3.1.2. Engine Creation:

Translation engine is the core part of the software and it is developed using machine learning and NLP (natural language processing) techniques. We train the translation engine with bilingual data (which we have collected) for improved accuracy.

Steps in translation engine creation are:

a) Algorithm Selection:

Choose suitable machine learning algorithms for diabetes prediction, considering factors such as decision trees, support vector machines, or neural networks based on the project's requirements.

b) Feature Integration:

Integrate relevant features identified during the feature selection phase into the predictive model, ensuring that the chosen variables contribute effectively to accurate predictions.

c) Data Preparation:

Preprocess the collected datasets, addressing missing values, normalizing numerical features, and encoding categorical variables to create a clean and standardized input for the machine learning model.

d) Model Training:

Train the selected machine learning model using the preprocessed datasets. This involves exposing the model to historical data to learn patterns and relationships that can later be applied to new, unseen data..

e) Hyperparameter Tuning:

Optimize the hyperparameters of the chosen machine learning algorithm to enhance the model's performance. This iterative process involves adjusting parameters to achieve the best predictive accuracy.

f) Validation and Cross-Validation:

Validate the model using a separate dataset not used during training to assess its generalization ability. Implement cross-validation techniques to ensure robustness and minimize overfitting.

g) Evaluation Metrics:

Utilize relevant evaluation metrics such as accuracy, precision, recall, and F1 score to assess the model's performance and its ability to correctly predict instances of diabetes.

3.1.3. User Interface Design:

User Interface Design focuses on creating an intuitive and visually appealing interaction platform. By understanding user needs and preferences, the design process involves sketching, wireframing, and developing a cohesive visual design with responsive elements. Navigation is designed to be user-friendly, and information architecture ensures logical organization. Interactive elements enhance engagement, and accessibility features cater to diverse user needs. Usability testing and prototyping validate the design, and iterative improvements incorporate user feedback. Consistency across platforms and scalability are prioritized, and comprehensive documentation ensures design guidelines for future reference and updates.

3.1.4. Content Analysis:

Systematically examining and interpreting textual, visual, or multimedia content to extract meaningful insights. This process employs various methodologies, including coding and categorization, to identify patterns, themes, or sentiments within the content. By scrutinizing and organizing information, content analysis provides a structured approach to extract valuable information, facilitating a deeper understanding of the subject matter. This phase is crucial for deriving meaningful conclusions, informing decision-making, and contributing to the overall knowledge base in the context of the project or research.

3.1.5. Quality Assurance:

Quality Assurance is a meticulous process ensuring that every aspect of the project meets predefined standards. It involves systematic checks, validations, and testing procedures to guarantee the reliability, functionality, and overall excellence of the project deliverables. Quality assurance encompasses various aspects, including code reviews, documentation accuracy, and adherence to project requirements.

a) Define Quality Metrics:

Quality metrics are quantifiable measures used to assess the excellence and compliance of a project or product against predefined standards.

b) Automated Assessment:

Automated Assessment involves the use of automated tools and algorithms to evaluate, analyze, and provide feedback on various aspects of a system or project. This process efficiently assesses performance, code quality, and adherence to predefined standards, enhancing accuracy and speed in evaluations.

c) Manual Assessment:

Manual assessment for diabetic prediction involves human-driven evaluation of data, models, and results to ensure accuracy, relevance, and clinical validity.

d) Feedback Loop:

continuous process where outputs of a system are circled back as inputs, enabling ongoing refinement and improvement. In various contexts, including software development.

e) Domain-Specific Review:

Domain-Specific Review tailors evaluations to industry standards, regulations, and unique requirements, ensuring project alignment and effectiveness within a specific domain.

f) Cultural Sensitivity Check:

A Cultural Sensitivity Check involves assessing content, communication, or actions to ensure they respect and align with diverse cultural perspectives, promoting inclusivity and avoiding unintentional offense.

g) Prediction Validation:

Prediction validation is the process of systematically assessing and confirming the accuracy, and effectiveness of predictions made by a model.

h) Final Review:

The Final Review is a comprehensive assessment conducted before project completion, ensuring all requirements are met. It identifies any remaining issues and confirms alignment with project goals, serving as a final quality check before deployment.

3.1.6. Simple prediction Generation:

Simple Prediction Generation involves the creation of straightforward predictive models to generate basic forecasts or insights based on available data. This phase focuses on developing models with simplicity and efficiency, often using fundamental algorithms, to quickly generate predictions and insights for initial assessments or basic decision-making processes.

3.1.7. Performance Optimization:

Performance Optimization is the process of refining and enhancing the efficiency, speed, and overall effectiveness of a system or application. This involves identifying bottlenecks, improving algorithms, and streamlining processes to achieve optimal performance, ensuring that the system operates at its highest potential and meets specified performance criteria. This phase is crucial for delivering a high-performing and responsive solution, enhancing user experience and overall system functionality.

3.1.8. Data Verification:

Data Verification is the process of confirming the accuracy, integrity, and reliability of collected data. This involves cross-referencing information against authoritative sources, checking for completeness and consistency, and ensuring that the data aligns with predefined standards. The goal of data verification is to enhance the overall quality of datasets, reduce errors, and ensure that the information is trustworthy and suitable for analysis or decision-making processes.

