

Plagiarism Scan Report



Characters:2793

Words:405

Sentences:16

Speak Time:
4 Min

Excluded URL

None

Content Checked for Plagiarism

Methodology Dataset We have collected the required Parkinson's disease dataset, created by the University of Oxford, called the Oxford Parkinson's Disease Detection dataset, containing 197 recordings and 23 features, created by Max Little from the University of Oxford, along with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals. The data contains biomedical voice recordings from 31 individuals, 23 of whom had Parkinson's disease (PD). Each column in the table is a voice measurement, and each row corresponds to one of the 195 voices recorded by that individual ("name row"). For patients with Parkinson's disease, the "Status" column is set to 0 for healthy patients and 1 for Parkinson's disease. The file is in ASCII CSV format. The rows of the CSV file contain examples for a recording. Data Analysis The data analysis process aims to provide an understanding of the data set showing important relationships, distributions and features. This knowledge will guide preprocessing steps and contribute to the development of effective machine learning models for diagnosing Parkinson's disease. Distribution Analysis Analyzing the distribution of target variable values (1 and 0) to understand the prevalence of patients with Parkinson's disease in the dataset and identify abnormalities that will affect model training and evaluation. Correlation Analysis: Analyzing the correlation between the different features and the target variable values (status 1 and 0) and finding out features that have a very strong correlation, because they can play an important role in the diagnostic process. Correlation Heatmap of features Proposed architecture The flowchart in Fig. 3. represents a pipeline for a machine learning approach on multi-classification models that can predict Parkinson's disease. Model Training: Several machine learning models are implemented within the pipeline. They are: LR (Logistic Regression) DT (Decision Tree) RF (Random Forest) SVM (Support Vector Machine) KNN (K-Nearest Neighbors) GNB (Gaussian Naive Bayes) BNB (Bernoulli Naive Bayes) VC (Voting Classifier) All these models are fitted with the training dataset, and each model attempts to predict whether the subject is diagnosed with the disease of Parkinson's. Splitting the dataset: The data is split into 80% training and 20% testing data as our applied ML algorithms first train themselves based on the data of the training dataset whereas the remaining testing dataset is used to text and predict the outcome. Performance Measure To compare the prediction performance of the algorithms, 4 metrics are used, namely precision, accuracy, recall, and f1 score respectively.

Sources



[Home](#)

[Blog](#)

[Testimonials](#)

[About Us](#)

[Privacy Policy](#)

Copyright © 2024 [Plagiarism Detector](#). All right reserved