Detecting Parkinson's disease using Data-Driven Classification Model

Author: Peter Osewingie [1807381]
Institution: University Of Essex
Unit: CE 888 Data Science and Decision Making
Date: 4/02/2021

Abstract This study anchored on predicting Parkinson's disease whilst utilizing voice Dataset which assists with treating the individuals in early phases. Parkinson's illness is a neurological issue that prompts shaking and difficulty in walking, balance, and coordination. In most pessimistic scenarios, Patients have a big challenge walking or standing to an extent they can't live on their own and require a wheelchair to move around whilst help is required in all every day exercises. Other than motor indications, the individual may see, hear, or experience things that are not genuine (hallucinations), or accept things that are false (delusions). Parkinson's disease patients commonly have a low-volume voice with a droning quality. The speech pattern of Parkinson's patient is frequently delivered in short overflows with unseemly hushes among words and long stops prior to starting discourse. The voice dataset have the variables like MDVP: FO (Hz) - Average vocal fundamental frequency, MDVP: Fhi (Hz) - Maximum vocal fundamental frequency, jitter, simmer and so forth. The dataset was split into train and test where the train dataset was utilized to prepare the model. The test dataset was utilized to test the XGB model which delivered a higher precision of 100 percent.

1 Introduction

Parkinson's disease (PD) is an ailment that influences the nerve cells in the cerebrum that make dopamine gives indications including muscle immovability, tremors, and changes in sync and talk. Parkinson's infection impacts a man's voice, aKecting them to murmur or have shuddering in talk. PD is simply second to alzheimer's disease in neurodegenerative sickness. It is relied upon to increment in the coming years subsequently it is important to create identification frameworks for down to earth examination and ideal treatment. As the indications of PD happen all around requested and generally, the older noticing the disorder utilizing assessments of dysphonia has an indispensable part in examination. The characterization calculations from AI and machine learning are utilized to predict and explore the Parkinson's disease. The ideal highlights from the dataset are passed as contribution to the models and the expectation results are goNen. The expectation execution can be approved from the precision acquired through the classification algorithm. The assurance of Parkinson's ailment has logically improved the exactness boundary through the diKerent analysis. This paper will therefore explore the eKectiveness of using supervised classification algorithms, such as XGBOOST classifier, to accurately diagnose individuals with the disease.

2 Background

The clinical diagnosis of Parkinson diseases (PD) can be aKirmed based on neuro-pathologic and histopathologic standards [1]. Clinical indicative grouping of PD should be possible on far reaching survey of the writing information and determination basing on the aKectability and particularity of the trademark clinical features. Imminent with center pathologic examinations in agent populace of patients demonstrating PD are expected to explore the clinical, pathologic, and nosologic studies dependent on recurrence of event, attributes, and danger factors in patients [2]. Neural Networks, DMneural, Regression and Decision Trees are recently utilized for ascertaining the presentation score of the classifiers dependable analysis of PD [3, 4]. R. Arefi Shirvan.et al [4] proposed a framework/system for detecting early stages of Parkinson's diseases. The data classification was completed by utilizing knn method. The least complex technique in gathering the comparability is knn. Among classification strategy knn is utilized at whatever point current realities for information dissemination are insuKicient [5]. In this technique it has two sections: a) decide k close neighbors, b) deciding class type utilizing these nearby neighbors. it was demonstrated that a 93.7 percent of precision for each 4 upgraded highlights, an exactness of 94.8percent per 7 streamlined highlights and 98.2 percent precision for 9 advanced highlights is accomplished which is a mo-

mentous outcome contrasted with diKerent investigations. In this study data from [6] from UCI archive is utilized. The information incorporate 192 voice test accounts from 32 male and female. Each subject has had 6 voice signal accounts. 23 individuals experience the ill eKects of PD and the rest are sound. individuals were around 46-85 years of age the primary inconvenience of the knn calculation is that it is a lethargic student, for example order is finished by utilizing preparing information and from the training data it did not get to learn anything from it. Mohammad s islam.et al [7] directed a comparative analysis to detect Parkinson's infection utilizing diKerent classifiers. Support vector machine (SVM), feed forward back-spread based artificial neural organization (fbann) and arbitrary tree (rt) classifiers were utilized and an examination between them is made to separate among PD and healthy patients. The study has utilized the UCI machine learning repository from [8],[9]. The dataset comprises of 195 voice tests from 31 people involving the two males and females. From the taken subjects 23 were resolved with PD and 8 were healthy. To improve the grouping precision with insignificant error rate a 10-fold cross approval which was repeated multiple times (100) has been executed for all the three classifiers. The knn classifier has accomplished a 97.37 percent acknowledgment exactness consequently outflanking the other two classifiers. Ddr. r. geetharamani.et al [10] has suggested a framework to order PD and Non-PD patients by the following classifiers; binary logistic regression, linear discriminant analysis (lda), partial least square regression (pls) , random tree (rnd tree) and support vector machine (svm). The Parkinson's disease dataset is retrieved from the UCI Repository. This data is exctracted from patients and comprises of 197 unique samples and 22 features. Fisher separating feature choice calculation was discovered to be a viable element positioning framework. The random tree calculation accomplished 100 percent arrangement precision while the lda, c4.5, cs-mc4 and k-nn yielded exactness results more noteworthy than 90 percent. Among all, the c-pls calculation accomplished minimal precision of 69.74 percent. The multilayer perceptron (mlp) with back-engendering learning algorithm.