3.1.9. Deployment and User Support:

Deployment involves the seamless rollout of the solution into the live environment, ensuring accessibility for users. Simultaneously, user support encompasses comprehensive training, technical assistance, and continuous monitoring to facilitate a smooth post-deployment experience. This includes ongoing feedback collection, scalability planning, security measures, and adaptation to user insights for sustained system efficiency and user satisfaction.

3.2. Implementation issues and challenges

In the implementation phase several issues and challenges may arise:

a). Technical Hurdles:

Overcoming technical complexities, such as integration challenges, system compatibility issues, or unforeseen technical constraints during implementation.

b) Resource Allocation:

Efficiently allocating resources, including human resources, time, and budget constraints, to meet project timelines and objectives.

c) Adaptation to Change:

Adapting to changes in project requirements, evolving technologies, or unforeseen external factors that may impact the implementation plan.

d) Stakeholder Collaboration:

Ensuring effective collaboration and communication among stakeholders to address concerns, gather feedback, and align expectations during the implementation process.

e) Quality Assurance:

Implementing robust quality assurance measures to detect and rectify any defects or issues that may arise during the implementation, ensuring a highquality final product.

f) Data Migration Challenges:

Managing the migration of data from existing systems to the new solution, addressing potential data inconsistencies, and ensuring data integrity.

g) User Training and Adoption:

Facilitating user training and ensuring a smooth transition to the new system, addressing any resistance or challenges in user adoption.

h) Scalability Concerns:

Planning for scalability and addressing potential challenges related to system growth and increased user load over time.

i) Security Measures:

Implementing robust security measures to safeguard the system from potential threats and ensuring data privacy and integrity.

j) Regulatory Compliance:

Navigating and adhering to relevant regulatory requirements and compliance standards applicable to the implementation domain.

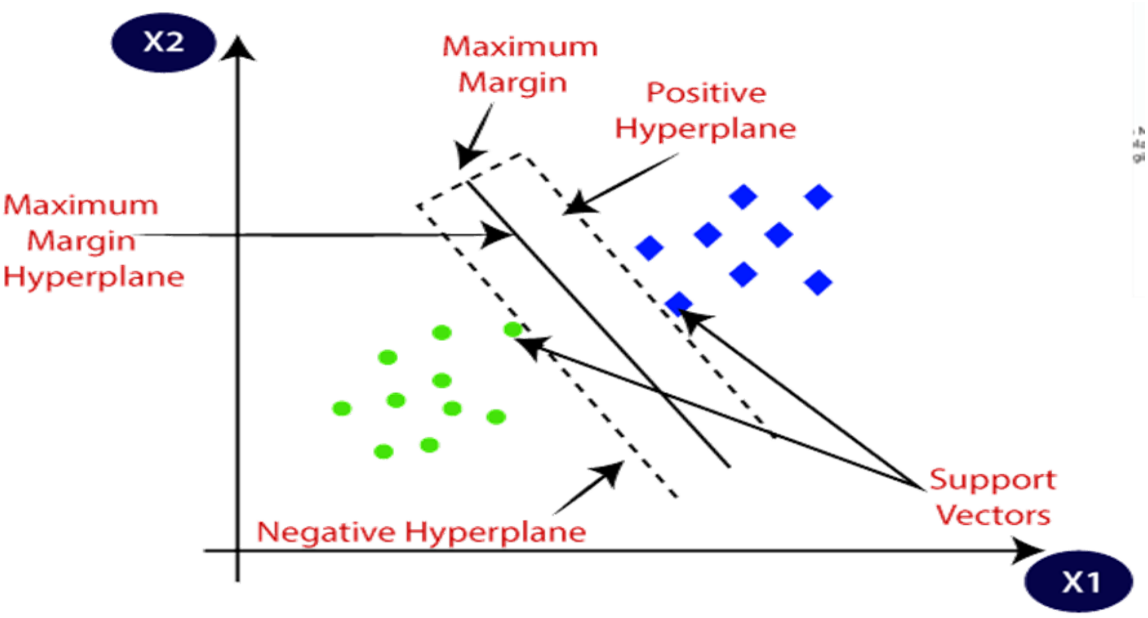


Fig 3.2. Support vector Machine

CHAPTER 4

METHOD AND METHODOLOGIES INVOLVED

4.1. Methodology

Methodology refers to the systematic approach or set of principles and procedures used to conduct research, implement projects, or solve problems. It outlines the steps, techniques, and tools employed to achieve specific objectives, ensuring a structured and organized process. A welldefined methodology provides a framework for planning, executing, and evaluating activities, contributing to the reliability and validity of outcomes in various fields such as research, software development, or project management.

