

**Department of Computer Science and Engineering**

**Graphic Era Hill University**

**Dehradun, Uttarakhand**

**2024-25**

**3rd Semester Mini Project Report on**

image3.png

**“DIABETES PREDICTION”**

image3.png

**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

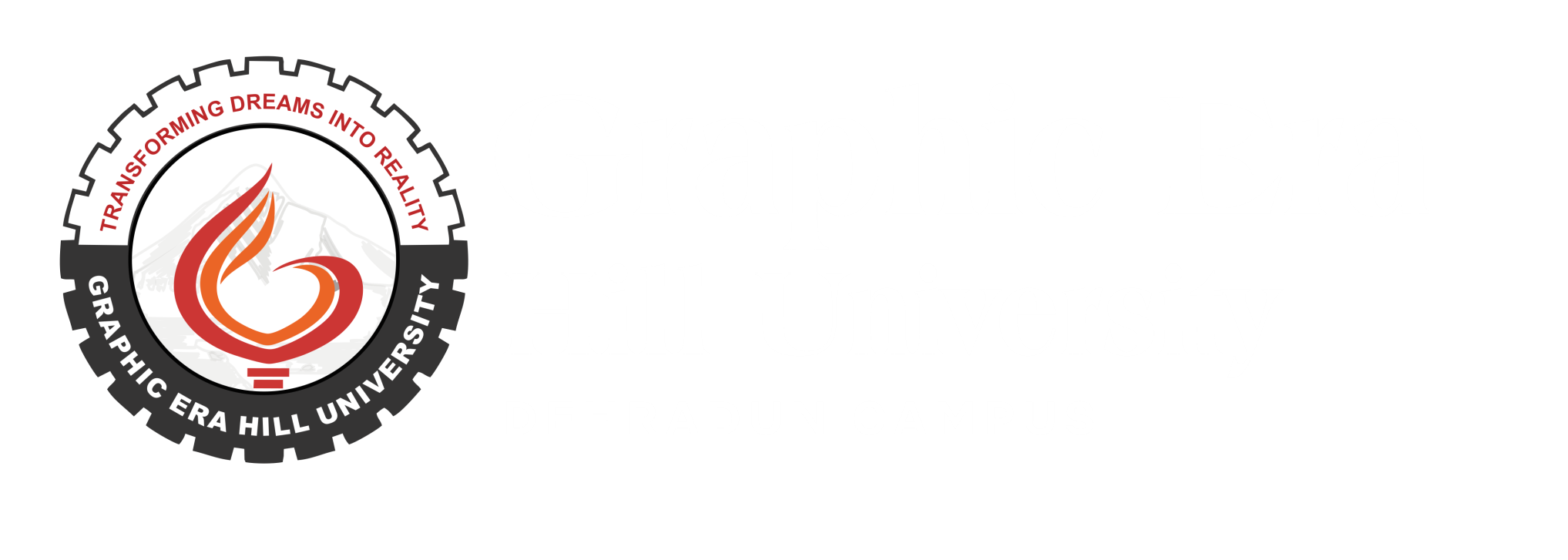
**COMPUTER SCIENCE & ENGINEERING (AI-ML)**

**Submited by Under the Guidance of**

Ashutosh Rauthan prof.Amit Gupta

**University rollno:**2318544

**Sec:**K1



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“DIABETES PREDICTION”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering **(AI-ML)** in the Department of Computer Science and Engineering of the Graphic Era Hill University, Dehradun shall be carried out by the undersigned under the supervision of **Amit Gupta, Professor**, Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun.

Name- Ashutosh Rauthan University Roll no-2318544

The above mentioned student shall be working under the supervision of the undersigned on the **“DIABETES PREDICTION”**

**Supervisor** **Head of the Department**



**CERTIFICATE**

### Certified that **Mr. Ashutosh Rauthan** University roll no-**2318544** has developed mini project on **“DIABETES PREDICTION ”** for the CSE III Semester Mini Project in Graphic Era Hill University, Dehradun. The project carried out by Students is their own work as best of my knowledge.

