SYNOPSIS

Project Team No: 23SOCU1050

Register No:	**Name:**
1. 124156079	Sankaranarayanan S
2. 124156080	Upendar S
3. 124157042	Peshwar Rajesh

Project Title: Early detection of Parkinson's disease using machine learning

Name of the Guide: JEYASHEELA RAKKINI M J

Abstract

First step of our project is to collect the required Parkinson's data from UCI. The data contains 23(columns) attributes of audio and 195(rows) patient data. The imported data is cleaned and pre-processing techniques have been applied. After performing exploratory data analysis, we discovered that as the disease advances in stages, speech noise increases, causing NHR to rise. The skew in the data and a low value of NHR imply poor voice quality, and NHR data points for PWP had the highest number of outliers because of the disease's increasing speech noise. Similar to PWP records, HNR records feature maximum data outliers below the median. After this, the data set is taken for model training where K- nearest – neighbours, Support Vector Machine, Random Forest, Logistic Regression techniques were used. For modal evaluation metrics we use Precision, Recall, Accuracy, ROC AUC Score. This is the basic implementation i.e., algorithm one. For the second algorithm we consider PCA (principal component analysis) approach where five number of components were chosen which identifies five most contributing attributes for our models. Those five features are: MDVP: Flo (Hz), DFA, D2, MDVP: Fo (Hz), MDVP: Shimmer. Similar model training on the Scaled data is trained and similar evaluation metrics have been used. For the Third algorithm we consider balancing the data by up sampling the minority class. Similar model training on the up sampled data is trained and similar evaluation metrics have been used. In all our cases no of neighbours in KNN is defaulted to five. Our findings indicate that random forest approach on basic implementation gives 91.83% accuracy and 1.00 recall, whereas KNN on PCA approach gives us 93.8% accuracy with 0.92 precision and 1.00 recall, Also KNN on Balanced approach gives us 92.3% accuracy with 0.93 precision and 0.96 recall.

Specific Contribution: Coding part

Specific Learning: KNN

Technical Limitations & Ethical Challenges faced: Understanding the attributes present in the data

Keywords: *KNN, PCA, MDVP*

Sankaranarayanan S JEYASHEELA RAKKINI M J

Name & Signature of the Student Name & Signature of Guide

Date: 8/5/2023