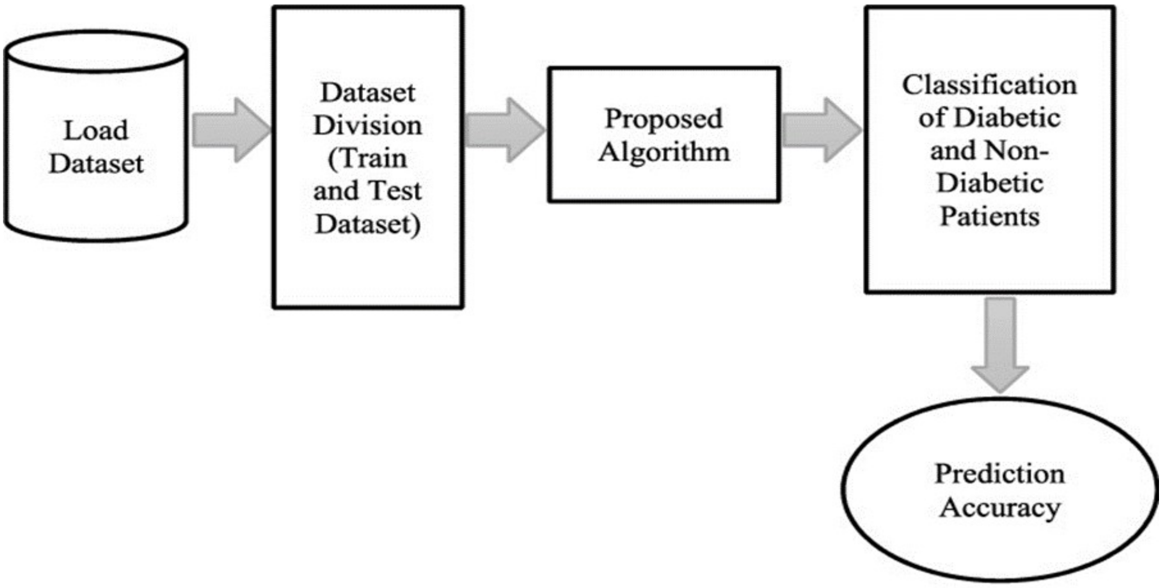


Fig 4.1. General flow chat for diabetics diagnosis

4.1.1. Planning:

Planning is a foundational phase that involves defining project goals, breaking down tasks, allocating resources, assessing risks, developing timelines, fostering collaboration, budgeting, setting quality standards, preparing for contingencies, and maintaining comprehensive documentation. This strategic organization ensures a systematic and wellcoordinated approach to project execution, fostering successful outcomes and effective problem-solving.

4.1.2. Requirement Analysis:

Requirement analysis for a diabetes prediction project involves engaging with stakeholders, defining data collection needs, ensuring regulatory compliance, specifying prediction model features, integrating with healthcare systems, designing a user-friendly interface, setting accuracy and reliability criteria, ensuring scalability and performance, defining training and support requirements, and incorporating a feedback mechanism. This comprehensive process ensures that the project aligns with user expectations, regulatory standards, and performance benchmarks.

4.1.3. Designing:

Designing is the creative process of conceptualizing, prototyping, and crafting the structure and visual elements of a system or product. It involves user experience (UX) and user interface (UI) design, functional definition, iterative refinement, ensuring compatibility and responsiveness, compliance with standards, collaborative efforts, and comprehensive documentation.

4.1.4. Building:

Building is the transformative phase where designed concepts materialize into functional entities through activities like coding, construction, manufacturing, assembly, testing, debugging, integration, quality control, and adherence to specifications. This dynamic process ensures the tangible realization of projects and products, culminating in deployment for end-users.

4.1.5. Testing:

Testing is a crucial phase in the development process where various aspects of a system or product are systematically evaluated. This includes checking individual components (Unit Testing), verifying their integration (Integration Testing), confirming features and functionalities (Functional Testing), assessing performance under different conditions (Performance Testing), addressing security vulnerabilities (Security Testing), validating user expectations (User Acceptance Testing), ensuring existing functionalities remain unaffected by changes (Regression Testing), implementing automated scripts for efficiency (Automated Testing),

4.2. Hardware and Software Requirement

4.2.1. Hardware Requirements

Hardware requirements specify the necessary physical components and specifications for a system to operate efficiently, encompassing details such as processors, memory, storage, graphics processing units, and connectivity options. Clear and accurate hardware requirements are essential for proper system functioning, user experience, and guiding users in selecting compatible configurations.

4.2.2. Software Requirements

- **Operating system** : Windows 8 / 10
- **Programming Language** : Python
- **DL Libraries** : Numpy, Pandas

4.3. Language specification

4.3.1. Python

Python is an easy to learn, powerful programming language. It has efficient high- level data structures and a simple but effective approach to object-oriented programming. Python’s elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting

- **Python is Interpreted** – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

- **Python is Interactive** – you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented** – Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

- **Python is a Beginner's Language** – Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

4.3.2. Jupyter Note Book :

Jupyter Notebook is the one of the applications of Anaconda Navigator .Anaconda Navigator has several other applications like CMD.exe Prompt, Datalore, IBM Watson Studio Cloud, Jupyter Lab, Jupyter Notebook, Powershell Prompt, Qt Console, Spyder ,Glueviz, Orange 3 ,PyCharm Professional and RStudio. These applications are useful for various developing and designing various systems.

