Title: **Medical Text Classification Using Supervised Machine Learning**

Abstract:
In this project, we aim to classify medical abstracts into predefined categories using machine learning models. The medical texts are first preprocessed by tokenizing, stemming, and removing stopwords, ensuring that the data is clean for further analysis. Two primary approaches are used to represent the cleaned text data: **Bag-of-Words (TF-IDF)** and **Word Embeddings (Word2Vec)**. These text representations are then used as features to train and evaluate three machine learning models: **Naive Bayes**, **Decision Trees**, and **Support Vector Machines (SVM)**.

The goal is to determine which machine learning algorithm performs best on this medical text classification task. Initially, the models are trained and evaluated on a subset of the training data, using cross-validation techniques to compare the effectiveness of each model. Finally, we will compare the performance of the models in terms of precision, recall, accuracy, and F1-score, and select the most effective model for medical text classification.