

**INTRODUCTION TO DATA SCIENCE**

**BITI2513**

**CAPSTONE PROJECT:**

**TASK 1 (Parkinson Disease Detection with XGBoost)**

|  |  |
| --- | --- |
| **NAME:** | **MATRIC NUM:** |
| Shamhinidevi Jayabalan | B031820098 |
| Ain Shazryna Binti Abdul Rahim | B031810157 |
| Azeera Fazreenaa Binti Aminuddin | B031810411 |
| Muhammad Farid bin Ed Noor | B031810455 |

# **1. Introduction**

Parkinson’s Disease (PD) is a progressive neurodegenerative movement disease affecting approximately 2% of people at the age of 65 and is the most second most commonly occurring neurodegenerative disease in the elderly (after Alzheimer’s Disease), with more than 6.3 million people worldwide with PD. In PD sufferers, loss of dopamine-producing neurons results in a range of both motor and non-motor symptoms and currently there is no cure, no means of slowing the disease progression, and no means of prevention. From the perspective of patient quality of life, PD is one of the most severe of all chronic diseases.

Therefore, in this project we will implement machine learning algorithm in python language to detect the presence of PD. XGBoost is a new Machine Learning algorithm designed with speed and performance in mind. XGBoost stands for eXtreme Gradient Boosting and is based on decision trees. In this project, we will import the XGBClassifier from the xgboost library; this is an implementation of the scikit-learn API for XGBoost classification.

# **2. Objective**

The objective of this project is:

1. To build a model to accurately detect the presence of Parkinson’s disease in an individual.
2. To prevent the disease from being worsen.
3. To analyze the Parkinson disease that affect individual.

# **3. Goals**

It has 5 stages to it and affects more than 1 million individuals every year. Generally, by the time of diagnosis of Parkinson Disease (PD), the disease is already well advanced, significant neuron loss and damage has already occurred, and any possibility of delaying further disease progression or providing neuroprotection is unlikely. The goal must be to diagnose and treat PD well before the irreversible destructive changes have taken place, ideally at least 5 years earlier than is currently the case. In addition, because the most severe symptoms occur in the advanced stages of the disease, strategies aimed at early detection and treatment will have the most benefit.

The detection presence of Parkinson’s disease in individuals using various factors can provide faster result without waiting for a longer time.

# **4. Question**

1. What is the scope of the disease and how many of them in total?
2. Which trusted sources are willing to cooperate in providing the quality dataset?
3. How many experts needed in order to make an early classification of the disease?
4. How long this system will be relevance due to the rapid growth of technology?

# **5. Success**

The system can be finalized as success when it performs a very good diagnosis in term of detection Parkinson’s disease as well as generating a proper reporting.

# **6. Measurement**

The main measurement of this system is the Accuracy. The higher the accuracy we get the better the result will been conclude. The accuracy is important because it will affect our analyzing of Parkinson disease.

# **7. Data Sources**

Our data will be has 24 columns and 195 records and is only 39.7 KB. Usually our data can get from the UCI ML or we can get the data from the hospital but need the approval from the staff before we provide the data.