**Software Requirement Specification**

**of**

**Parkinson’s disease classification**

**Introduction:**

Parkinson’s Disease is one of the most wide spread diseases in elderly people. This disease largely limits the patient’s movement and speech abilities. The patient develops a tendency to fall frequently hence, ending up hurt with various injuries. Thus, it is very important to monitor and notify either the patients or their caregivers about the severity of the disease. This work showcases a comparative study of the various datasets, algorithms and techniques available for the classification of Parkinson’s Disease. This project aims to apply multiple algorithms for classification on UCI Spiral dataset for Parkinson’s Disease.

The data set of Parkinson’s patient is available from the UCI repository and classification algorithms are applied for classifying the healthy patients from the people having Parkinson’s disease using a set of voice recording values as the attributes of the data set.

**Existing System:**

The existing system uses Support Vector Machine (SVM) one of the main issues with this is that it need the data to be linearly separable. The system also does not provide enough preprocessing and visualization or Exploratory Data Analysis(EDA).

**Disadvantages of Existing System:**

The limitations of available systems are not sufficient to deal with the complex data. In this section, we present some of the limitations that are present in the existing system.

* The model suffers from over fitting due to no generalization of data.
* The error on test data is high due to over fitting.
* The system also requires data extensive data preprocessing and Exploratory Data Analysis(EDA) in order to perform feature engineering.

**Proposed System:**

We aim to build other classification models like logistic regression, Naïve Bayes, Decision Trees and others and also fine tune the parameters of the model. These models would be trained on a data set which will be engineered carefully after performing the feature engineering.

**Advantages:**

* The requirement is to come up with novel features based on the functional understanding of the dataset. It is important to keep in mind to avoid correlated features during this process. Each feature should only improve the information contained in the dataset.
* visualize the dataset without cleaning the data and understand the distribution of the dataset.
* Find the attributes that needs to be handled based on the data visualization task.

**Software requirements:**

Operating System : Windows 7 , Windows 8, (or higher versions)

Language : Python 3.5 and other libraries likes numpy, pandas, matplotlib, seaborn and scikitlearn.

Mozilla Firefox(or any browser)

**Hardware requirements:**

Processor : Pentium 3,Pentium 4 and higher

RAM : 2GB/4GB RAM and higher

Hard disk : 40GB and higher

