# Architecture

**Parkonson’s disease prediction**

**Version 1.0**

**Date of Revision 15 April 2022**

# Contents

Document Version Control 1. Introduction

1.1 What is an architecture design document?

1.2 Scope

1.3 Constraints 2. Technical specifications

2.1 Dataset

2.1.1 Dataset overview

2.1.2 Input schema

2.2 User Input Requirements

2.3 Logging

2.4 Deployment 3. Technology Stack

1. Proposed Solution
2. Data pre-processing workflow 6. Admin data setup workflow
3. Machine learning training workflow
4. Machine learning recommendation workflow 9. Test cases
5. Performance

**Document Version Control**

**Date Issued Version Description Author**

08-05-2022

1 Initial Architecture – V1.0

sachchit s kolekar

## Introduction

### What is an architecture design document?

The architecture design document is a technical document describing the components and specifications required to support the solution and ensure that the specific business and technical requirements of the design are satisfied.The architecture design document is a technical document describing the components and specifications required to support the solution and ensure that the specific business and technical requirements of the design are satisfied.

### Scope

This Parkinson’s Disease Prediction System web application will recommend books based on the user's interested book.

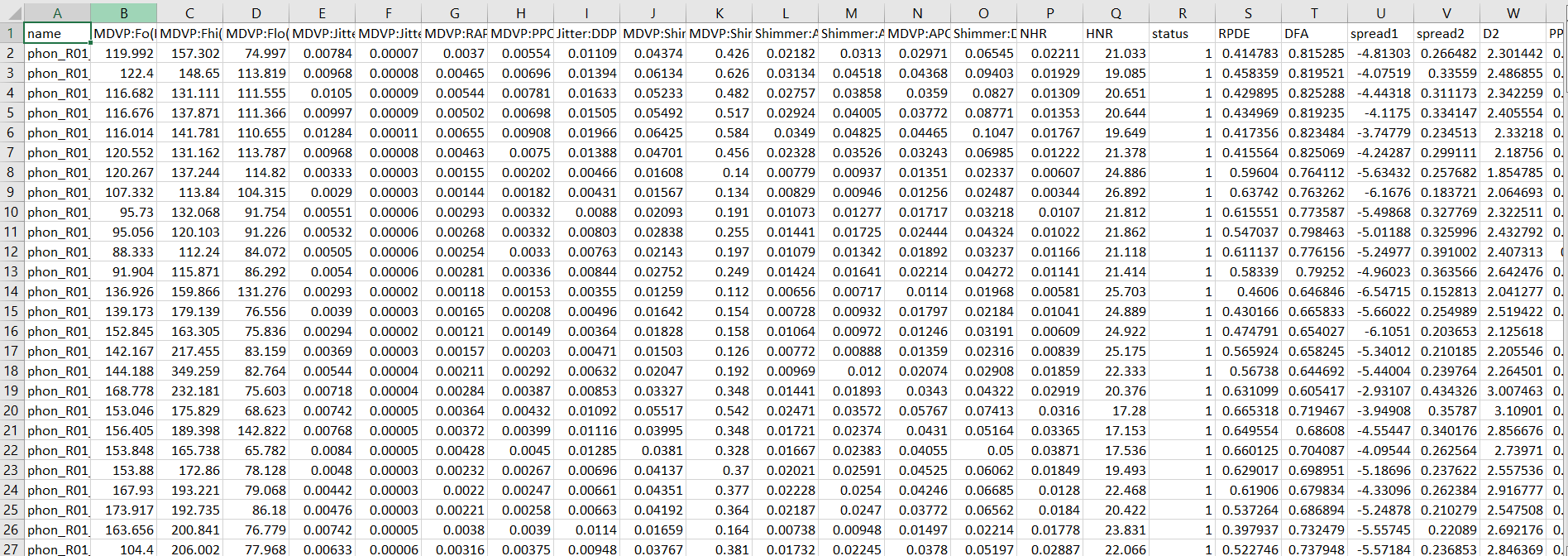
### Constraints

The Parkinson’s Disease Prediction System application must be user friendly, as automated as possible and users should not be required to know any of the workings.

