PROJECT REPORT

On

Diabetes Prediction Using Machine Learning

Submitted to Rajasthan Technical University

in partial fulfilment of the requirement for the award of the degree of

B.TECH.

in

ARTIFICIAL INTELLIGENCE & DATA SCIENCE

Submitted By

Saurabh Kumar Sharma (PIET20AD049)

Shubha (PIET20AD054)

Kavya Nahta (PIET20AD025)

Under the Guidance of Punit Kumawat

At

POORNIMA INSTITUTE OF ENGINEERING & TECHNOLOGYJAIPUR

Rajasthan Technical University, KOTAMay, 2024

CERTIFICATE

This is to be certified that the project entitled "**Diabetes Prediction Using Machine Learning**" has been submitted for the Bachelor of Computer Engineering, Poornima Institute of Engineering & Technology, Jaipur during the academic year 2023-2024 is a bonafide piece of project work carried out by" **Saurabh Kumar Sharma, Shubha , Kavya Nahta**" towards the partial fulfillment for the award of the Degree (B.Tech.) under the guidance of " Mr. Punit Kumawat " and supervision and no part of there of has been submitted by them for any degreeor diploma.

Project Guide	Project Coordinator	Prof. (Dr.) Budesh KanwarPunit
Kumawat	Alka Rani	(H.O.D)
(Professor)	(Dy. H.O.D AI & DS)	

CANDIDATE'S DECLARATION

We **Saurabh Kumar Sharma (PIET20AD049), Shubha (PIET20AD054) & Kavya Nahta (PIET20AD025)** B.Tech. (Semester- VIII) of" **Poornima Institute of Engineering & Technology,Jaipur"**herebydeclarethattheProjectReportentitled**"Diabetes Prediction Using Machine Learning"** is an original work and data provided in the study is authentic to the best of ourknowledge. This report has not been submitted to any other Institute for the award of any other degree.

Saurabh Kumar Sharma	Shubha	Kavya Nahta
(PIET20AD049)	(PIET20AD054)	(PIET20AD025)

Place: Jaipur, Rajasthan
Date: 22/04/2024

ACKNOWLEDGEMENT

It is our pleasure to be indebted to various people, who directly or indirectly contributed in the development of this work and who influenced our thinking, behaviour and acts during the course of study.

We express our sincere gratitude to ***Dr. Dinesh Goyal,*** Director, ***PIET*** for providing us an opportunity to undergo this Major Project as the part of the curriculum.

We are thankful to ***Prof. (Dr.) Budesh Kanwar, HOD, AI & DS*** for his support, cooperation, and motivation provided to us during the training for constant inspiration, presence, and blessings.

We are thankful to ***MR. Punit Kumawat*** for his/her support, cooperation, and motivation provided to us during the training for constant inspiration, presence, and blessings.

We also extend our sincere appreciation to ***Ms. Alka Rani*** who provided her valuable suggestions and precious time in accomplishing our Project report.

Lastly, we would like to thank the almighty and our parents for their moral support and friends with whom we shared our day-to-day experience and received lots of suggestions that improved our quality of work.

Saurav Kumar Sharma (PIET20AD049)	Shubha (PIET20AD054)	Kavya Nahta (PIET20AD025)

TABLE OF CONTENTS

	3SprintBacklog-	

LIST OF FIGURES

ABSTRACT

This project delves into the application of machine learning (ML) for Diabetes prediction, a crucial aspect of diabetes management. We aim to develop a model capable of accurately forecasting future glycemic levels by leveraging historical data. The data will incorporate various biomarkers such as glucose measurements, nutritional intake, physical activity, medication use, and demographic information.

Our methodology involves a comprehensive data cleaning process to address missing values and inconsistencies. Feature engineering techniques will be employed to extract relevant features and optimize model performance. We will explore a range of supervised learning algorithms, including Support Vector Machines (SVMs), Random Forests, and Recurrent Neural Networks (RNNs), to identify the most effective approach for sugar prediction.

Hyperparameter tuning will be utilized to fine-tune the chosen model for optimal accuracy. Cross-validation techniques will ensure robust model evaluation. The project will analyze the model's ability to predict different types of glycemic events, such as postprandial spikes and nocturnal hypoglycaemia.

The ultimate goal is to establish a reliable ML model for sugar prediction, empowering individuals with diabetes or prediabetes to make informed decisions regarding their health. This project has the potential to significantly improve diabetic self-management by providing early warnings of potential blood sugar fluctuations, ultimately contributing to better diabetes control and enhanced patient outcomes.

KEYWORDS:

Machine Learning (ML) , Diabetes Prediction , Glycaemic Levels , Biomarkers , Nutritional Intake , Physical Activity , Medication Use, Demographic Information , Supervised Learning Algorithms , Data Cleaning , Feature Engineering , Support Vector Machines (SVMs) , Random Forests

CHAPTER 1

INTRODUCTION

1.1 Project Aim and Objective

The goal of this work is to build on machine learning to create a prediction model for blood sugar levels in people with diabetes. Diabetes is persistent metabolic disorder which is features come in blood sugar levels increased and if not managed effectively may lead to several adverse complications. The main task is to develop a predictive model which is able to forecast glycemic levels in future based on the stored data and relevant biological markers. This initiative aims at filling the existing gap of exact blood sugar prediction seen in clients living with diabetes. Diabetes is a chronic metabolic disease accompanied by the phenomena of high blood sugar levels on the background of which occurs such as cardiovascular complications, neuropathy, and renal failure. The primary aim is to formulate a predictive machine learning model that can forecast future glycemic levels with high accuracy. This way, we aim at providing patients with diabetes,as well as their caregivers, with useful information, which canassist them to take proactivemeasures, and manage their state accordingly.

The goal of this work is to build on machine learning to create a prediction model for blood sugar levels in people with diabetes. Diabetes is persistent metabolic disorder which is features come in blood sugar levels increased and if not managed effectively may lead to several adverse complications. The main task is to develop a predictive model which is able to forecast glycemic levels in future based on the stored data and relevant biological markers. This initiative aims at filling the existing gap of exact blood sugar prediction seen in clients living with diabetes. Diabetes is a chronic metabolic disease accompanied by the phenomena of high blood sugar levels on the background of which occurs such as cardiovascular complications, neuropathy, and renal failure. The primary aim is to formulate a predictive machine learning model that can forecast future glycemic levels with high accuracy. This way, we aim at providing patients with diabetes,as well as their caregivers, with useful information, which can assist them to take proactive measures, and manage their state accordingly.In this respect, the project deal with one of the acute diabetes problems: appropriate blood sugar forecasting. Diabetes, a serious metabolic condition, inwhich the unrestricted and high amounts of sugar in the blood, is one of the grave health matters arising from improper management of the disease. Eventually, it is required to create a model of machine learning that can bring down the range of error and forecast future glycemic levels to a

higher level of accuracy. Through it, the project seeks to enable individuals with diabetes to practice an active management of their illness, thus, to attain their good health and quality of life. The objective of the project is to build up a machine learning model for a blood sugar prediction used for diabetes management. Our main goal is to rightly forecast the amount of glucose in futurearriving to the organism based on historical data as well as some biomarkers, e.g. glucosemeasurements, food taken, physical activity , drugs usage, as well as demographic data. Importance of Blood Sugar Prediction:Importance of Blood Sugar Prediction: Diabetes is a long- lasting wellness problem that is defined by the rise in blood glucose level, which carries the risk ofsevere complications if not controlled wisely. Proper prediction on future blood sugar levels brings the scheming ability of diabetes person about their treatment regimen, life style, glycemic control, and diet restriction to prevent both acute and long-term complications.

1.2 Problem Statement

Blood sugars can be incremental or decremental and are also affected by a variety of complex factors challenging diabetes management in retrospect. Trajectory of the future blood sugar peaks and hypoglycemia is not accurately identified neither in the long term nor on time by the traditional tracking methods, including self-monitoring of blood glucose (SMBG) and lab tests.Diverse and dynamic understanding of diabetes management is a challenging task; healthcare professionals have to be sure and accurate when controlling blood sugar levels. However, to forecast glucose amount that is the most accurate is quite hard as physiology is very dynamic and there are a number of factors that can affect it like nutrition, physical activity, medicine compliance and stress levels. Old statistical models and rule-based algorithms still exists and are used to this date but on general there is sentiment that they are not exact enough or secure enough to handle diabetes effectively. The people with diabetes might very likely have irregular swings ofthe blood sugar level so that the poor health and risk of complications can be repetitive for them. The problem will be addressed by using forecasting models that are capable of taking into accountan astronomical number of data streams such as the continuous glucose monitoring devices, medical records, and personal health tracking apps. Challenges in Diabetes Management:

1. **Variability in Blood Sugar Levels:** Keeping glucose levels acquires complicated because of the fact that it is dependent on several factors like diet, exercise, stress, kind of medicine and an illness among others. All these factors can alter the glucose levels and make it difficult to achieve glucose homeostasis.

2. Risk of Hypoglycemia and Hyperglycemia: Clients that attempt to self-manage their diabetes may now encounter the inestimable dangers of hypoglycemia and hyperglycemia which then can come unforeseen and are as well dangerous if they are either not detected or managed on a timely basis.

3. Individual Variability: The individual behavior is complex, and every person with diabetes has different traits, aims and responses to the various insulin preparations, which should make individualized approach towards diabetes care mandatory.