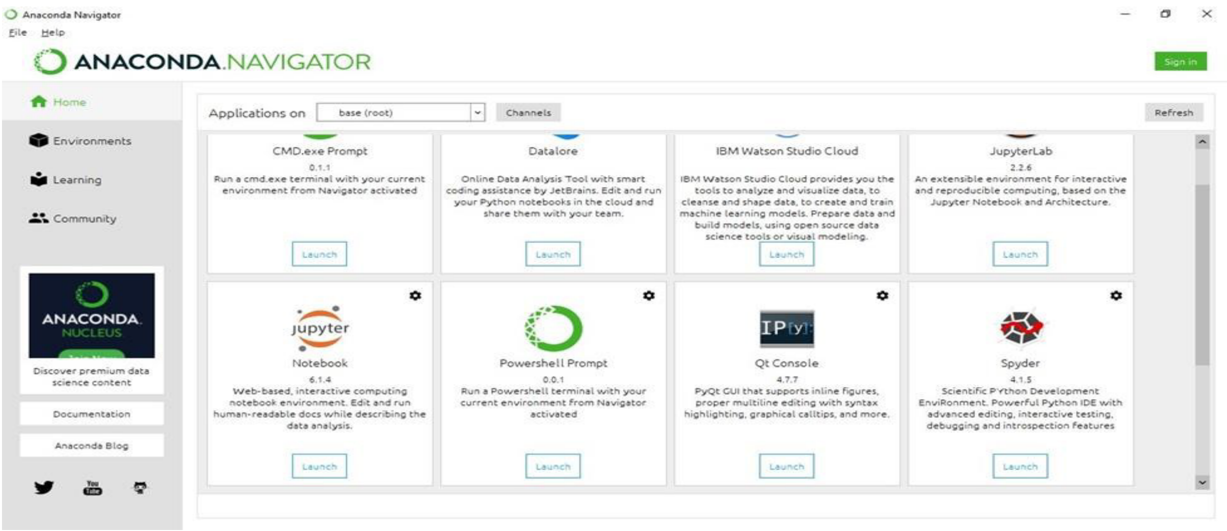


Fig 4.3.2. Anaconda Navigator Home page with various application.

The Anaconda software is Open Source, which means the code, the schematics, design, etc. are all open for anyone to take freely and do what they like with it. This means there is nothing stopping anyone from taking the schematics and anyone can contribute to the open source. The main purpose of open source anyone contribute the code. This increases the reusability of the code.

4.3.3. CMD.exe Prompt:

Run a cmd.exe terminal with your current environment from Navigator activated

4.3.4. Anaconda Navigator Environments:

The Navigator consists of Environment tab which specifies various environments available.

There were of four types of packages possible

- Installed
- Not Installed
- Updatable

- Selected

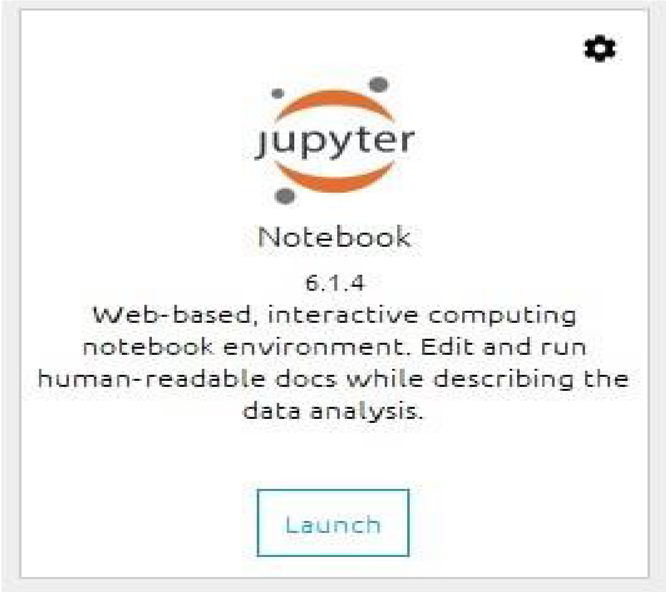


Fig 4.3.4. Jupyter Notebook Launch

Jupyter NoteBook:

Jupyter notebook is a web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis. Since it is an web-based application it will be launched from anaconda navigator .

The main page of Jupyter notebook contains three functionalities Files, Running and Clusters **Files**: In the files we can access all the system and we can upload a new file or create a new file. we will also have details like Name, Last Modified, File Size.

Running: In the running we can see the terminals and notebooks that are currently running.

Clusters: Clusters tab is now provided by IPython parallel.

When we open the required file for working the tool bar contains various operations.