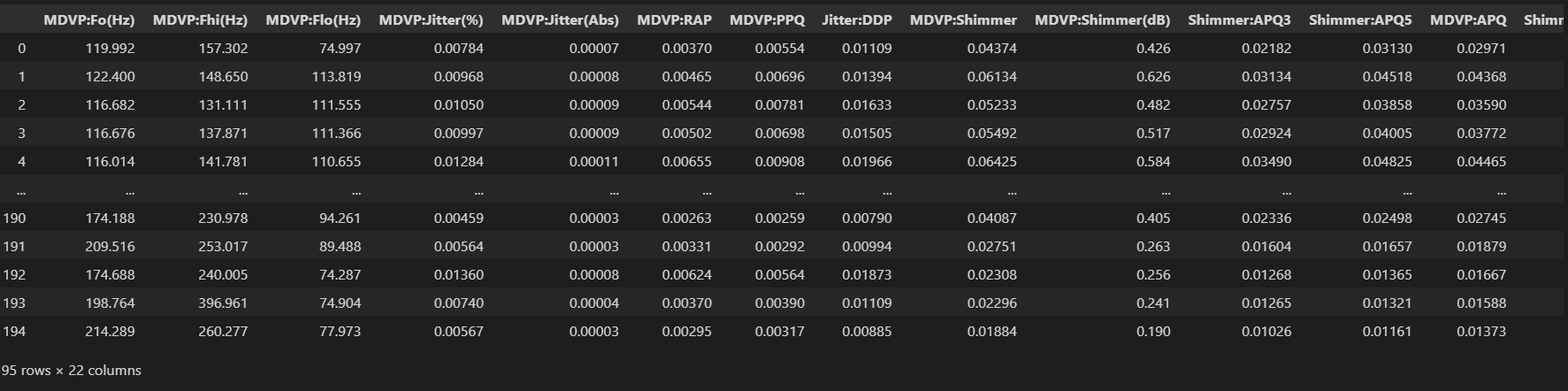
## Technical specifications

### Dataset

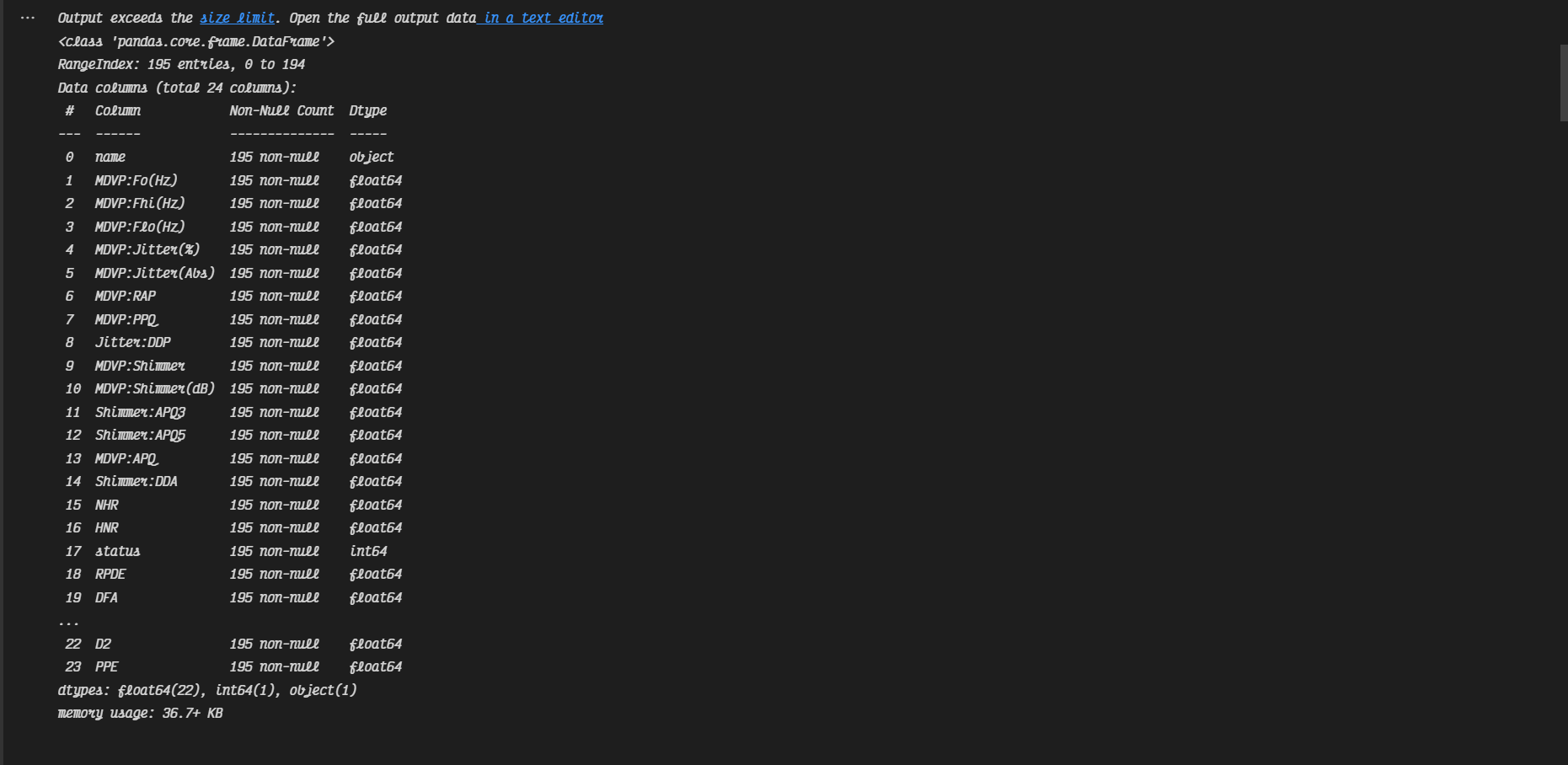
**Diagnosis table:**



**Transcrip table:**

****

**Description table:**

****

* + - 1. **Dataset overview**

**TMDVP:Fo(Hz) - Average vocal fundamental frequency  
MDVP:Fhi(Hz) - Maximum vocal fundamental frequency  
MDVP:Flo(Hz) - Minimum vocal fundamental frequency  
MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter: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**

### Input schema

|  |  |  |
| --- | --- | --- |
| **Feature name** | **Null / Required** | **data type** |
| MDVP:Fo(Hz) | Required | Int |
| Jitter:DDP |  | Int |
| NHR | Required | Int |
| Spread1 | Required | int |
| PPE |  | Int |

* + 1. **User Input Requirements**

Users must give input ,and the disease will be predicted.

### Logging

We should be able to log every activity done by the user.