Name - ASCII subject name and recording number
MDVP: Fo(Hz): Average vocal fundamental frequency
MDVP: Fhi (Hz): Maximum vocal fundamental frequency
MDVP: Flo(Hz): Minimum vocal fundamental frequency
MDVP:JiNer(percent),MDVP:JiNer(Abs), MDVP:RAP,MDVP:PPQ,JiNer:DDP– Several measures of variation in fundamental frequency.
MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3, Shimmer:APQ5,MDVP:APQ,Shimmer:DDA – Several measures of variation in amplitude
NHR, HNR - Two measures of ratio of noise to tonal components in the voice
Status - Health status of the subject (one) - Parkinson's, (zero) - healthy
RPDE, D2 - Two nonlinear dynamical complexity measures

DFA - Signal fractal scaling exponent
spread1, spread2, PPE - Three nonlinear measures of fundamental frequency variation

Data Preprocessing
This section entails two cycle which is Normalization and adjusting the dataset and is given in finer detail below:

Normalization
Normalization is a procedure which is applied as a stage of preparing a dataset for machine learning model. The need of normalization is to adjust the estimations of numeric sections in the dataset to a typical scale, without changing contrasts in the scopes of values.

Xnew= (X- Xmin)/(Xmax- Xmin)
Where Xnew specific component spoke to by a segment in the dataset, x is a value of this column. The minimum value of the column is represented as Xmin and the maximum value of the column is Xmax. However for this study I used the MinMaxScaler function from sklearn to normalize the features.

Modeling and Analysis
XGBOOST: XGBoost is a gradient boosting library.

assemble the proficient model for early detection of Parkinson's illness:

3 Methodology

The following are the means that has been taken to

It helps to implements machine learning algorithms under the Gradient Boosting framework. XGBoost a parallel tree boosting which solves many Machine Learning problems in a fast and simple way. The jupyter notebook code aNached demonstrates how XGB classifier has solved this machine learning problem.

Dataset Detail

The dataset used to build the model for this study is retrieved is collected from UCI website. This dataset has 195 unique values and 24 columns. Matrix column entries (aNributes):

4 Results and Discussion

This machine learning project analysis has utilized various factors/variables to detect the presence of Parkinson's disease. The XGBClassifier was used for the classification and made use of the sklearn library to prepare

the dataset. This XGBClassifier model produced an accuracy of 100 percent, which is great considering the number of lines of code and the size of the dataset in this python project.

5 Conclusion

In conclusion, this study has leveraged on the use of the XGBoost classifier which gives eKicient Parkinson's disease prediction model with high accuracy 100 percent. This will go a long way in assisting detecting and predicting Parkinson disease prior to geNing it to most exceedingly awful. Analysis of voice data is significant in the current decade to comprehend and indicative techniques for human infections. The current technique gives the finding of PD utilizing voice dataset through machine learning algorithms. Early recognition of Parkinson's illnesses is valuable as it will assists with keeping the patients from most noticeably terrible stage

6 References

[1] D.J. Gelb, E. Oliver, and S. Gilman, "Diagnostic criteria for Parkinson disease." Archives of neurology, vol. 56, no.1, pp. 33 1999.

[2] D. Aarsland, K. Andersen, J.P. Larsen, and A. Lolk, "Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study." Archives of Neurology. Vol. 60, no. 3, pp. 387, 2003.

[3] R. Das, "A Comparison of multiple classification methods for diagnosis of Parkinson disease." Expert Systems with Applications, vol. 37, no. 2, pp. 1568-1572, 2010.

[4] Deepak Gupta, Optimized cuNlefish algorithm for diagnosis of Parkinsons disease, Cognitive Systems Research, Volume 52, December 2018, Pages 36-48 hNps://doi.org/10.1016/j.cogsys.2018.06.006.

[5] Mathur, R., Pathak, V., Bandil, D. (2019). Parkinson Disease Prediction Using Machine Learning Algorithm. InEmerging Trends in Expert Applications and Security (pp.357-363). Springer, Singapore.

[6] Benba, A., Jilbab, A., Hammouch, A., Sandabad, S.(2015, March). Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. In2015 International conference on electrical and information technologies (ICEIT) (pp. 300-304). IEEE.

[7] M. A. LiNle, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Trans. Biomed. Eng., vol. 56, no. 4, pp.1010–1022, 2009.

ease patients. In Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India (p. 14).ACM.

[9] Leandro A. Passos, Parkinson Disease Identification using Residual Networks and Optimum-Path Forest, SACI 2018, IEEE 12th International Symposium on Applied Computational Intelligence and Informatics, May 17-19, TimiÃºoara, Romania.

[10] R. Polikar, A. Topalis, D. Green, J. Kounios, and C. M. Clark, "Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease." Computers in biology and medicine, vol. 37, no. 4, pp. 542-558, 2007.

[8] BhaNacharya, I., Bhatia, M. P. S. (2010, September).SVM classification to distinguish Parkinson dis-