Date: 11-01-2025

**Class Co-ordinator**

**Prof Akash Chauhan Project Guide**

**CSE-K1-III-Sem Prof Amit Gupta**

(CSE Department) ( CSE Department )

GEHU Dehradun GEHU Dehradun

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction and Problem Statement | **5** |
| Chapter 2 | Methodology | **7** |
| Chapter 3 | Project Work Carried Out | **8** |
| Chapter 4 | Results and Discussion | **9** |
| Chapter 5 | Conclusion and Future Work | **10** |
| Chapter 6 | References | **11** |

**Chapter 1**

**Introduction and Problem Statement**

In the following sections, a brief introduction and the problem statement for the work has been included.

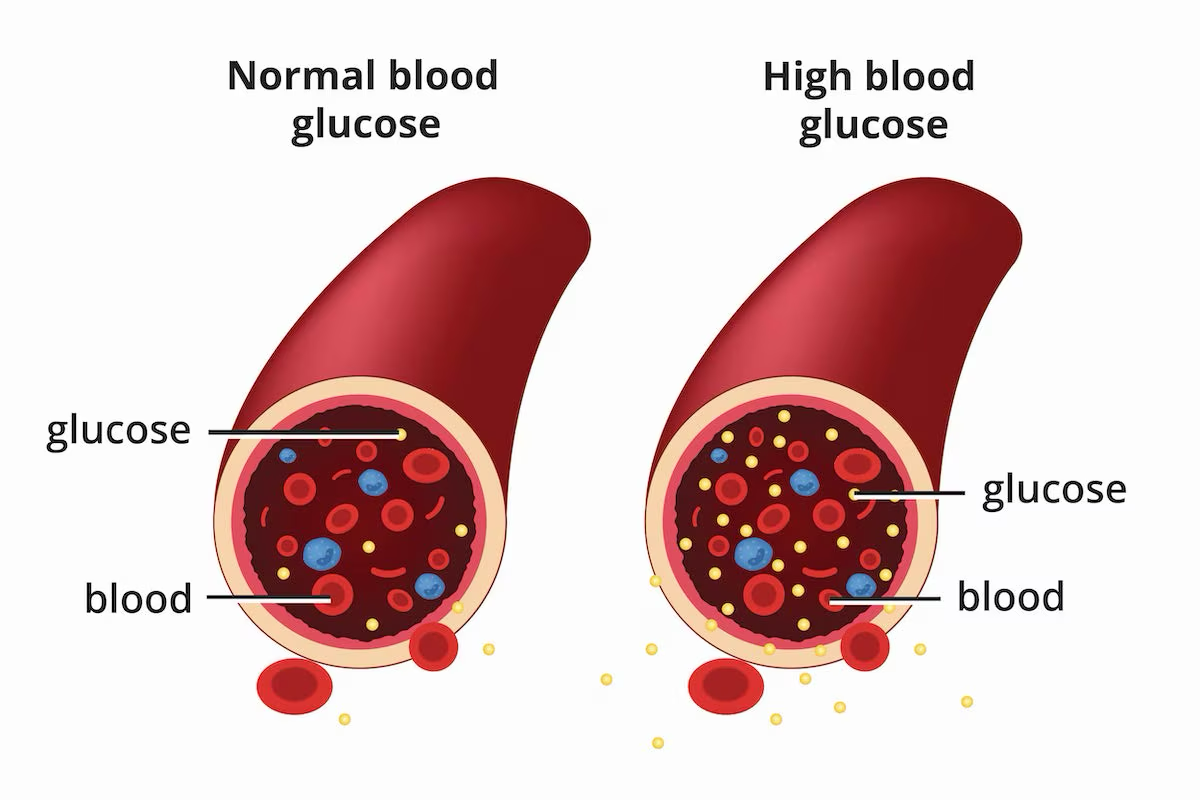
* 1. **Introduction**

**What is diabetes ?**

Diabetes is a chronic disease that occurs when the body doesn't produce enough insulin or can't use insulin properly, resulting in high blood sugar levels.

**Facts:** Diabetes, a chronic disease affecting millions worldwide, poses substantial challenges to healthcare systems. The International Diabetes Federation (IDF) projects a substantial rise in the global prevalence of diabetes among adults by 2045.

Early prediction and diagnosis of diabetes are pivotal in mitigating its impact and enhancing patient’s quality of life. Machine learning techniques have emerged as powerful tools in analyzing medical data and predicting diseases like diabetes. The potential of these techniques to discern patterns and correlations in intricate datasets has garnered extensive research attention in recent years**.**



**Fig1.1** : High glucose level in blood

* 1. **Problem Statement**

The problem statement for the present work can be stated as follows:

**Why This Project?**

* Diabetes affects millions globally, leading to severe health complications if untreated.
* Early prediction helps in timely intervention and reduces healthcare costs.