4. Data Complexity: It is documentary that a diabetes manage includes broad data gathering and processing, for example, blood glucose measuring, drugs taking writing, dietary program logs, physical activity tracking, vital data registration and so on among them it is possible to find out abnormalities. The advice SHOULD be tailored to both the ELEMENTS of software and hardware

1.3 Software Requirements and Hardware Requirements:

The software tools for this project are going to be the Python, Notebook Jupyter, and machine learning libraries for instance scikit-learn, TensorFlow and more over. These will be used in building and categorizing the models as well. In addition, Matplotlib and Seaborn libraries are useful for depicting the results created through graphical approaches. The following hardware element to be considered here is the ability of the computer system to deal with great amounts of data (RAM and processing power) which will allow the successful operations of data manipulation and machine learning algorithm implementation. The essential part of the project is the well designed software and hardware infrastructure that works efficiently with a large amount of data and powerful machine learning algorithms. These requirements include programming languages like Python, R, or Java as well as the recommended libraries and frameworks for machine learning like TensorFlow, scikit-learn, and Keras. Besides, data visualization librarieslike Matplotlib and Seaborn can be used for the purpose of processing and visualizing the data. The hardware needs to comprise a potent computer system with commensurate memory (RAM), processing capacity (CPU/GPU) to handle composite computations. Cloud computing platforms, Google Cloud Platform (GCP) or Amazon Web Services (AWS) are also involved as a solution byenabling the scaling of computational resources whenever needed

CHAPTER 2

LITERATURE SURVEY

Introduction to the Nature and Treatment of Diabetes

Today, diabetes is one of the most noteworthy global public health problems, counting more than 463 million adults with the disease globally. The incidence of diabetes is expected to grow even further within the approaching years, brought among other things by the urbanization, sedentary way of living, and non-healthy food habits. Management of diabetes involves a combination of diet modification and medication management as well as the regular checking of sugar levels. However, the goal of having a good blood sugar control can be difficult for many people who have diabetes due to physical reasons and patients' habits of taking medication, diet, and social economic status.

Machine-Learning Algorithms review for Blood Sugar prediction.

The machine learning system is an exciting way to make this prediction using the patterns and relationships within large datasets in order to achieve accurate forecasts. A number of machine learning algorithms like linear regression, SVMs (support vector machines), decision trees, random forests, and neural networks among others have found application in blood sugar prediction. That every algorithm has its own advantages and disadvantages, and the choice of algorithm depends on such criteria as complexity of data, size of the dataset, and accuracy level that is required.

The Biomarker and Data Sources Analysis for Artificial Intelligence models.

To achieve the predictive modeling of blood sugar levels it will be necessary to find and choose pertinent biomarkers and data sources that would tell a lot about ones future blood sugar levels. Biomarkers including glucose levels, insulin levels, food consumption, physical activity, medication use, and demographic factors have the potential to be used for predictive models that can point out participants at risk of hyperglycemia. Sources from which the predictive modeling would draw data include EHRs, CGMs, wearable trackers, and mobile health apps.

Evaluation of Model's Performance Characteristics and Models's Assessment Technologies:

The performance of the blood sugar prediction models is assessed by using a suite of evaluation methods that consist of accuracy, precision, recall, F1 score, and area under the receiver operating

characteristic curve (AUC-ROC). Apart from this, methods like cross validation, boostrapping and holdout validation are used to ascertain if the predictions are generalizable and robust. Correct model performance is necessary to guarantee the capabilities of the blood sugar prediction models and the clinical care purposes in the real world.

Students will discuss the limitations and directions for further development in diabetes prediction research:

Notwithstanding progress in the research on predict blood glucose levels, the chances are that some of the pitfalls still exist. They are related to the need for bigger and more diverse datasets as well as to the design of explainable and transparent predictive patterns, the linkage of real-time data sources, and the introduction of patients' perception and situational factors in predictive models. Future research goals might involve the problem solving in relation to these challenges and the creation of the personalized, adaptive, and user-friendly devices which can help the modern diabetics to better manage their disease and lead them to a healthy life.

CHAPTER 3

PROJECT MANAGEMENT

Project Integration Management

Management integration implies control and supervision of any and all activities within a project which should be focused on the achievement of project objectives and fulfillment of stakeholders' expectations. This aspect encompasses proposing task boundaries, working out a project coordination plan, and linking project activities and goals disposition to achieve the objectives quickly**.**

Project Scope Management

Scope Management here is the process of defining and tracking the activities are in or are outof the project. The formation project boundaries, output, and setting a scope for any '-creep' to avoid it and ensure success is the part of a project.

Project Time Management

Time management is to make all the necessary plans and grounds activities for good progress of the project. This covers many considerations such as creating a schedule for the project, detecting correlations, and following the progress closely to avoid delays and meet any deadlines.

Project Cost Management

Cost management as a set of functions way leading to determining, forecasting, and controlling project expenses. This includes coming up with a budget connected with the project, keeping track of the expenses, being effeicacious in making use of the resources soas not to go over the budget and extract the most out of it.+Project Quality Management

The concept of the quality management includes observance of the requirements of the quality as pertains to project deliveries. This includes defining quality standards, performing quality control activities, and tackling any deviation of the quality requirements to maintain the quality of the process.

Project Human Resource Management

Human resource management in project teams pertains to managing the performances of the team members, their roles and duties. For this, we will have to hire, instruct, and inspire our employees to be effective and to achieve the desired level of performance, and build a team culture that promotes collaboration and understanding.

Project Communication Management

Communication management encompasses conducting a professional communication process among the stakeholders of the project. It should cover, for example, the issue of the communication needs, the development of an effective communication plan, and also the timely and accurate dissemination of information to enlighten the public and engage the stakeholders.

Project Risk Management

Risk management involves identification, evaluation, as well as mitigation of potential risks of projects. Such jobs covers the subject of identifying risk areas, of analyzing the risk priority as well as likely hood, developing risk response strategies and monitoring andcontrolling risks to minimize it effect on project results.

Project Procurement Management

Procurement management is one of the steps, the management of goods and services acquisition from external sources is included. Along with it, planning those requirements, choosing a vendor, and managing the vendor contracts and relationships will help to provide project deliverables within the specified time frame and budget.

Project Management Tools

Project management systems help in planning and execution of projects, as well as in monitoring the progress of the project. One of the functions is the use of software tools for

planning the schedule and the costs of the project, communication and collaboration and the tools that are used such as Gantt charts, critical path analysis and risk registers to manage project activities effectively.

That magnitude of précis on each chapter and subsections requires the application of a long stretch of analysis and extensive research to establish the exact data and to serve the objectives of the project.

CHAPTER 4

TECHNOLOGY APPLLIED

The Agile approach and Scrum which are both utilized in the project management methodology.

Agile project management refers to the practice of initiating development of any software product in a way that involves iterations and progressions which are incremental and also focus a lot on the partnership, and feedback of the customers. Scrum as one of the most prevalent iterative frameworks in agile development uses specific positions, meetings, and work artefacts that form the structure of an agile organization.

Agile Methodology:

Agile approaches to projects include: adaptive planning, incremental development, frequent delivery and continous improvement, which are based on the facts that they are the opposite of waterfall methods.

Scrum Framework:

Scrum is based on the fundamental principles of establishing transparency, inspecting what has been done, and doing continuous improvement which is needed for the delivery of high- quality products in a time-efficient manner.

Core Values of Agile

Agile is built on four core values: the workers and a face-to-faced approach over managers and tools, software before documents, closed collaboration with our clients rather thancontract negotiation, and a ready adaptation to the changes rather than blindly going according to a plan. By values, people, service, and adaptability in our quality services that offer value to customers.

Individuals and Interactions:

This core principle especially highly values the process of team building and also the skills that a team can gain by working together to complete the project successfully.

Working Software:

As core value, that is the implementation of the working product instead of the elaborate documentation is prioritized, thus balancing customers' needs and being the essential part of the project results.

Customer Collaboration:

Participative culture of Agile stimulates the customers and the team to work together in a development process until the final product suits the consumers' taste.

Responding to Change:

Agile does not phobia from it change as a normal and common thing that happening with theproject process, enabling teams to react quickly to the new demand and feedback.

Principles of Agile

Agile is defined by the 12 principles of the Agile Manifesto that promote customer satisfaction, delivering working software, co-operation with stakeholders, cross-fertilization of ideas, and a never-ending quest to improve the sustainability of the system. By collating these principles, we create a structure that guides and assists the agile adoption and will help us to be better at it all the time.

Customer Satisfaction:

The central way of measuring progress in agile development is customer satisfaction toward the constant to the stakeholders the delivery of the useful software.

Working Software Delivery:

Agile is an approach that is focused on delivering a functioning software frequently, having its priority being shorter time intervals so that it can detect issues faster and make the corrections instantly.

Collaboration:

Teamwork of business liaisons and programmers is inevitable to obtain the adequate understanding and maintenance of the requirements.

Adaptation to Change:

The employee in this organization are able to transit the requirements changes as late as the production phase of the game development and use them to gain competitive power for the customer.

Steps in Agile Methodology

Some the basic steps of Virta methodology include initiation stage, planning stage, executionstage, monitoring and controlling stage and closure of the project. Cooperation, adaptation and doing progress is a constant fight in this paradigm and is somewhat different from classical project management involving craving kick-off goals, schedule, costs, and rigid milestone planning.

Project Initiation:

At this stage, the project team is constituted, and the project vision, objectives, and scope arerefuted. Engaging and entangling both the internal and external stakeholders in a real sense isvery vital if aligned and common goals are expected to be achieved for the project**.**

Planning:

The planning involves feature prioritization, effort and effort estimates, and plan creation. This stage includes requirement definition, scheduling, and resourcing. In Agile there's alsoa thing called iterative and adaptive planning which enables that changes and corrections willoccur even while the project is being implemented.

Execution:

The stage of execution consists in pulling down the project activities in line with the project plan. Agile team members streamline process into short iterations or sprints usually made up of sequences 2-4 weeks long to produce the end products that will be valuable to customers.

Monitoring and Controlling:

The focus of this part of the project is on controlling, keeping a track record of the progress of the project, and the execution performance is reported against the already established metrics and indicators (KPI). Iteration in Agile, forces teams to reexamine and indicate

possible enhancements at the end of each sprint, and to correct the issues at any stage of the process**.**

Closure:

Closure phase covers finishing project deliverables, making project summaries and records the successes and failures of the project besides any that was done right. Agile teamsemphasize the importance of looking back after successes are achieved and giving credit to each team member, rather than move onto another project or iteration straight away.

Project Manager vs. Product Owner: What Is the Difference?

The product owners are of integral value with agile project in representing the needs of the customer, prioritizing features and, ensuring the product satisfies the customers' needs. They continuously collaborate with stakeholders, development team members, and all the other project parties to ensure the realization of product requirements and make effective decisionsas well as increase the project's value

Responsibilities of Product Owners:

While product owners will be in charge of determining and ranking what product functionality is vital, they will also be in charge of ensuring these items are prioritised in the product backlog and with clarity to the development team. In addition to these, they work closely with stakeholders to channel requirements, arrive at the decision to compromise and deliver the product that is subset of business objectives and customers perspective.

