Instructor:Ali Sabra

**Alzheimer’s Disease Detection**

**By Omar Al Zoghbi**

**Objective:**

**This project aim to build 2 machine learning model that will predict if a patient has Alzheimer's disease or not based on a given dataset. The model will analyze the data to identify patterns that indicate the presence of the disease. By doing so, it aims to assist in early detection and improve diagnostic accuracy.**

**Dataset source:**

**Obtained from Kaggle.**

**Dataset Info:**

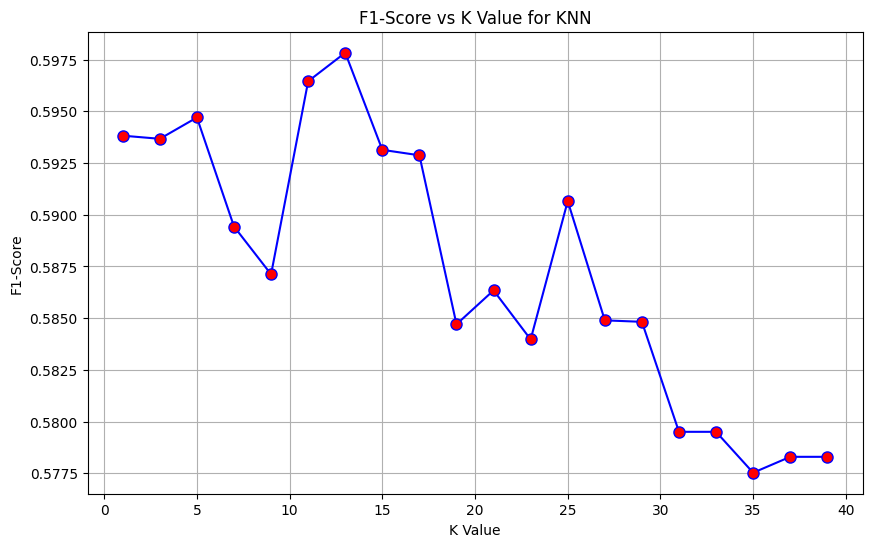
**This dataset contains extensive health information for 2,149 patients, each uniquely identified with IDs ranging from 4751 to 6900. The dataset includes demographic details, lifestyle factors, medical history, clinical measurements, cognitive and functional assessments, symptoms, and a diagnosis of Alzheimer's Disease.**

**Dataset feature engineering:**

**It refers to the process of transforming raw data into meaningful features that improve the performance of machine learning models (feature selection, handling missing/duplicated values, encoding categorical variables, and normalizing or scaling data).**

**K Best Value:**

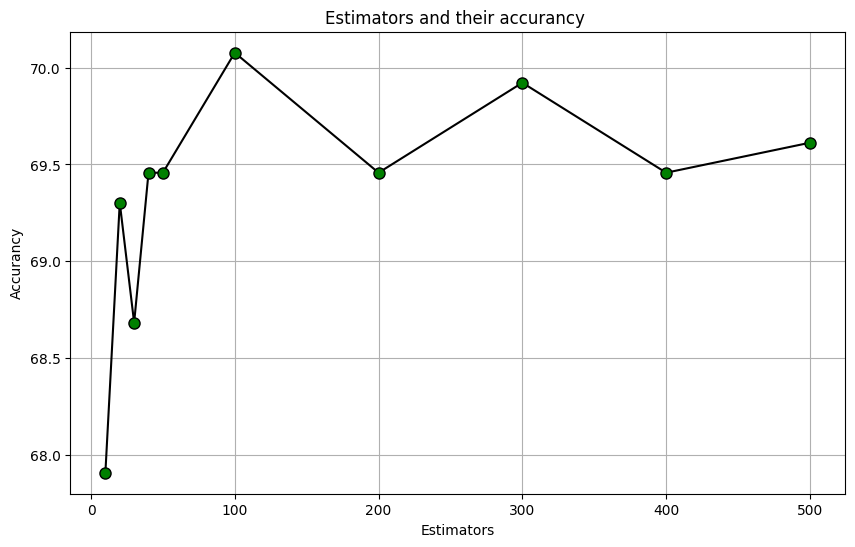
**After applying feature engineering and splitting the data into training and testing we need to find the best value of K which is the number of nearest neighbors to make an accurate prediction through a performance metric called F1-Score.**

****

**As the f1 score is near to 1, the K value is better and vice versa.**

**N estimators Best Value:**

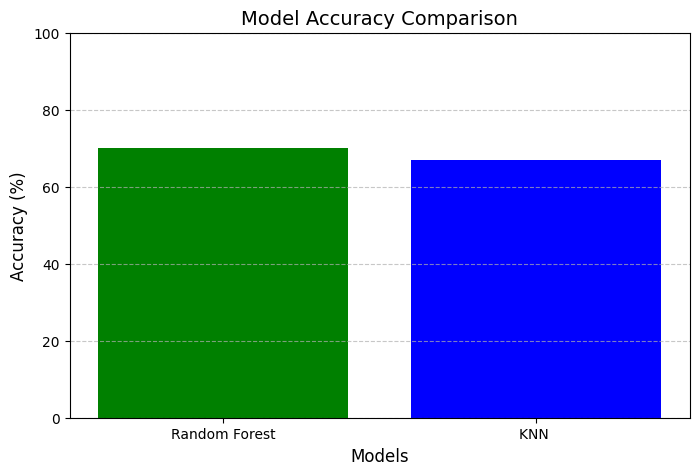
**For the Random Forest Classifier we need to find the best n value which is the number of trees in random forest to make an accurate prediction through a performance metric called accuracy score.**

****

**The best number of trees was 100 giving as an accuracy of approximately 70%.**

**Comparing the two models:**

**After inserting the k value and the n estimator to KNN and Random Forest Classifier respectively, a comparison take place to compare these two algorithms.**

****

**Results Overview:**

**In this project, we evaluated the performance of two machine learning models k-Nearest Neighbors (KNN) and Random Forest Classifier for detecting Alzheimer’s disease. The models were trained on a dataset containing various features related to patient health. After evaluating both models using metrics such as accuracy score and F1 score, we observed that while both models performed well, their strengths varied. Further fine-tuning and optimization could enhance the accuracy of both models in real-world applications.**