UNIVERSITI TEKNIKAL MALAYSIA MELAKA
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

BITI 2233 MACHINE LEARNING
PROJECT PROPOSAL

A	**TITLE OF PROPOSED PROJECT:**	**GROUP NO:** **7**
	DIABETES PREDICTION SYSTEM USING ML	

B **DETAILS OF STUDENT** / *MAKLUMAT PELAJAR*

B(i)

Name of Student:	**Student matric no.:**	**E-mail Address:**
1. HARIHARAN A/L R.MOHAN	B031910477	b031910477@student.utem.edu.my
2. PAVEETHERAN A/L THINAGARAN	B031810470	b031810470@student.utem.edu.my
3. KRISHNAR RAM A/L MURUGAN	B031910013	b031910013@student.utem.edu.my
4. AZFAR RAHMAN BIN FAZUL RAHMAN	B031910467	b031810467@student.utem.edu.my

C **PROJECT INFORMATION /** *MAKLUMAT PROJEK*

C(i) **Executive Summary of Project Proposal**

Diabetes is considered as one of the deadliest and chronic diseases which causes an increase in blood sugar. Many complications occur if diabetes remains untreated and unidentified. The tedious identifying process results in visiting of a patient to a diagnostic center and consulting doctor. But the rise in machine learning approaches solves this critical problem. The motive of this study is to design a model which can prognosticate the likelihood of diabetes in patients with more accuracy.

Therefore, machine learning classification algorithm namely Random Forest Tree is used in this experiment to detect diabetes at an early stage. Experiments are performed on Pima Indians Diabetes Database (PIDD) which is sourced from Kaggle.

C(ii) **(a) Project Background**

Classification strategies are broadly used in the medical field for classifying data into different classes according to some constrains comparatively an individual classifier. Diabetes is an illness which affects the ability of the body in producing the hormone insulin, which in turn makes the body mass index abnormal and raise the levels of glucose in the blood. In Diabetes a person generally suffers from high blood pressure and many complications occur if diabetes remains untreated.

Many researchers are conducting experiments for diagnosing the diseases using various classification algorithms of machine learning approaches like SVM, Naive Bayes, Decision Tree, Decision Table as researches have proved that machine-learning algorithms works better in diagnosing different diseases. Machine learning algorithms gain its strength due to the capability of managing a large amount of data to combine data from several different sources.

.

This research work focuses on female suffering from diabetes. In this work, Random Forest Tree, classification algorithm is used and evaluated on the PIDD (Pima Indians Diabetes Database) dataset to find the prediction of diabetes in a patient. Experimental performance of the algorithm are measures and achieved good accuracy.

(b) Objective (s) of the Project
Objektif Projek

1. To build an efficient machine learning model.
2. To predict the accurate classification algorithm method.
3. To predict whether the patient has diabetes or not.
4. To develop a system and apply machine learning algorithm.

(c) Scope
Skop

1. Module to be developed

a. User Management Module –Interface: user-friendly, easier to use for input controls such as using range slider and for showing the result whether have diabetes or not.

2. Target User

a. This system can be widely used by female over 21 years old.

(d) Project Significance
Kepentingan projek

1. *Motivation and inspiration for the project.*
The findings of this study will redound to the benefit of identifying early symptoms of diabetes in order to identify the occurance of diabetes and also it is an easier task for the target to identify whether have diabetes or not. For the researcher, this project will help them to use or upgrade for the future finding process without any problems.

The main advantages of the system are:
- When it is compared to manual diabetes testing process, this system saves the time and also helps target users to analyze it without any problems.

(e) Expected Outcome: *Data and Statistic Visualization, Graph chart, User Input Table, Train Accuracy, Model Test Accuracy Score,Classification Result*

(f) Tools And Algorithm To Use

TOOLS	ALGORITHM
PYTHON **ANACONDA NAVIGATOR** **GOOGLE COLAB**	**RANDOM FOREST TREE**

(g) Appendix

UCI Machine Learning. (2016). *Pima Indians Diabetes Database*. Kaggle.com. https://www.kaggle.com/uciml/pima-indians-diabetes-database

Chaki, J., Thillai Ganesh, S., Cidham, S. K., & Ananda Theertan, S. (2020). Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2020.06.013

Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders*, *19*(1), 391–403. https://doi.org/10.1007/s40200-020-00520-5

Asel Mendis. (2018, November 6). *Data Retention: Handling Data with Many Missing Values and Less Than 1000 Observations*. Medium; Towards Data Science. https://towardsdatascience.com/easy-data-science-with-r-and-python-diabetes-among-the-pima-indians-an-exploratory-analysis-d7bc321d0fa7

Koehrsen, W. (2018, August 30). *An Implementation and Explanation of the Random Forest in Python*. Medium; Towards Data Science. https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76

Schulz, L. O., Bennett, P. H., Ravussin, E., Kidd, J. R., Kidd, K. K., Esparza, J., & Valencia, M. E. (2006). Effects of Traditional and Western Environments on Prevalence of Type 2 Diabetes in Pima Indians in Mexico and the U.S. *Diabetes Care*, *29*(8), 1866–1871. https://doi.org/10.2337/dc06-0138