Names of toolbars in above image in the order

1)Save and Check Point 2) Insert cell below

3)Cut selected cells 4) Copy Selected cells

5) Paste cells below 6) move selected cells up

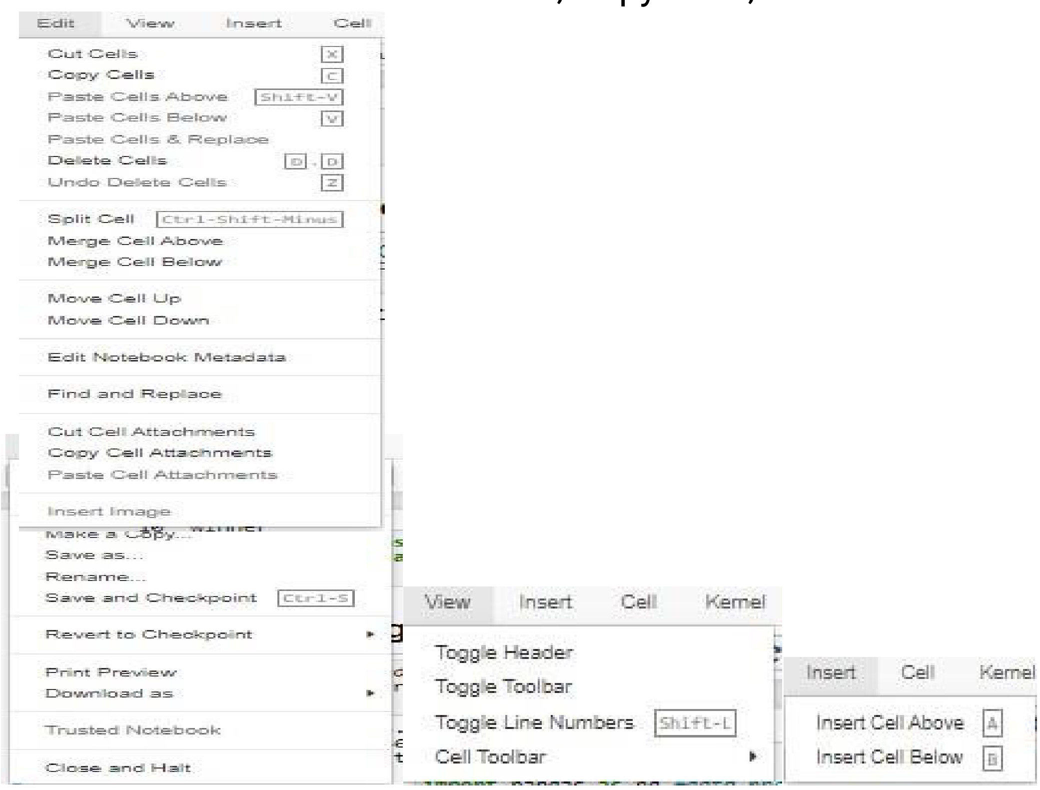
7) Move selected cells down and select below 8) Run current cell

9)Interrupt the kernel 10) Restart the kernel 11)Restart the kernel and re-run the whole note book 12) open the command pallete The table shows the functionalities of the various toolbars.

4.3.5. Functionalities of various toolbars

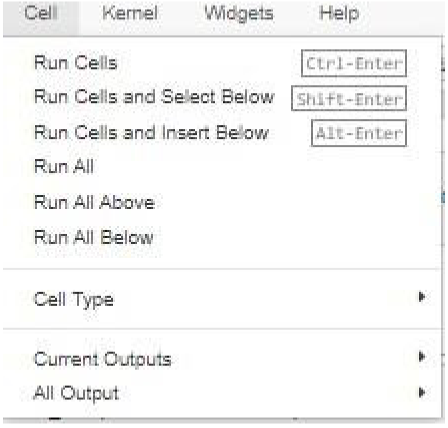
It also provides some more functionalities through menu bar.

- File menu contains functionalities like New Notebook, open, make a copy, save as ,Rename , Revert check point etc.
- Edit menu consists of Cut Cells, Copy Cells, Delete Cells etc.



- View menu consist of Toggle header , Toggle toolbar , Toggle Line Numbers, Cell Toolbar
- Insert menu consists Insert cell above, Insert Cell Below
- Run menu consists of run cells, run cells and select below. run cells and insert below , run all

, run all above , run all below , cell type , current outputs , all output



- Kernel menu contains Interrupt, restart, restart and clear output , restart and run all, Reconnect, shutdown.

CHAPTER 5

SOFTWARE TESTING

5.1. TESTING

Software testing plays a crucial role in offering an unbiased and independent evaluation of software, enabling businesses to comprehend and assess the risks associated with software implementation. Test techniques encompass the systematic execution of a program or application with the aim of identifying software bugs, errors, or defects, and validating the software product's suitability for use..

In the current project, six types of testing are conducted, each contributing to a comprehensive evaluation of the software's performance and robustness.

There are six kinds of testing is done in our project.

1. Unit testing
2. Load Testing
3. Beta Testing
4. Acceptance Testing
5. System Testing

5.1.1. Unit Testing

Unit testing is a critical phase in software development, focusing on evaluating individual units or components of an application. As an application comprises various units and modules, detecting errors at the unit level is relatively straightforward and consumes less time, given the smaller scale of these components.