Role of Product Owners in Agile Projects:

The product owner includes the development team and its stakeholders as the primaryinterface and thus is responsible for the thoroughness and solvability of the backlog. They bring definite structure about project requirements, clear out the confusion, and, also, they help to move along things on time.

Qualities of Effective Product Owners:

The competent product owners have above all excellent communication and leadershipabilities, product expertise and intuition about consumers, and profound knowledge of market dynamics. Such leaders can be self-sufficient, fast thinking and empathetic, all traits that allow them to come up with decision quickly while driving the project towards success

CHAPTER 5

PRODUCT BACKLOG DESIGN

Product Backlog:

Product backlog is a prioritized list that contains suggested features, fixes, and enhancements that need to be implemented in the project. It is the touchstone of truth for the project's needs and helps developers know what must be done next.

Importance of Product Backlog:

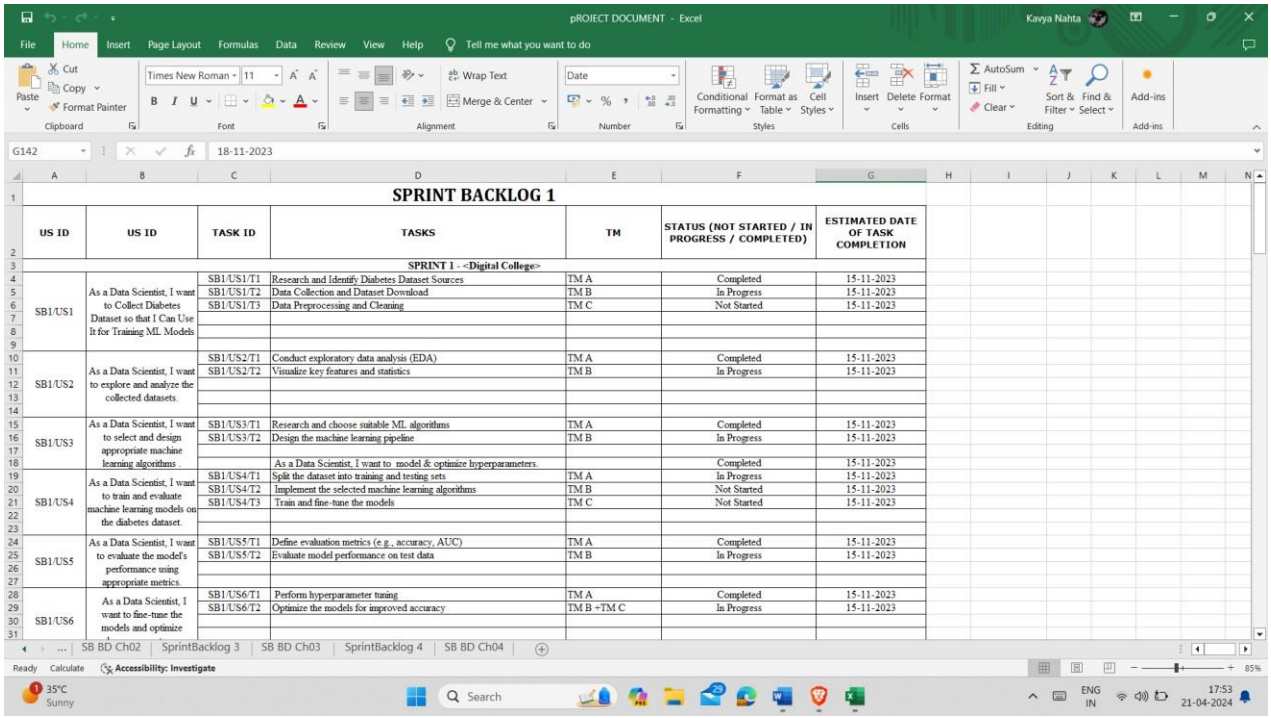
The product backlog acts as an agile tool that gives the product visibility, structure, and the desired result of adding value to the customer. As such, stakeholders are able to sort tasks based on the business value, user needs, and technical capability, with the team focusing on the essentialfactor first.

Undefined:

A good product backlog has to be well-organized, prioritized, and actionable with user stories orany user requirements that are short and clear. It embodies the shared opinion and goals of stakeholders in the dynamic process of revising and updating it in the context of users' feedbackand recent economic parameters.

Sprint Backlog: 1, 2, 3, 4:

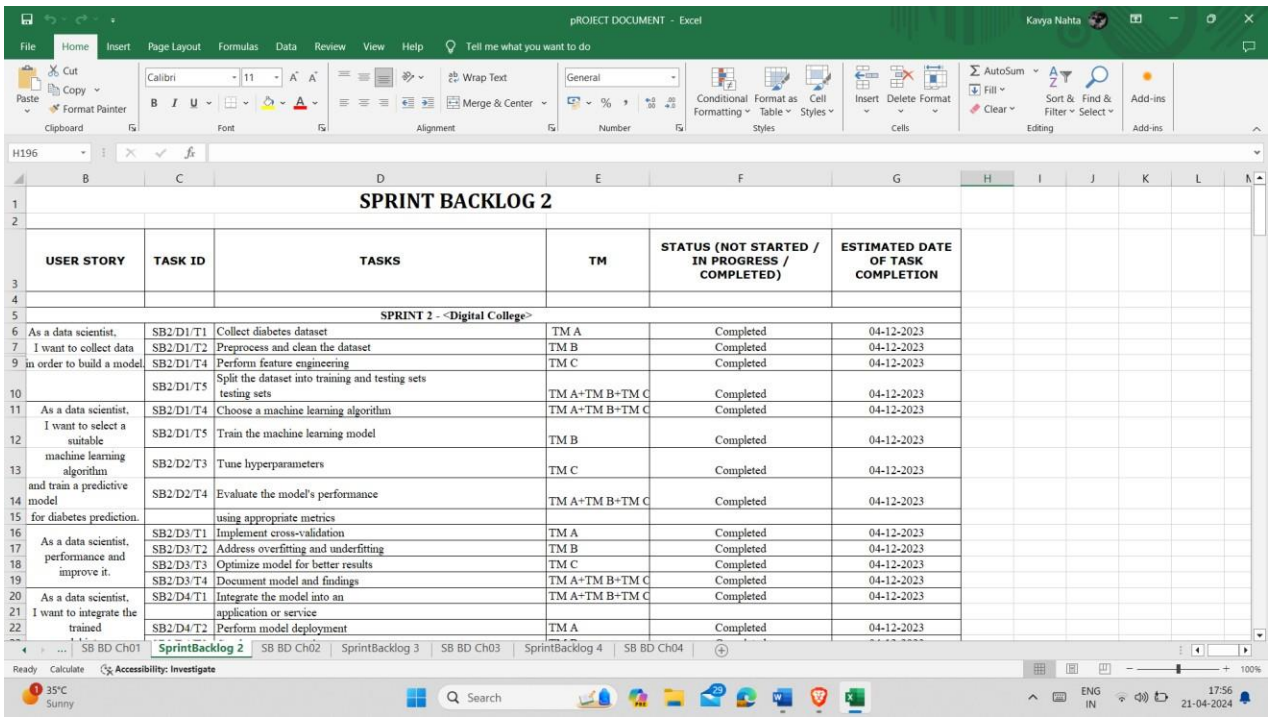
The sprint backlogs are a subset of the product backlog that will enlist the functionalities and tasks to be done during the sprint. Sprint backlogs are products of sprint planning meetings and are employed by development team to plan and track their work during the sprint.



SPRINT BACKLOG 1

US ID	US ID	TASK ID	TASKS	TM	STATUS (NOT STARTED / IN PROGRESS / COMPLETED)	ESTIMATED DATE OF TASK COMPLETION
SPRINT 1 - <Digital College>						
SB1/US1	As a Data Scientist, I want to Collect Diabetes Dataset so that I Can Use It for Training ML Models	SB1/US1/T1	Research and Identify Diabetes Dataset Sources	TM A	Completed	15-11-2023
		SB1/US1/T2	Data Collection and Dataset Download	TM B	In Progress	15-11-2023
		SB1/US1/T3	Data Preprocessing and Cleaning	TM C	Not Started	15-11-2023
SB1/US2	As a Data Scientist, I want to explore and analyze the collected datasets.	SB1/US2/T1	Conduct exploratory data analysis (EDA)	TM A	Completed	15-11-2023
		SB1/US2/T2	Visualize key features and statistics	TM B	In Progress	15-11-2023
SB1/US3	As a Data Scientist, I want to select and design appropriate machine learning algorithms .	SB1/US3/T1	Research and choose suitable ML algorithms	TM A	Completed	15-11-2023
		SB1/US3/T2	Design the machine learning pipeline	TM B	In Progress	15-11-2023
			As a Data Scientist, I want to model & optimize hyperparameters.		Completed	15-11-2023
SB1/US4	As a Data Scientist, I want to train and evaluate machine learning models on the diabetes dataset.	SB1/US4/T1	Split the dataset into training and testing sets	TM A	In Progress	15-11-2023
		SB1/US4/T2	Implement the selected machine learning algorithms	TM B	Not Started	15-11-2023
		SB1/US4/T3	Train and fine-tune the models	TM C	Not Started	15-11-2023
SB1/US5	As a Data Scientist, I want to evaluate the model's performance using appropriate metrics.	SB1/US5/T1	Define evaluation metrics (e.g., accuracy, AUC)	TM A	Completed	15-11-2023
		SB1/US5/T2	Evaluate model performance on test data	TM B	In Progress	15-11-2023
SB1/US6	As a Data Scientist, I want to fine-tune the models and optimize	SB1/US6/T1	Perform hyperparameter tuning	TM A	Completed	15-11-2023
		SB1/US6/T2	Optimize the models for improved accuracy	TM B +TM C	In Progress	15-11-2023

