

**The Superior University Gold Campus**

**A.I Lab**

**Task 13**

**Project Report: Diabetes Prediction System**

**1. Introduction**

* **Objective**: Develop a machine learning system to predict the likelihood of diabetes based on user-provided health data.
* **Scope**: Create an end-to-end pipeline that includes data processing, model training, evaluation, and deployment as a web application.
* **Key Features:**
  + User-friendly interface for health data input.
  + Real-time predictions based on a machine learning model.

2. **Methodology**

2.1 **Data Collection**

* **Dataset:** Medical data sourced from publicly available datasets, containing features like glucose level, BMI, and more.
* **Features used:**
  + Pregnancies
  + Insulin
  + Age
  + BMI
  + Blood Pressure
  + Glucose
  + Skin Thickness
  + Diabetes Pedigree Function

**2.2 Data Preprocessing**

* Cleaned missing or inconsistent data.
* Normalized features for uniform scaling.

**2.3 Data Visualization**

* Explored correlations between features.
* Identified trends using tools like matplotlib and seaborn.

**2.4 Model Training**

* **Algorithms Used:**
  + Logistic Regression
  + Decision Trees
  + Random Forest
* **Tools and Libraries:** scikit-learn, pandas, numpy.
* **Training Approach:**
  + Split data into training and testing sets (e.g., 80-20 split).
  + Evaluated models on accuracy, precision, recall, and F1-score.

**2.5 Model Deployment**

* Framework: Flask.
* Hosting Platform: Heroku.
* Functionalities:
  + Input: User health data via a web form.
  + Output: Displays prediction results (Diabetic/Non-Diabetic).

**3. Results**

* Model Performance:
  + Logistic Regression: X% Accuracy
  + Decision Trees: Y% Accuracy
  + Random Forest: Z% Accuracy
* Chosen Model: [Insert best-performing model here].
* Web Application: Successfully deployed and tested for usability.

**4. Challenges**

* Handling missing data and ensuring clean preprocessing.
* Optimizing model performance without overfitting.
* Streamlining deployment for scalability on Heroku.

**5. Future Enhancements**

* Incorporate advanced algorithms like neural networks.
* Add user authentication for a personalized experience.
* Expand to include more comprehensive health data.
* Enhance UI/UX for a better user experience.

**6. Conclusion**

* Successfully created and deployed a machine learning-based diabetes prediction system.
* The project demonstrates the integration of data science and software development to solve a real-world problem.