5.1.2. Load testing

Load testing is a crucial process involving the application of demand to a system or device to assess and measure its response. In the professional application testing community, the term "load testing"

5.1.3. Beta Testing

Load testing is a vital facet of application testing, where demand is applied to assess a system's response. Within the professional testing community, "load testing" typically denotes simulating anticipated usage scenarios for an application program. This process helps gauge the system's performance under various workloads and user interactions, ensuring stability and reliability. Load testing is crucial for identifying performance bottlenecks, predicting scalability, and optimizing resource allocation.

5.1.4. Acceptance Testing

In engineering and its sub-disciplines, acceptance testing is a crucial evaluation conducted to ascertain whether the stipulations of a specification or contract are fulfilled. This testing may encompass chemical, physical, or performance assessments. When customers undertake acceptance testing for applications, it is termed user acceptance testing, end-user testing, site testing, or field testing. This process ensures that the product or system aligns with the agreed-upon criteria, meeting user expectations and adhering to specified standards. User acceptance testing is pivotal in validating the functionality and performance of engineering solutions, playing a key role in ensuring the successful implementation of projects across various domains.

5.1.5. System Testing

System testing, alternatively known as system-level testing or system integration testing, is a pivotal phase in quality assurance. During this process, a QA team systematically assesses the interactions among diverse components within an application within the complete, integrated system. The primary objective is to ensure seamless collaboration among different elements, validating that the integrated system or application functions cohesively and meets specified requirements. System testing encompasses a comprehensive evaluation of the overall system, focusing on its performance, functionality, and interoperability.

a) **Improved product quality:**

The exhaustive system testing process significantly enhances product quality. As an integrated system undergoes scrutiny through multiple test sets in the product development cycle, it offers insights into the product's capability to function effectively across diverse platforms and environments. This comprehensive evaluation ensures that the product not only meets individual component specifications but also excels in the broader context of the entire system.

b) **Error reduction:**

In the development of complex systems, errors are inevitable, and system testing plays a crucial role in identifying these issues. This phase verifies the system's code and functionality against its specified requirements. Importantly, system testing serves as a final check, revealing errors that might not have been detected during earlier stages like integration and unit testing.

c) **Cost savings:**

Detecting system defects later in the project lifecycle can be more time-consuming to address. Timely and continuous system testing not only helps in reducing unexpected costs and project delays but also grants project managers better control over budgets.

e)Security:

Systematic testing helps ensure that the tested system is free from weaknesses that could pose risks to end users and system data. By identifying and addressing potential threats during the testing phase.

f) Customer satisfaction:

System testing provides crucial visibility into the stability of a product at every stage of development. This transparency not only builds customer confidence but also enhances the overall user experience. By systematically assessing the product's functionality, performance, and integration, system testing ensures that potential issues are identified and addressed early in the development process.

h) Easier code modification:

system testing is a critical phase in software development that plays a vital role in identifying code problems. Detecting and rectifying issues during the system testing phase is generally more efficient than

System testing includes various testing which are as:

Performance testing:

Performance testing measures the speed, average load time, stability, reliability and peak response times of the system under various conditions. It's typically coupled with stress testing and may include both hardware and software testing tools.

Usability testing:

These are tests to evaluate if a system is easy to use and functional for the end user. Metrics, including user error rates, task success rates, the time it takes a user to complete a task and user satisfaction, are used during testing.

Load testing:

This is testing to determine how a system or software performs under a real-life extreme load and test scenarios. Metrics, such as throughput, number of users and latency, are measured through this testing.

Regression testing:

Also known as sanity testing, it ensures that all changes introduced into an application or code during system testing, recent code changes or updates haven't caused any new bugs or issues.

Migration testing:

This is conducted to ensure smooth migration of legacy systems to new systems without disruptions, data loss or downtimes.

Scalability testing:

This measures an application's or system's capability to scale up or down when trying to meet the changing user requirements.

5.2. TESTING METHODOLOGIES

5.2.1. White-Box Testing (GLASS-BOX TESTING)

White-box testing is a test methodology that concentrates on the internal control structures of a program. Test cases are designed to ensure comprehensive coverage, aiming to execute every statement at least once and exercise all logical conditions.

5.2.2. Black-Box Testing:

Black-box testing concentrates on the functional requirements of software, serving as a complementary rather than an alternative approach to white-box techniques. It aims to uncover a different class of errors than white-box methods and specifically targets

5.2.3. Unit Testing:

Unit testing is a verification process that concentrates on the smallest unit of software design. By using the procedural design description, critical control paths are tested within the module's boundaries to uncover errors..

5.2.4. Integration Testing

Integration testing is a systematic technique for constructing the program structure while conducting tests to uncover errors associate with interfacing. The objective is to take unit tested modules and build a program structure that has been dictated by design.

5.3. Verification and Validation:

Validation is a process of finding out if the product being built is right? Whatever the software product is being developed, it should do what the user expects it to do. The software product should functionally do.