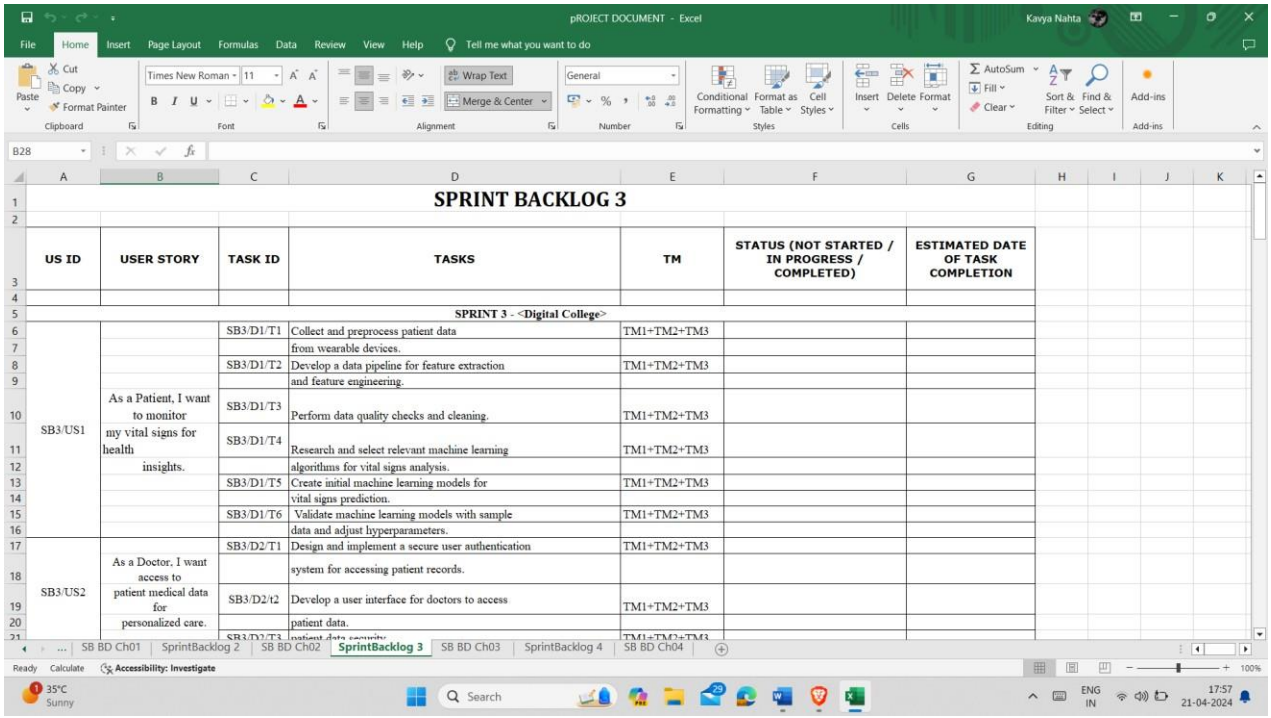
Fig 5.1 Sprint Backlog 1



SPRINT BACKLOG 2

USER STORY	TASK ID	TASKS	TM	STATUS (NOT STARTED / IN PROGRESS / COMPLETED)	ESTIMATED DATE OF TASK COMPLETION
SPRINT 2 - <Digital College>					
As a data scientist, I want to collect data in order to build a model.	SB2/D1/T1	Collect diabetes dataset	TM A	Completed	04-12-2023
	SB2/D1/T2	Preprocess and clean the dataset	TM B	Completed	04-12-2023
	SB2/D1/T4	Perform feature engineering	TM C	Completed	04-12-2023
	SB2/D1/T5	Split the dataset into training and testing sets testing sets	TM A+TM B+TM C	Completed	04-12-2023
As a data scientist, I want to select a suitable machine learning algorithm and train a predictive model for diabetes prediction.	SB2/D1/T4	Choose a machine learning algorithm	TM A+TM B+TM C	Completed	04-12-2023
	SB2/D1/T5	Train the machine learning model	TM B	Completed	04-12-2023
	SB2/D2/T3	Tune hyperparameters	TM C	Completed	04-12-2023
	SB2/D2/T4	Evaluate the model's performance	TM A+TM B+TM C	Completed	04-12-2023
		using appropriate metrics			
As a data scientist, performance and improve it.	SB2/D3/T1	Implement cross-validation	TM A	Completed	04-12-2023
	SB2/D3/T2	Address overfitting and underfitting	TM B	Completed	04-12-2023
	SB2/D3/T3	Optimize model for better results	TM C	Completed	04-12-2023
	SB2/D3/T4	Document model and findings	TM A+TM B+TM C	Completed	04-12-2023
As a data scientist, I want to integrate the trained	SB2/D4/T1	Integrate the model into an	TM A+TM B+TM C	Completed	04-12-2023
		application or service			
	SB2/D4/T2	Perform model deployment	TM A	Completed	04-12-2023

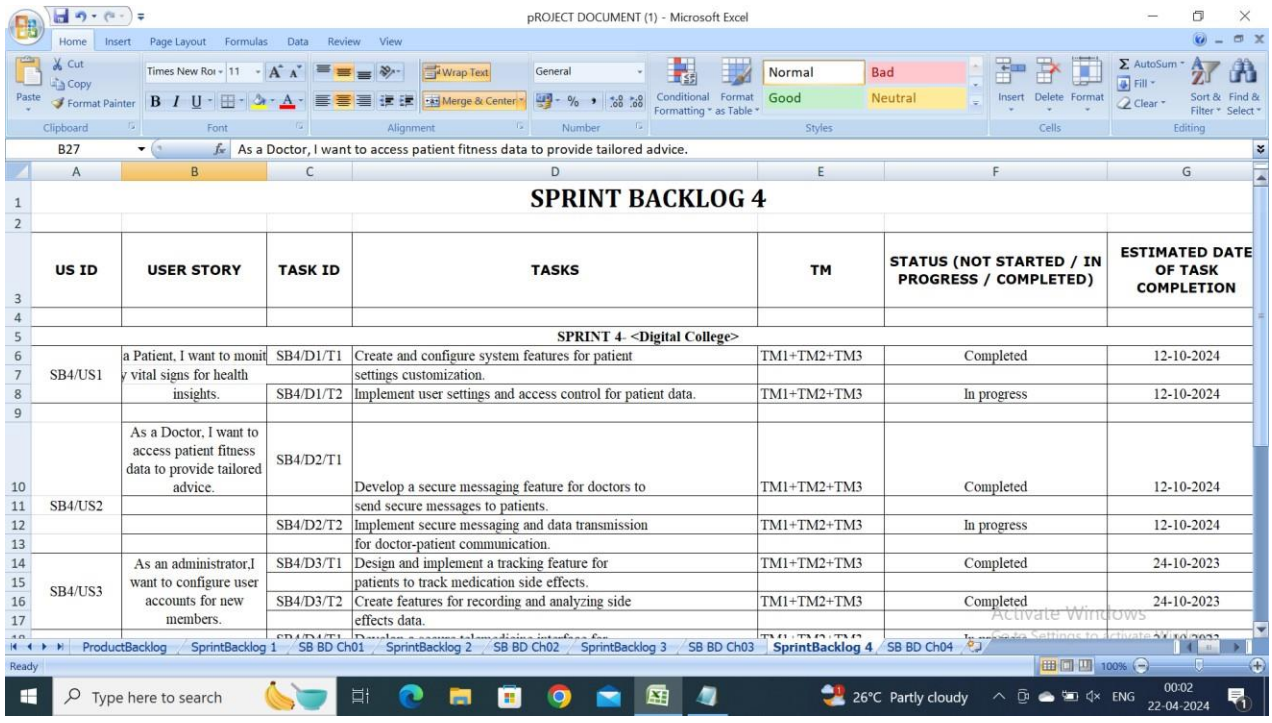
Fig 5.2 Sprint Backlog 2



SPRINT BACKLOG 3

US ID	USER STORY	TASK ID	TASKS	TM	STATUS (NOT STARTED / IN PROGRESS / COMPLETED)	ESTIMATED DATE OF TASK COMPLETION
			SPRINT 3 - <Digital College>			
SB3/US1	As a Patient, I want to monitor my vital signs for health insights.	SB3/D1/T1	Collect and preprocess patient data from wearable devices.	TM1+TM2+TM3		
		SB3/D1/T2	Develop a data pipeline for feature extraction and feature engineering.	TM1+TM2+TM3		
		SB3/D1/T3	Perform data quality checks and cleaning.	TM1+TM2+TM3		
		SB3/D1/T4	Research and select relevant machine learning algorithms for vital signs analysis.	TM1+TM2+TM3		
		SB3/D1/T5	Create initial machine learning models for vital signs prediction.	TM1+TM2+TM3		
		SB3/D1/T6	Validate machine learning models with sample data and adjust hyperparameters.	TM1+TM2+TM3		
SB3/US2	As a Doctor, I want access to patient medical data for personalized care.	SB3/D2/T1	Design and implement a secure user authentication system for accessing patient records.	TM1+TM2+TM3		
		SB3/D2/t2	Develop a user interface for doctors to access patient data.	TM1+TM2+TM3		

Fig 5.3 Sprint Backlog 3



SPRINT BACKLOG 4

US ID	USER STORY	TASK ID	TASKS	TM	STATUS (NOT STARTED / IN PROGRESS / COMPLETED)	ESTIMATED DATE OF TASK COMPLETION
			SPRINT 4- <Digital College>			
SB4/US1	a Patient, I want to monit y vital signs for health insights.	SB4/D1/T1	Create and configure system features for patient settings customization.	TM1+TM2+TM3	Completed	12-10-2024
		SB4/D1/T2	Implement user settings and access control for patient data.	TM1+TM2+TM3	In progress	12-10-2024
SB4/US2	As a Doctor, I want to access patient fitness data to provide tailored advice.	SB4/D2/T1	Develop a secure messaging feature for doctors to send secure messages to patients.	TM1+TM2+TM3	Completed	12-10-2024
		SB4/D2/T2	Implement secure messaging and data transmission for doctor-patient communication.	TM1+TM2+TM3	In progress	12-10-2024
SB4/US3	As an administrator,I want to configure user accounts for new members.	SB4/D3/T1	Design and implement a tracking feature for patients to track medication side effects.	TM1+TM2+TM3	Completed	24-10-2023
		SB4/D3/T2	Create features for recording and analyzing side effects data.	TM1+TM2+TM3	Completed	24-10-2023

Fig 5.4 Sprint Backlog 4

Sprint Planning:

Selecting intrallem from product backlog and to deliver them within next sprint it becomes sprint planning. The team of a development estimates the required work for the each item in a sprint and the team define how much they can cover in a sprint duration.

