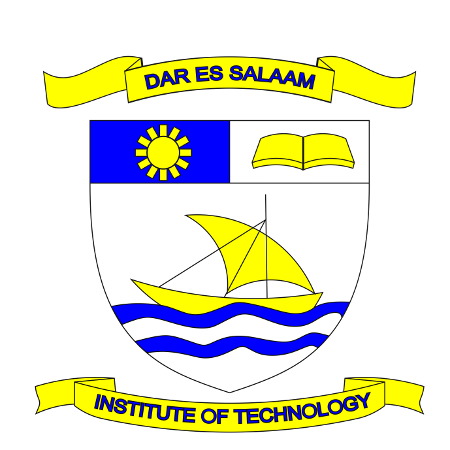
**DAR ES SALAAM INSTITUTE OF TECHNOLOGY**



**ARTIFICIAL INTELLIGENCE GROUP PROJECT**

**PROJECT TITLE: AI-BASED DIABETES PREDICTION**

**GROUP MEMBERS**

|  |  |  |
| --- | --- | --- |
| NO. | NAME | REGISTRATION NO. |
| 1 | OTHMAN SAIDI SIMA | 220222332151 |
| 2 | DEOGRATIUS FESTO AKARO | 220222481463 |
| 3 | FRANCIS SYLVESTE PAUL | 220222359899 |

**Phase 1**

Deliverable: Problem Identification & Title Justification

1. Problem Identification

Diabetes is a chronic disease that is the no. 7 cause of deaths in 2021 (World health statistics 2024: monitoring health for the SDGs, sustainable development goals), with delayed diagnosis leading to severe complications like cardiovascular disease and kidney failure. Current diagnostic methods rely on blood tests (e.g., fasting glucose, HbA1c) and clinical evaluations, which are:

* Time-consuming for both patients and healthcare providers.
* Cost-prohibitive in low-resource settings.
* Reactive, often detecting diabetes only after symptoms manifest.

Selected Problem:  
Develop an interpretable and lightweight AI model to predict diabetes risk using diagnostic measurements (e.g., glucose, BMI, insulin levels). This tool aims to complement (not replace) medical expertise, enabling proactive healthcare interventions.

2. Relevance Justification

Public Health Need: Early detection can reduce treatment costs.

AI Advantages:

* Pattern Recognition: Machine learning can uncover non-linear relationships (e.g., age × BMI) that manual analysis might miss.
* Scalability: A deployable model could integrate with electronic health records (EHRs) or mobile health apps, democratizing access to screening.
* Ethical Alignment: Addresses healthcare disparities by providing low-cost screening tools for underserved populations.

3. Title Defense & Scope

Scope:

Binary Classification: Predict "diabetic" or "non-diabetic" outcomes.

Target Audience:

Primary care clinics and telehealth platforms.

Limitations:

Focuses on structured clinical data; excludes genetic or lifestyle factors.

AI Approach:

* Dataset: Pima Indians Diabetes Dataset (Kaggle), containing 768 instances with 8 features (e.g., pregnancies, glucose, BMI) and labeled outcomes.
* Preprocessing: Address missing values (e.g., insulin = 0 in some cases) and normalize features.

Model Choice:

* Logistic Regression: Prioritized for interpretability, allowing clinicians to trust predictions (e.g., identifying glucose as a top risk factor).
* Decision Trees: Backup option for visualizing decision pathways.

Evaluation Metrics:

* Recall: Minimize false negatives (critical in healthcare).
* F1-Score: Balance precision and recall.

Deliverable Summary

This project combines healthcare accessibility and AI transparency to create a practical tool for early diabetes detection. By leveraging supervised learning on a well-established dataset, the model aims to provide actionable insights for healthcare providers while adhering to ethical standards in medical AI.