CHAPTER 6

SOURCE CODE AND SCREENSHOTS

6.1. Source Code

```
# Importing the Dependencies

import numpy as np import pandas
as pd import matplotlib.pyplot as
plt from sklearn.preprocessing
import StandardScaler from
sklearn.model_selection import
train_test_split from sklearn import
svm
from sklearn.metrics import accuracy_score

# Data Collection and Analysis
# PIMA Diabetes Dataset
# loading the diabetes dataset to a pandas DataFrame

diabetes_dataset =
pd.read_csv('diabetes.csv')
get_ipython().run_line_magic('pinfo', 'pd.read_csv')

 # 1 --> Diabetic
diabetes_dataset.groupby('Outcome').mean()

# separating the data and labels
X =   diabetes_dataset.drop(columns   =   'Outcome',
  axis=1)
Y =   diabetes_dataset['Outcome')   print(X)   print(Y)
  scaler.fit(X) standardized_data = scaler.transform(X)
print(standardized_data)
```

```
X = standardized_data Y
=
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3,
random_state=63) from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier() knn.fit(X_train,y_train) knn.predict(X_test)
score = knn.score(X_train,y_train) score1 = knn.score(X_test,y_test)
print("Accuracy score of KNN Training data is:",score) print("Accuracy
score of KNN Test data is:",score1)

  # Decision Tree Classifier

from sklearn.tree import DecisionTreeClassifier tre =
DecisionTreeClassifier() tre.fit(X_train,y_train) tre.predict(X_test) score=
tre.score(X_train,y_train) score1 =
['Pregnancies','Glucose','BloodPressure','SkinThickness','Insulin','BM
I','Diabetics','Age' ] features = tre.feature_importances_ features
plt.figure(facecolor='y'
)
plt.barh(featur_names,
features) plt.show()

X=standardized_data
Y=diabetes_dataset['Outcome']
X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.2,
stratify=Y, random_state=2) print(X.shape, X_train.shape, X_test.shape)

# Training the
Model       from
sklearn     import
svm
classifier = svm.SVC(kernel='linear')

#training the support vector Machine Classifier
classifier.fit(X_train, Y_train)
```

6.2. IMPLEMENTATION PROCESS SCREENSHOTS

Reading the Data from csv file

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Data distribution

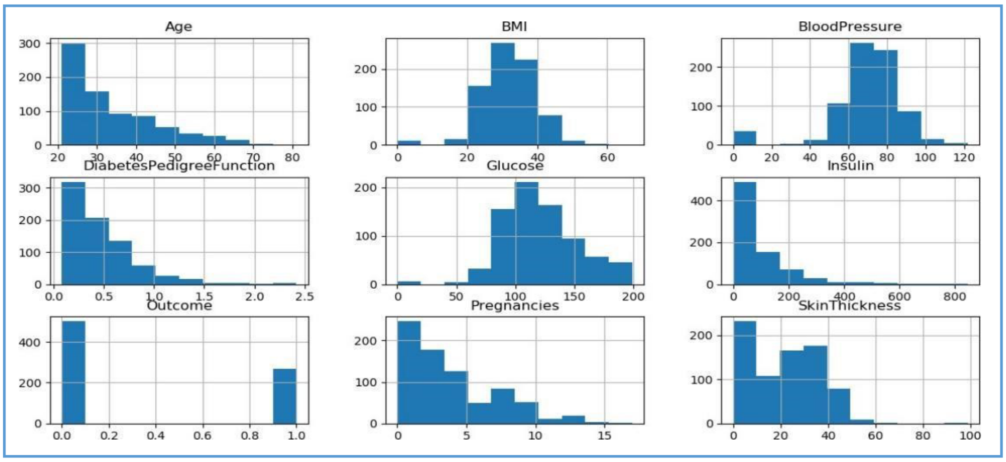


Fig 6.2. Data Distribution

Seperating data and labels

```
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0              6      148             72             35        0  33.6
1              1       85             66             29        0  26.6
2              8      183             64              0        0  23.3
3              1       89             66             23       94  28.1
4              0      137             40             35      168  43.1
..           ...      ...            ...            ...      ...   ...
763           10      101             76             48      180  32.9
764            2      122             70             27        0  36.8
765            5      121             72             23      112  26.2
766            1      126             60              0        0  30.1
767            1       93             70             31        0  30.4

     DiabetesPedigreeFunction  Age
0                       0.627   50
1                       0.351   31
2                       0.672   32
3                       0.167   21
4                       2.288   33
..                        ...  ...
763                     0.171   63
764                     0.340   27
765                     0.245   30
766                     0.349   47
767                     0.315   23

[768 rows x 8 columns]
```

Diabetic or non-diabetic

```
0    500
1    268
Name: Outcome, dtype: int64
```

Preprocess Data

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.348958
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

Standardize the data

```
[[ 0.63994726  0.84832379  0.14964075 ...  0.20401277  0.46849198
   1.4259954 ]
 [-0.84488505 -1.12339636 -0.16054575 ... -0.68442195 -0.36506078
  -0.19067191]
 [ 1.23388019  1.94372388 -0.26394125 ... -1.10325546  0.60439732
  -0.10558415]
 ...
 [ 0.3429808   0.00330087  0.14964075 ... -0.73518964 -0.68519336
  -0.27575966]
 [-0.84488505  0.1597866  -0.47073225 ... -0.24020459 -0.37110101
   1.17073215]
 [-0.84488505 -0.8730192   0.04624525 ... -0.20212881 -0.47378505
  -0.87137393]]
```