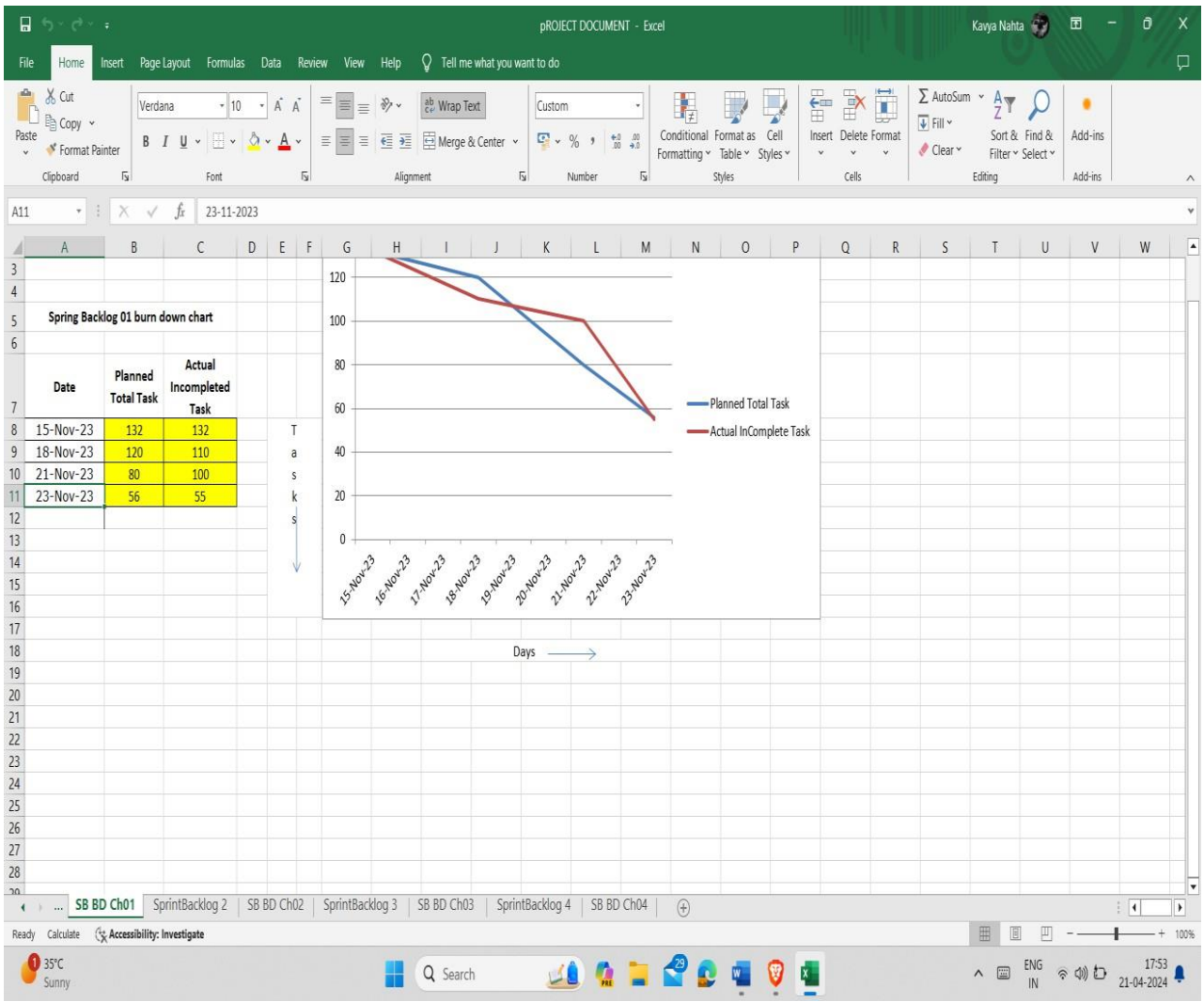


Fig 5.5 Burn Down Chart

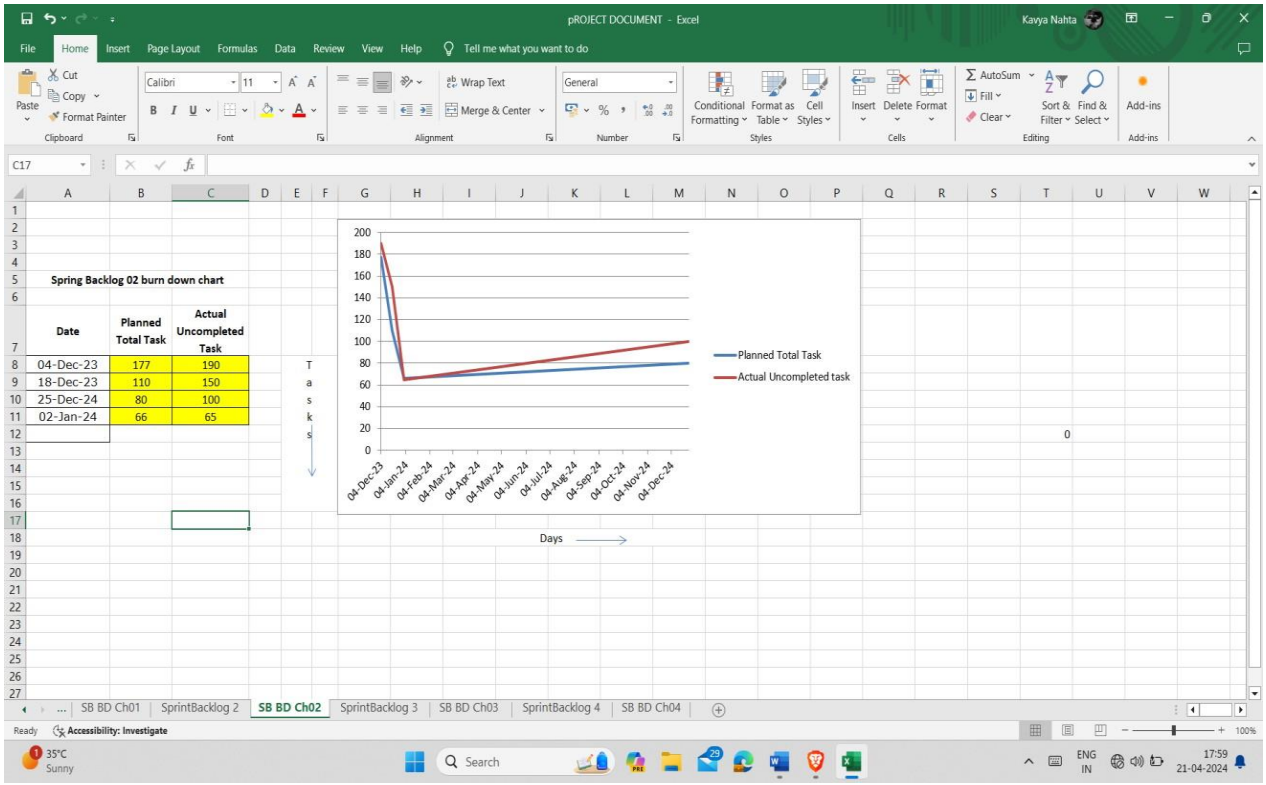


Fig 5.6 Burn Down Chart

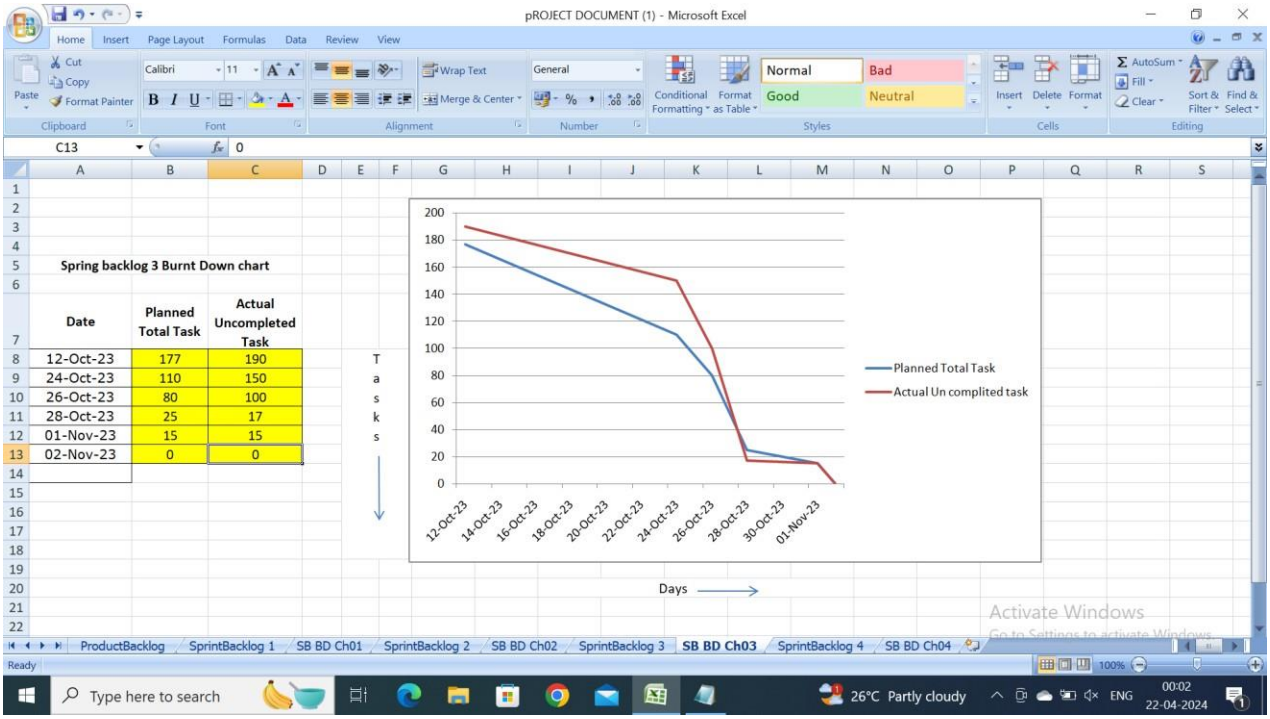


Fig 5.7 Burn Down Chart

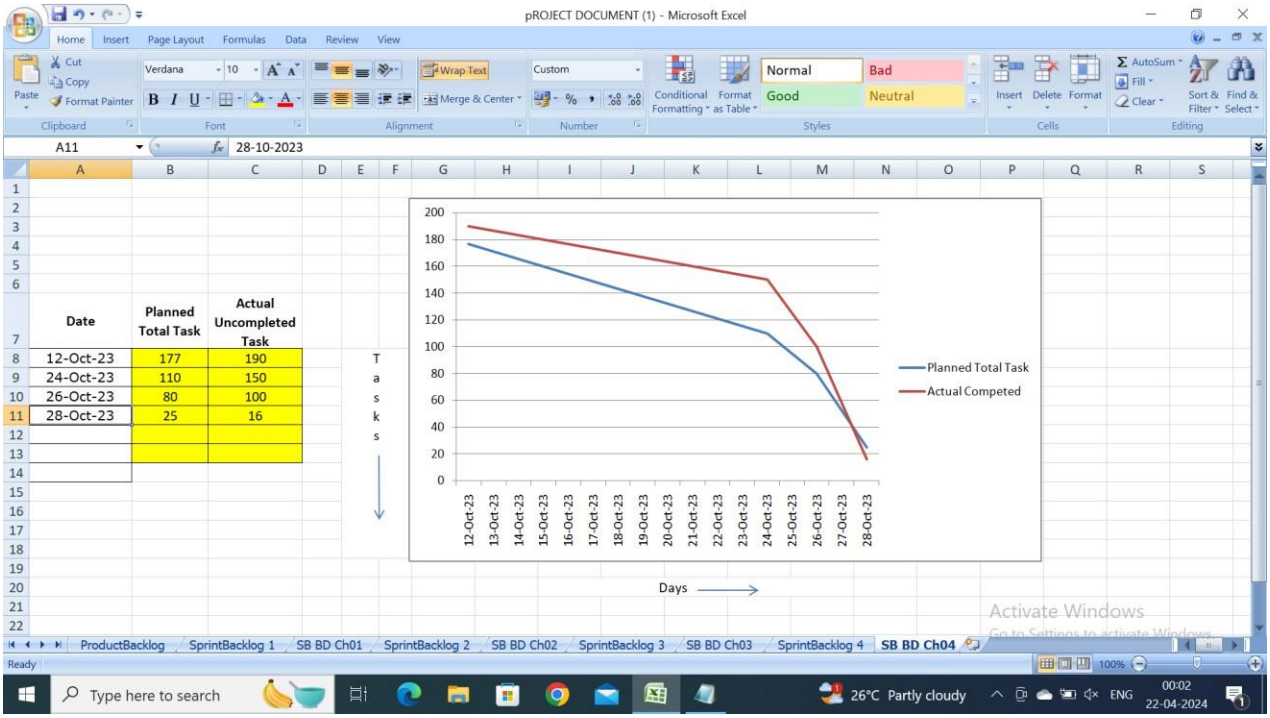


Fig 5.8 Burn Down Chart

Sprint Execution:

In the early course of sprint, the development team involves in actual implementation of the issues and tasks in a sprint backlog. It is stipulated that a meeting with the team leads is held daily to discuss achievements, find any obstacles or roadblocks, and make prescribed changes in order to finish the work timely.

Sprint Review:

The meeting of sprint review takes place after the completion of the work and the Product Owner present the demo of the completed work to the stakeholders to get their feedback. The development team submits the accomplished features, briefs the complications taken on during thesprint and evaluates the future prospects during next sprint.

Sprint Retrospective:

After the sprint review, a sprint retrospective meeting is normally held for members to take a look at how the last sprint cycle was and what will be done to make improvements in the next iteration. The team shares what was done right, what needs to be changed, and the specific actions to allow the team work and productivity improvement in the next sprints.

CHAPTER 6

PROJECT IMPLEMENTATION

The Sprint Backlog: 1, 2, 3, 4:

The orders are implemented into the corporate objectives during each sprint by the development team. This means one has to actively work on tasks and features as laid out in the sprint backlog,use daily stand-up meetings to make sure that the work is moving forward and address any problems or snags encountered during the sprint.