Diabetes remains a significant global health concern, often leading to severe complications if not detected and managed early. Conventional diagnostic approaches are typically invasive, costly, and time-consuming,

**Objectives**

The objective of this project is to develop a machine learning-based diabetes prediction system that analyzes readily available health indicators such as Pregnancies, Age, BMI, Glucose Levels, Insulin, Blood Pressure, etc this system aims to provide a non-invasive, cost-effective, and accurate prediction model, enabling healthcare providers and individuals to identify at-risk patients early and implement timely interventions. By leveraging advanced algorithms and feature engineering, the proposed solution seeks to contribute to more proactive and data-driven diabetes management.

* Develop a machine learning model to predict diabetes.
* Use healthcare data for training and testing.
* Handling missing values.
* Identify significant features for diabetes prediction
* train machine learning models
* Predict the likelihood of diabetes for unseen data and analyze contributing factors.
* Evaluate the model's accuracy and reliability.

**Chapter 2**

**Methodology**

The methodologies used are as follows:

**Used methodologies**

1. **Data Collection:** Acquired dataset with key health indicator.
   * Features : Pregnancies, Age, BMI, Glucose Levels, Insulin, Blood Pressure, etc.
   * Size: (768, 9)
   * Source- Kaggle

1. **Preprocessing:**  Cleaned and prepared the data for analysis.

Preprocessing Steps:

* + Handling missing values.
  + Normalization and scaling.

1. **Model Selection:** Tried multiple machine learning algorithms:
   * Logistic Regression
   * Random Forest
   * Support Vector Machine (SVM)
2. **Model Used:** Support Vector Machine (SVM)

* + High accuracy
  + Effective in handling high-dimensional data
  + Parameters Tuned: kernel type- linear

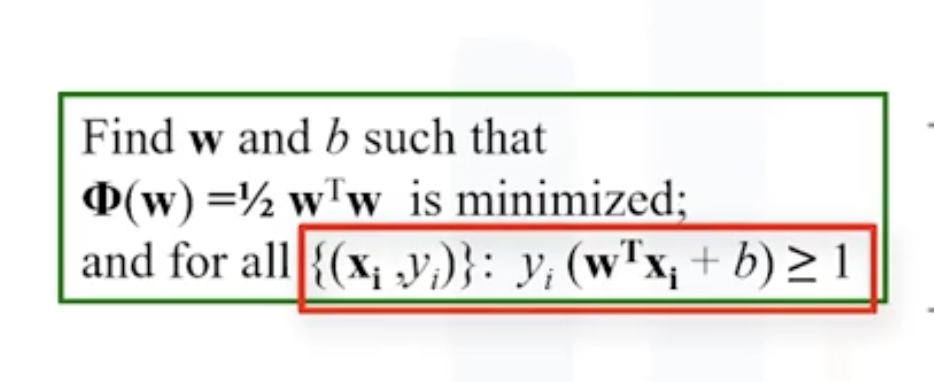
1. **Evaluation:** Used metrics like accuracy, precision, recall, and F1-score.

**Chapter 3**

**Project Work Carried Out**

This section include implementation of objectives, pseudo code/ algorithms.

|  |
| --- |
| **Input.**  df the dataset, *cleaning and preprocessing*  **begin**   1. **Model**-SVM 2. **Kernel** type-‘**rbf**’(Radial Basis Function) 3. **Operations:**   Mapping data to a high dimensional feature space  Finding a separator( fig3.1 )  Classification using the equation  (These operations is predefined inside the model form library)   1. Model is trained 2. Testing the model 3. **Evaluation:** Used metrics like accuracy, precision, recall, and F1-score. 4. **Input:** unseen data input 5. **Output**: class   **End** |