* The System identifies at what step logging required
* The System should be able to log each and every system flow.
* Developers can choose logging methods. You can choose database logging/ File logging as well.
* System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.
  + 1. **Deployment**



## Technology Stack

|  |  |
| --- | --- |
| **Frontend** | Html, CSS, Javascript, JQuery |
| **Backend** | Python, Flask |
| **Database** | Cassandra |
| **Machine Learning** | SK Learn |
| **Deployment** | Heroku |

* 1. **Proposed Solution**

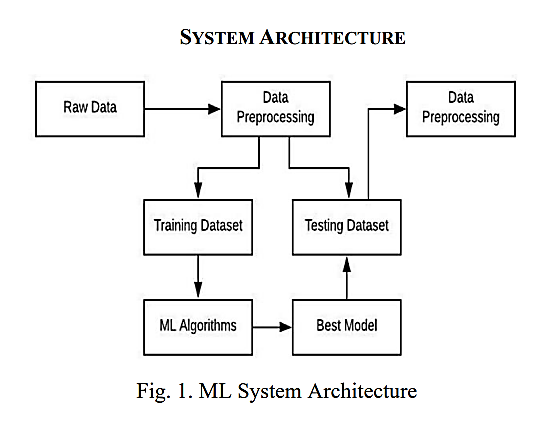
In multi disease model prediction, it is possible to predict more than one disease at a time. So user no need to traverse many models to predict the diseases. It will reduce time and also due to predicting multiple diseases at a time there is a chance of reducing mortality rate.