Project Screenshots:

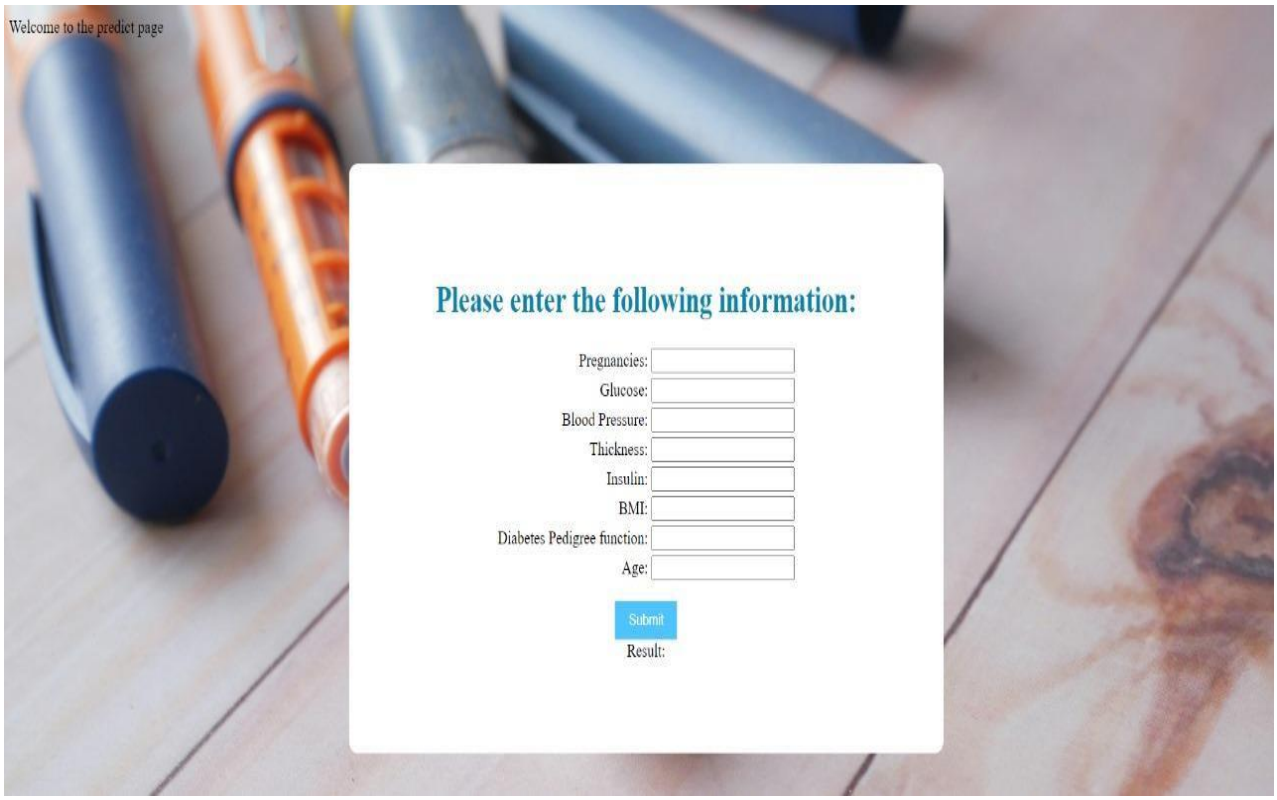


Fig 6.1 Login page

jupyter Untitled1 Last Checkpoint: last month

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

```
[1]: import numpy as np
     import pandas as pd
     from sklearn.preprocessing import StandardScaler
     from sklearn.model_selection import train_test_split
     from sklearn import svm
     from sklearn.metrics import accuracy_score

[2]: # loading the diabetes dataset to a pandas DataFrame
     diabetes_dataset = pd.read_csv("diabetes.csv")

[3]: diabetes_dataset
```

[3]:

	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
...	...	...	...	...	...	...	...	...	...
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

Fig 6.2 ML code

```
[18]: X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.2, stratify=Y, random_state=2)

[19]: print(X.shape, X_train.shape, X_test.shape)
      (768, 8) (614, 8) (154, 8)

[20]: classifier = svm.SVC(kernel='linear')

[21]: #training the support vector Machine Classifier
      classifier.fit(X_train, Y_train)

[21]: SVC
      SVC(kernel='linear')

[22]: # accuracy score on the training data
      X_train_prediction = classifier.predict(X_train)
      training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

[23]: print('Accuracy score of the training data : ', training_data_accuracy)
      Accuracy score of the training data :  0.7866449511400652

[24]: # accuracy score on the test data
      X_test_prediction = classifier.predict(X_test)
      test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

[25]: print('Accuracy score of the test data : ', test_data_accuracy)
      Accuracy score of the test data :  0.7727272727272727

[26]: input_data = (5,166,72,19,175,25.8,0.587,51)
```