**Table 3.1** Pseudo code of the SVM model algorithm

**fig 3.1** : equation to find separator hyperplan.

**Accuracy=**TP+TN/TP+TN+FP+FN

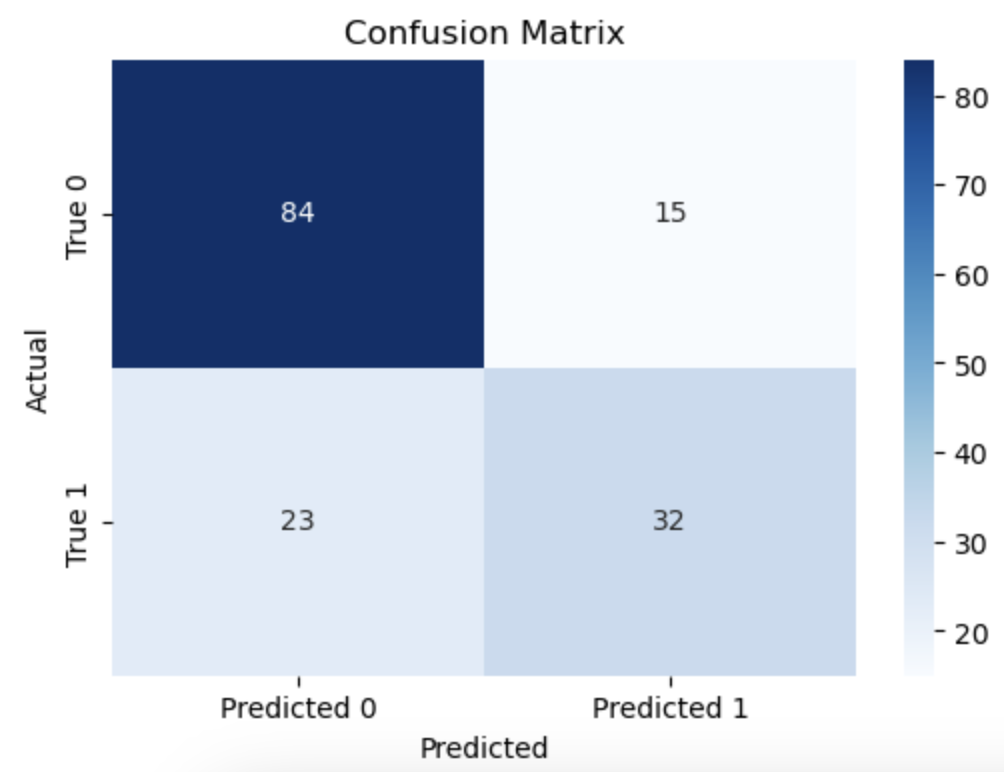
**Precision**=TP/TP+FP

**Recall**=TP/TP+FN

**F1 Score**=2⋅Precision⋅Recall/Precision+Recall

**Chapter 4**

**Results and Discussion**

The model is trained and tested on the datasets with the following results.

**Accuracy=**0.75 or 75%

**Precision**=0.75

**Recall**=0.75

**F1 Score**=0.75

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classification | Report |  |  |  |
|  |  |  |  |  |
|  | precision | recall | f1-score | support |
|  |  |  |  |  |
| 0 | 0.79 | 0.85 | 0.82 | 99 |
| 1 | 0.68 | 0.58 | 0.63 | 55 |
|  |  |  |  |  |
| accuracy |  |  | 0.75 | 154 |
| macro avg | 0.73 | 0.72 | 0.72 | 154 |
| weighted avg | 0.75 | 0.75 | 0.75 | 154 |

**Chapter 5**

**Conclusion and Future Work**

Diabetes prediction model is implemented and tested .

In this mini-project, an **SVM model with an RBF kernel** was successfully implemented to predict diabetes. The model demonstrated good performance in classifying the data, with avg results in terms of accuracy, precision, recall, and F1-score. The **confusion matrix** confirmed the model’s ability to differentiate between diabetic and non-diabetic individuals.

While the model showed promising results, hyper parameter tuning and data preprocessing could further improve its performance. This project highlights the potential of machine learning in healthcare, specifically for early diabetes detection, and serves as a foundation for future enhancements.

* **Challenges Faced:**
  + imbalanced dataset.
  + Handling missing or inconsistent data.
  + Deciding on the best model for the given datasets.
* **Key Takeaways:**
  + Machine learning can effectively predict diabetes.
  + Data quality significantly impacts model performance.
* **Future Scope:**
  + Improve accuracy with larger datasets.
  + Explore deep learning models.

**References**

**Chapter 6**

1. IBM (Machine Learning with Python)
   * IBM. (n.d.). Machine Learning with Python: A Practical Introduction. Retrieved from https://www.ibm.com/training
2. YouTube
   * [YouTube](https://www.youtube.com/) for tutorial videos on machine learning and Python libraries..