



**UTM**  
UNIVERSITI TEKNOLOGI MALAYSIA

**SCHOOL OF COMPUTING**  
Faculty of Engineering

Project Proposal Form MCST1043  
Sem: 2 Session: 2024/25

## SECTION A: Project Information.

---

Program Name: **Masters of Science (Data Science)**

Subject Name: **Project 1 (MCST1043)**

Student Name: NURHAFIZAH BINTI MOHD YUNOS

Metric Number: MCS241048

Student Email & Phone: [nurhafizah99@graduate.utm.my](mailto:nurhafizah99@graduate.utm.my) & 0143680885

Project Title: Early Prediction of Stroke Risk Using Machine Learning for Preventive Healthcare

Supervisor 1: \_\_\_\_\_

Supervisor 2 / Industry  
Advisor(if any): \_\_\_\_\_

## SECTION B: Project Proposal

---

### Introduction:

Stroke is a leading cause of death and disability worldwide. The early detection of individuals at high risk can greatly enhance preventive care and alleviate healthcare burdens. This project aims to develop a machine learning-based model designed to predict stroke risk using clinical and lifestyle data. By analyzing patient records with key indicators such as age, hypertension, heart disease, BMI, and smoking status, the model will uncover patterns associated with stroke occurrences. The ultimate objective is to support early intervention strategies and advance intelligent, data-driven preventive healthcare systems.

Strokes are medical emergencies that arise when the blood supply to a part of the brain is interrupted or reduced, often resulting in long-term disability or death. According to the World Health Organization (WHO), strokes rank among the top causes of death globally, with millions of new cases reported each year. The early identification of at-risk individuals is essential for implementing timely preventive measures. With the increasing availability of electronic health records and health data, machine learning is poised to provide more efficient and accurate predictions of stroke risk compared to traditional methods. This project explores machine learning techniques to analyze patient data and assess the likelihood of stroke, thereby contributing to the field of preventive healthcare.

### Problem Background:

Despite advancements in medical science, many stroke cases remain undetected until it is too late for effective intervention. Traditional risk assessment models often depend on manual evaluation, which can be time-consuming, subjective, and inconsistently applied across different healthcare providers. The absence of scalable, automated systems for assessing stroke risk underscores the necessity for intelligent tools capable of processing vast datasets and identifying complex patterns that may elude human observers. The integration of machine learning in healthcare presents a valuable opportunity to develop predictive models that enhance clinical decision-making and public health initiatives focused on early detection and prevention.

**Problem Statement:**

There is currently a gap in effective and scalable solutions for the early prediction of stroke risk using data-driven methods. Traditional approaches often struggle to manage large, multidimensional datasets and frequently fail to generalize across diverse populations. This project aims to develop a robust machine learning model that can accurately predict stroke risk based on key health indicators. By doing so, it will facilitate proactive medical intervention and improve patient outcomes.

**Aim of the Project:**

The aim of this project is to develop and evaluate a machine learning-based system that can predict the risk of stroke in individuals using clinical and demographic data, thereby aiding in early intervention and enhancing preventive healthcare efforts.

**Objectives of the Project:**

The objectives of the project are:

1. To collect and preprocess relevant healthcare data for stroke prediction.
2. To explore and select appropriate machine learning algorithms for classification.
3. To build and train predictive models using historical patient data.
4. To identify and analyze key features contributing to stroke risk.

**Scopes of the Project:**

In the Data Handling, collecting and preprocessing a pertinent dataset sourced from reputable platforms such as Kaggle or UCI, ensuring it includes essential clinical and lifestyle features associated with strokes. For Feature Engineering, identify and transform critical variables, including age, hypertension, diabetes, body mass index (BMI), and smoking status. Regarding Algorithm Selection, we implement a range of machine learning models, including Logistic Regression, Decision Trees, Random Forests, and Support Vector Machines, to effectively address binary classification challenges. In the area of Model Evaluation, we employ various performance metrics, such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC, to assess the validity of our predictions. For Visualization and Interpretation, utilizing advanced data visualization tools to clarify the model's decision-making processes and to assess feature importance.

**Expected Contribution of the Project:**

This project seeks to enhance the field of preventive healthcare by showcasing the use of machine learning for early stroke prediction. It aims to provide insights into the critical health factors that contribute to stroke risk, as well as deliver a prototype tool to aid clinicians in identifying at-risk individuals. Additionally, it encourages further research into AI-driven solutions for early diagnosis.

**Project Requirements:**

Software: Python(Jupyter/Google Colab), Pandas, Scikit-learn, Matplotlib, Seaborn, Power BI(dashboard)

Hardware: Computer with 8GB RAM and 4GB GPU

Technology/Technique/  
Methodology/Algorithm: Data Collection (Kaggle), Data Preprocessing, Exploratory Data Analysis (EDA),  
Model Development, Model Evaluation

**Type of Project (Focusing on Data Science):**

☐ ☐ Data Preparation and Modeling

☐ ☐ Data Analysis and Visualization

☐ ☐ Business Intelligence and Analytics

☐ ☐ Machine Learning and Prediction

☐ ☐ Data Science Application in Business Domain

**Status of Project:**

☐ ☐ New

☐ ☐ Continued

If continued, what is  
the previous title?

**SECTION C: Declaration**

I declare that this project is proposed by:

☐ ☐ Myself

☐ ☐ Supervisor/Industry Advisor ( )

Student Name: NURHAFIZAH BINTI MOHD YUNOS

*Nurhafizah*

Signature

17/4/2025

Date

**SECTION D: Supervisor Acknowledgement**

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1:

Signature

Date

Name of Supervisor 2 (if any):

Signature

Date

**SECTION E: Evaluation Panel Approval**

The Evaluator(s) shall complete this section.

**Result:**

☐ ☐ FULL APPROVAL

☐ ☐ CONDITIONAL APPROVAL (Major)\*

☐ ☐ CONDITIONAL APPROVAL (Minor)

☐ ☐ FAIL\*

\* Student has to submit new proposal form considering the evaluators' comments.



.....

.....

.....

.....

.....

.....

.....

.....

.....

Name of Evaluator 1: .....

.....  
**Signature** .....

.....  
**Date** .....

Name of Evaluator 2: .....

.....  
**Signature** .....

.....  
**Date** .....