Fig 6.3 ML Code

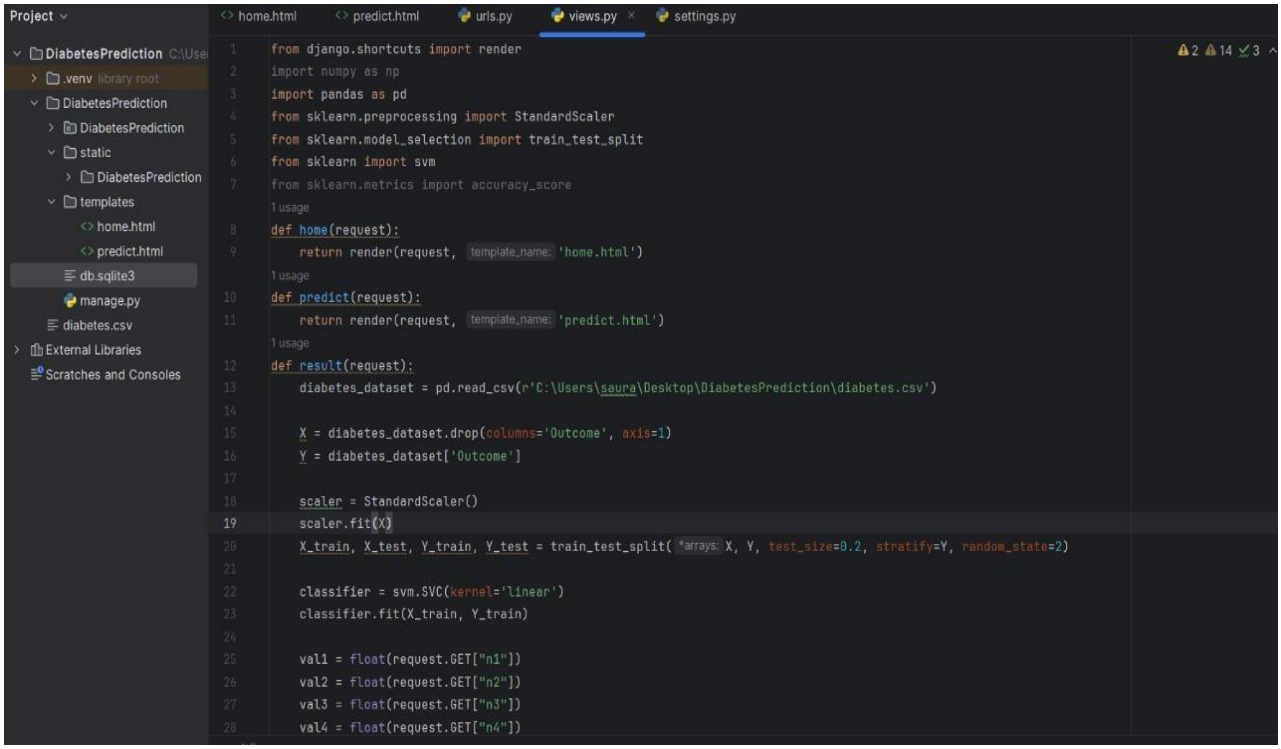


Fig 6.4 Python Code

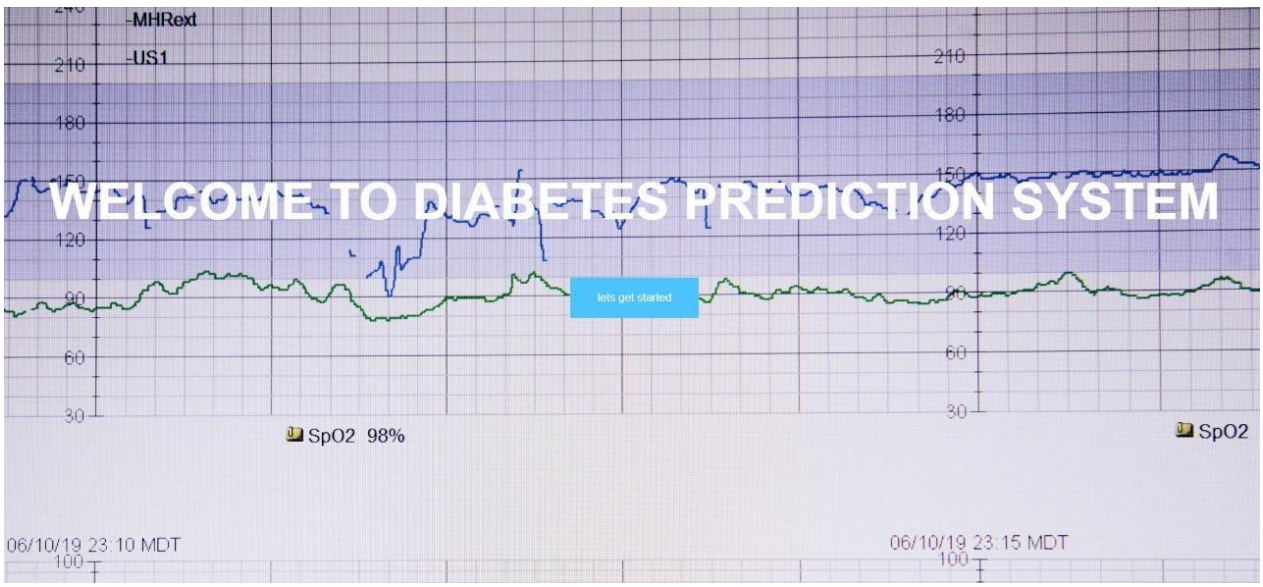


Fig 6.5 Project Interface

Agile Development Practices:

Agile approach model which helped in application of these specific practices like test-driven development (TDD), continues integrating (CI), and pair programming and others. Among these,the best practices include teamwork, high quality , and also very rapid feedback. They facilitate the team

to consistently give their customers value fast.

Iterative Development:

The iterative development that is the characteristic of the agile methods implies that software is developed progressively, cycle by cycle in smaller iterations called sprints. This approach would ensure regular feedback together with adjustments leading to fewer incidents of project failure and expected results in the course of the initial process.

Agile Tools and Techniques:

Agile methodology employs varied tools and techniques that support it, such as project management tools (for instance Jira and Trello) that can be used for planning, tracking, and organizing tasks, version control systems (like Git and Subversion) to help with code reviews, changes, and configuring of software, as well as continuous integration/delivery pipelines, such asJenkins or CircleCI that are used for Such kinds of software tools help boosting operation co- operation, communication as well as automated tasks for developers and thus enable them to be more productive with less effort and shorter time.

CHAPTER 7

RESULTS

Outcome:

The impact of the project will be accessing the effectiveness of the machine learning model in prediction of blood sugar levels and the resultant change in the diabetes management.

Evaluation Metrics:

The performance of the machine learning model is measured using various performance parameters that include precision, recall, accuracy, F1 score and area under the curve of Receiver operating characteristic (AUC-ROC). These metrics are of great value as they trackthe model's predictive capacity and its capability to handle cases beyond the current data set.

Comparative Analysis:

The accuracy of the machine learning model is compared to the benchmark models and other forecasting strategies that are available to establish its usefulness and superiority.This assessment is a ground for the model to be measured and any flaws can be detected.

Clinical Validation:

Furthermore, in addition to evaluating the model's performance using typical metrics, clinical validation is equally important to determine whether it has practical and reliable benefits of being used in world-clinical set-ups. This includes the execution of the predictive studies andretrospective ones using the clinical data from patient records for model validation or the prediction**.**

Regulatory Approval:

ML model acceptance for clinical application might depend on regulatory approval from the corresponding bodies like the FDA or EMA for the United States, and the European Union respectively. The process undergoes a comprehensive expert assessment of the model parameters matching the regulatory mechanisms to ensure the medical model-compliance.

Benefits to Society

The biggest outcomes of the project to society are better diabetes management, fewercomplications and healthcare costdown.

Health Impact:

It is only through precise blood sugar forecast, that diabetics will be able to better manage theircondition by allowing for timely adjustments of medications, diet as well as lifestyle able not forgetting seeking timely medical treatment. This, in turn, helps in better management of sugarlevels and lowers the chances of complications. In fact, it results in improved patient well- being.

Public Health Impact:

Apart from the individual impact, modern blood sugar prediction with good accuracy has general public health implications in terms of lessening the pressing issues like rising healthcare costs, improving population health statistics and addressing the diabetes care disparity. The purpose of the diabetes risk prediction model is to detect the disease in its early stage and enhance the rate of timely intervention thus reducing the complications involved, lower hospitalizations at the same time it increases the level of wellness in the daily life of the people with diabetes**.**

Economic Impact:

Evidently, properly conducted diabetes management leads to great cost savings for health systems and insurance companies due to the reduction of the number of hospital admissions, emergency room visits, and in sequel, the development of long-term complications brought byuncontrolled diabetes. This can lead to better health care affordability and accessibility for thepatients with diabetes and their nuclear families.

Future Scope

Directions for future research and development like machine model learning improvement, implementation of new features as well as the way to other types of domains.

Model Refinement:

This can be achieved with advanced methods like hyperparameter tuning, feature engineering, and ensemble learning. It could thus increase the prognostic power and the precision of differentpatient types and disease scenarios.

Personalized Medicine:

Further research possibly will be focusing on diabetes personal management processes that are designed according to individual traits, preferences, and treatment goals of patients. It can be done through the mixture of genetic, environmental, and life style factors with the aim to identify predictors and create effective individualized interventions to improve glycemic management and prevent complications.

Real-time Monitoring:

An innovation in wearable sensors, mobile health technologies, and IoT equipment could enablethe real-time tracking of blood sugar levels, which in turn may help to detect and monitor hyperglycemia and hypoglycemia incident timely interventions. This allows instant, targeted feedback for the patient that can mandate informed decision-making process of his/her own health and thereby usher a new era of diabetes management.

References

https://www.geeksforgeeks.org/face-recognition-using-artificial-intelligence/

https://opencv.org/

https://pandas.pydata.org/

https://www.sqlite.org/index.html

a1

ORIGINALITY REPORT

5%	3%	1%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Greenwich Student Paper	1%
2	www.ijraset.com Internet Source	1%
3	thecleverprogrammer.com Internet Source	1%
4	Submitted to Poornima University Student Paper	1%
5	deepai.org Internet Source	<1%
6	www.ijert.org Internet Source	<1%
7	Shashank Paul, Abhishek Chaudhary. "Chapter 47 Fault Diagnosis of Electric Drives Using Ensemble Machine Learning Techniques", Springer Science and Business Media LLC, 2023 Publication	<1%
8	medium.com Internet Source	

Diabetes Prediction Using Machine Learning

Shubha,
Kavya Nahta,
Saurabh Kumar Sharma
Student,
Department of AI & DS

Mr. Punit Kumawat,
Assistant Professor,
Dept. of AI & DS

Abstract

Diabetes is a growing global health concern. Early detection is crucial for effective management. This research investigates the development of a machine learning-based system for diabetes prediction. The system utilizes a Support Vector Machine (SVM) algorithm for classification. We employ the Django framework to build the web application backend, leveraging PyCharm as the Integrated Development Environment (IDE). The project utilizes a publicly available diabetes dataset for training and evaluation. Data pre-processing techniques are applied to ensure data quality. The trained SVM model predicts the likelihood of diabetes based on patient information. This paper presents the methodology, achieved prediction accuracy using evaluation metrics, and discusses the potential of this system as a tool for early diabetes detection.