Outcome	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	3.298000	109.980000	68.184000	19.664000	68.792000	30.304200	0.429734	31.190000
1	4.865672	141.257463	70.824627	22.164179	100.335821	35.142537	0.550500	37.067164

KNN Classifier

Accuracy score of KNN Training data is: 0.7914338919925512

Accuracy score of KNN Test data is: 0.7186147186147186

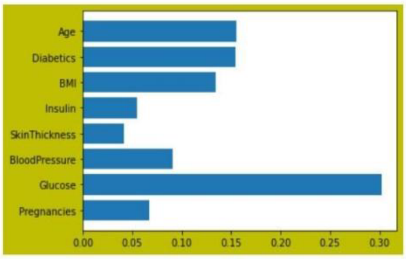
Logistic Regression

Accuracy score of Logistic Training data is: 0.776536312849162

Accuracy score of Logistic Test data is 0.7662337662337663

Decision Tree Classifier

```
Accuracy score of Decision tree Classifier Training data is: 100.0
Accuracy score of Decision tree Classifier Test data is: 69.26406926406926
```



SVM Algorithm

```
Accuracy score of SVM Training data is: 0.7746741154562383
Accuracy score of SVM Test data is: 0.7619047619047619
```

```
print('Accuracy score of the training data : ', training_data_accuracy)
```

```
Accuracy score of the training data :  0.7866449511400652
```

```
# accuracy score on the test data
X_test_prediction = classifier.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
```

```
print('Accuracy score of the test data : ', test_data_accuracy)
```

```
Accuracy score of the test data :  0.7727272727272727
```

6.3. Diabetes Result

Predict wheather person affected with Diabeic or Not

Making a Predictive System

```
input_data = (5,166,72,19,175,25.8,0.587,51)

# changing the input_data to numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the array as we are predicting for one instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the input data
std_data = scaler.transform(input_data_reshaped)
print(std_data)

prediction = classifier.predict(std_data)
print(prediction)

if (prediction[0] == 0):
  print('The person is not diabetic')
else:
  print('The person is diabetic')
```

```
[[ 0.3429808   1.41167241  0.14964075 -0.09637905  0.82661621 -0.78595734
   0.34768723  1.51108316]]
[1]
The person is diabetic
```

```
input_data=(4,110,92,0,0,37.6,0.191,30)
input_data_as_numpy_array = np.asarray(input_data)

# reshape the array as we are predicting for one instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the input data
std_data = scaler.transform(input_data_reshaped)
print(std_data)

prediction = classifier.predict(std_data)
print(prediction)

if (prediction[0] == 0):
  print('The person is not diabetic')
else:
  print('The person is diabetic')
```

```
[[ 0.04601433 -0.34096773  1.18359575 -1.28821221 -0.69289057  0.71168975
  -0.84827977 -0.27575966]]
[0]
The person is not diabetic
```

Fig 6.3. Diabetes Result

CHAPTER 7

CONCLUSION

The diabetes prediction project represents a significant stride towards advancing healthcare strategies, particularly in the realm of diabetes management. Through a comprehensive exploration of predictive models, machine learning algorithms, and data analytics, the research aimed to enhance the precision and timeliness of diabetes detection. The problem statement underscored the urgency of addressing the rising challenge of diabetes prevalence, emphasizing the need for a more accurate and proactive approach to interventions.

Technological innovation played a pivotal role in the project, highlighting a dedication to staying at the forefront of scientific inquiry. By integrating cuttingedge approaches, such as machine learning, The emphasis on clarity and formal accuracy underscored the commitment to transparent scholarly discourse. Clear communication and adherence to formal standards were prioritized to enhance the credibility and reliability of the research findings.

In conclusion, the diabetes prediction project strides beyond conventional approaches, leveraging technological innovation, inclusivity, and a commitment to clarity. By enhancing accessibility, understanding, and the overall quality of diabetes-related information, this project aspires to leave a lasting impact on healthcare practices, contributing to a more proactive and informed approach to diabetes management.

CHAPTER 8

REFERENCES

[1] https://www.researchgate.net/publication/339543101_Diabetes_Pre diction_using_ Machine_Learning_Algorithms

[2] https://www.ijert.org/research/diabetes-prediction-using-machine-learningtechniques-IJERTV9IS090496.pdf

[3] https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.411.5264

[4] https://www.researchgate.net/publication/336570929_Predictive_models_for_diab etes_mellitus_using_machine_learning_techniques

[5] https://www.ijert.org/diabetes-prediction-using-machine-learningtechniques

[6] https://www.sciencedirect.com/science/article/pii/S1877050918308 548/pdf?md5= 9fce

[7] https://www.researchgate.net/publication/338581650_Diabetic_Prediction_Syste m_Using_Data_Mining

[8] https://research.vit.ac.in/publication/a-method-for-classification-

using-machinelearning-technique

[9] https://www.kaggle.com/uciml/pima-indians-diabetes-database

[10] https://hbr.org/2011/09/how-to-solve-the-cost-crisis-in-

healthcare