## Data pre-processing workflow

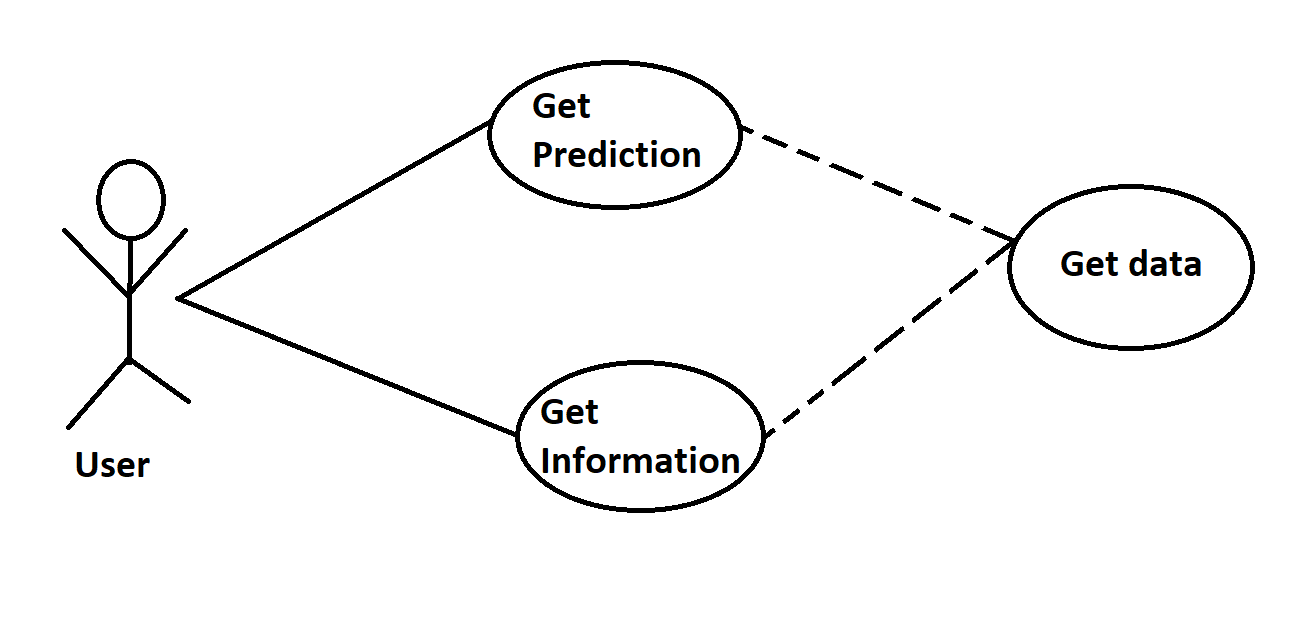
In data pre-processing the following steps are taken.

* **Create tags:** by combining feature columns we are creating tags.
* **Apply Stemming:** in tags we are applying stemming so the model will learn reliably. Ex: applying stepping on `dancing`,`danced` will become `danc`
* **Remove Stop Words:** in tags we are removing stop words like `the`, `and` etc… so, learning accuracy will increase.

## **Model training/validation workflow**

****

* 1. **User I/O workflow**

****

* 1. **Test cases**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr No** | **User Input** | **Output** | **Test Result** |
|  |  |  |  |
| 1 | Statistical Learning Theory | 1. Nature of Statistical Learning Theory 2. Machine Learning for Hackers 3. Statistical Decision Theory' 4. Pattern Classification 5. Neural Networks | PASS |

## Work-flow implementation

## 

## 10.Key performance indicators (KPI)

Baseline Model: Logistic Regression, since this is a classification problem.

Actual model: LSTMs.

● Number of times a patient visits the hospital.

● Time between symptom onset and detection of illness/visit to hospital.

● Immunity of patient varies.

● Vaccines the patient .

● prediction depicts the model credibility.