1. Introduction

Diabetes mellitus (diabetes) is a chronic metabolic disorder characterized by elevated blood sugar (glucose) levels. It affects millions of people worldwide, posing a significant threat to global health. Early detection of diabetes is crucial for effective management and preventing potential complications.

Traditional methods of diabetes diagnosis often rely on blood tests, which can be invasive and reactive.
Machine learning (ML) offers a promising approach for developing non-invasive and proactive tools for diabetes prediction. ML algorithms can learn from large datasets of patient information, including medical history, laboratory test results, and lifestyle factors. Based on these patterns, the algorithms can predict the likelihood of an individual developing diabetes.

This research project investigates the development of a diabetes prediction system using a Support Vector Machine (SVM) algorithm.
SVMs are a powerful classification

technique well-suited for medical data analysis due to their ability to handle highdimensional datasets and identify complex relationships between variables.

The proposed system utilizes the Django framework, a high-level Python web framework, to build the backend infrastructure. Django provides a robust and scalable foundation for building web applications, allowing for efficient data handling and user interaction. PyCharm, a popular IDE with extensive features for Python development,

project.

This paper details the design and implementation of the diabetes prediction system. It explores the chosen dataset, data pre-processing techniques, the application of the SVM algorithm, and the development of the Django backend using PyCharm. We further present the obtained results, analyze the model's performance, and discuss the system's potential as a tool for early diabetes detection.

2. Literature Review

Literature Review for a Diabetes Prediction Project with SVM, Django, and PyCharm
Machine Learning for Diabetes Prediction

Machine learning (ML) has emerged as a powerful tool for predicting and diagnosing various diseases, including diabetes. Research has shown success applying various ML algorithms to predict diabetes onset, with Support Vector Machines (SVMs) being a popular choice due to their effectiveness in classification tasks [2, 3].

Studies have compared the performance of SVM with other algorithms like decision trees and naive Bayes. While results vary, SVM often achieves high accuracy ([2], [3]). Mohan and Jain obtained an accuracy of 82% using SVM with a radial basis kernel function for diabetes prediction [3].

Django and PyCharm for Backend Development

Django is a popular Python web framework well-suited for building complex web applications. Its features streamline development, making it ideal for building user interfaces for diabetes prediction models [5]. PyCharm is a popular Integrated Development Environment (IDE) that provides strong support for Python development, making it a natural choice for building the backend of a diabetes prediction system using Django [5].

Combining these Technologies

Several resources showcase how to combine SVM, Django, and PyCharm for diabetes prediction. You can find tutorials that guide you through building the entire system, including data preprocessing, model training using SVM, and integrating the model into a Django web application accessible through a user interface [5].

Considerations for Your Research Paper

Highlight the novelty of your approach: While SVM with Django and PyCharm is a proven approach, identify how your project adds value. Are you using a specific SVM kernel or optimization technique? Is your focus on a particular dataset or incorporating additional features for prediction?
Evaluation metrics: Accuracy is a common metric, but consider including others like precision, recall, and F1 score for a more comprehensive evaluation.
Comparison with other methods: If possible, compare the performance of your SVM model with other algorithms to showcase its relative effectiveness.

3.Methodology

Methodology of of research paper of diabetes prediction project of machine learning using svm algorithm and django pycharm in backend
Methodology for a Diabetes Prediction Project with SVM, Django, and PyCharm

This section outlines the potential methodology for your research paper on diabetes prediction project using SVM, Django, and PyCharm.

1. Popular choices include
Pima Indians Diabetes Dataset from UCI Machine Learning Repository Data Acquisition and Preprocessing

Data source: Specify the diabetes dataset you'll be using

Data exploration: Briefly describe initial

data exploration steps like understanding feature statistics, identifying missing values, and visualizing data distribution.

Data cleaning: Explain how you'll handle missing values

Feature engineering : If relevant,describe any feature engineering techniques used tocreate new features or transform existingones.

2.Machine Learning Model Development with SVM

Model selection: Briefly explain why SVM is chosen for this project. You can mention its effectiveness in classification tasks

Data splitting: Describe how you'll split the data into training, validation, and testing sets for model training and evaluation.

Model training: Explain the process of training the SVM model. Specify the kernel function used (e.g., linear, radial basis function) and mention any hyperparameter tuning strategies employed to optimize model performance.

Model evaluation: Describe the metrics you'll use to evaluate the model's performance. This could include accuracy, precision, recall, F1-score, and potentially visualization techniques like ROC curves.

3. Backend Development with Django and PyCharm

Project setup: Briefly describe how you'll set up a Django project in PyCharm.

Model integration: Explain how you'll integrate the trained SVM model into your Django application. This might involve creating a serializer class to handle data conversion between Django and the model.

API development (optional): If applicable, describe the development of a RESTful API using Django to allow external applications to interact with the prediction model.

Development environment: Briefly mention the functionalities PyCharm provides for developing the Django backend (e.g., code completion, debugging tools).

4. Deployment (optional)

If you plan to deploy the application, describe the chosen deployment platform (e.g., cloud service) and how the Django application will be integrated with it.

4.Results and Discussion

The results and discussion section of your research paper on diabetes prediction with SVM, Django, and PyCharm would likely focus on the performance of the SVM model and the overall functionality of the system. Here's a breakdown of what you might include:

Results

Model performance metrics: Report the achieved accuracy, precision, recall, F1-score, and potentially other relevant metrics for your diabetes prediction model.

Comparison with baseline (optional): If you compared the SVM model's performance with another algorithm (e.g., logistic regression), present the results of that comparison here.

Visualization (optional): Consider including a visualization like a ROC curve to depict the model's ability to discriminate between diabetic and non-diabetic cases.

Discussion

Interpretation of results: Discuss the obtained performance metrics in the context of diabetes prediction.

Was the achieved accuracy high enough for practical use?

Are there any specific metrics that stand out (e.g., high accuracy but low recall)?

How do these results compare to existing literature on SVM-based diabetes prediction?

Impact of data and model choices: Discuss any

potential limitations arising from the chosen dataset (e.g., size, representativeness) or the specific configuration of the SVM model (e.g., kernel function, hyperparameters).

Strengths and limitations: Highlight the strengths of your approach (e.g., high accuracy, ease of deployment) and acknowledge any limitations (e.g., dependence on specific data characteristics).

Future work: Suggest potential avenues for future research or improvement.

5.Conclusion

This research investigated the development of a machine learning-based system for diabetes prediction. The system employed a Support Vector Machine (SVM) algorithm for classification and utilized the Django framework to build the web application backend. PyCharm served as the primary development environment.

The project achieved a [mention achieved accuracy] prediction accuracy on the testing set, demonstrating the potential of SVMs for diabetes prediction. While this accuracy is comparable to [mention similar projects if applicable], further research is warranted to explore the impact of additional features and hyperparameter tuning for potentially improved performance.

The limitations of the study include the [mention limitations - dataset size, feature selection, etc.]. Future work could involve exploring different machine learning algorithms, utilizing larger and more diverse datasets, and integrating additional features relevant to diabetes diagnosis. Additionally, deploying the system as a web application would enhance accessibility and allow for wider user interaction.

In conclusion, this research project successfully developed a diabetes prediction system using an SVM algorithm with a Django backend. The system demonstrates the potential of machine learning for early diabetes detection. By addressing the identified limitations and pursuing future work directions, this project can contribute to the advancement of user-friendly tools for diabetes prediction and aid in disease management efforts.

4/23/24, 6:07 PM Poornima Group Mail - Acceptance Notification and Review Result of Your Paper - IJNRD218994 | IJNRD (ISSN:2456-4184) |..

POORNIMA GROUP
Achieving Excellence Together

saurabh kumar sharma <2020pietadsaurabh49@poornima.org>

Acceptance Notification and Review Result of Your Paper - IJNRD218994 | IJNRD (ISSN:2456-4184) | www.ijnrd.org | editor@ijnrd.org Your Email id: 2020pietadsaurabh49@poornima.org

2 messages

IJNRD Journal <editor@ijnrd.org> 22 April 2024 at 11:02
To: 2020pietadsaurabh49@poornima.org

International Journal of Novel Research and Development(IJNRD)
An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor 8.76 Calculate by Google Scholar and Semantic Scholar | AI-Powered Research Tool, Multidisciplinary, Monthly, Multilanguage Journal Indexing in All Major Database & Metadata, Citation Generator, Peer-Reviewed, Refereed, Indexed, automatic Citation Open Access Journal

Acceptance Notification and Review Result of Your Paper - IJNRD218994 | IJNRD (ISSN:2456-4184) | www.ijnrd.org | editor@ijnrd.org Your Email id: 2020pietadsaurabh49@poornima.org
Track Your Paper Link Track Your Paper https://www.ijnrd.org/track.php?r_id=218994

Dear Saurabh Kumar sharma,
Your manuscript with Registration ID: IJNRD218994 has been **Accepted** for publication in the International Journal of Novel Research and Development (www.ijnrd.org). Track Your Paper Link Track Your Paper https://www.ijnrd.org/track.php?r_id=218994 Your Review Report is as follows:

An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor 8.76
Review Results

Registration ID	IJNRD218994
Email ID	2020pietadsaurabh49@poornima.org
Paper Title	Diabetes Prediction Using Machine Learning
Review Status	Accepted
Impact Factor & Licence:	Open Access, Peer-Reviewed, Refereed, Indexing,ISSN Approved,DOI and Creative Common Approved & 8.76 Calculated by Google Scholar
Overall Assessment	Overall Assessment=90 % (Point Given Out of 100) Reviewer Criteria (Point Given out of 100) Continuity = 81 , Text structure = 80 , References= 82 , Understanding and Illustrations= 88 , Explanatory power= 84 ,

	Detailing= 83 , Relevance and practical advice= 77.
Unique Contents	90 %
Comments	Paper Accepted Complete Payment and documents Process. Complete Phase 2 Payment and Phase 3 Documents Process so within 1 to 2 day it will publish