Prediction of Diseases using Different Machine Learning Approaches

Abstract:

The medical system is overworked and becoming more and more expensive in many countries each year as the number of patients and diseases rises. In order to be treated, the bulk of the ailment requires a visit to the doctor. A simple and affordable algorithm can predict diseases if there is enough data. A key part of treatment is the ability to diagnose disease based on symptoms. Using the patient's symptoms, we sought to accurately predict a disease in our project. We are going to use four different algorithms to accomplish this to obtain better accuracy. The use of such a technology in medical care in the future is quite promising. In order to facilitate interaction with the data, we are also planning to develop an interactive interface.

Dataset:

The dataset is an open-source one hosted in Kaggle Data Repository.
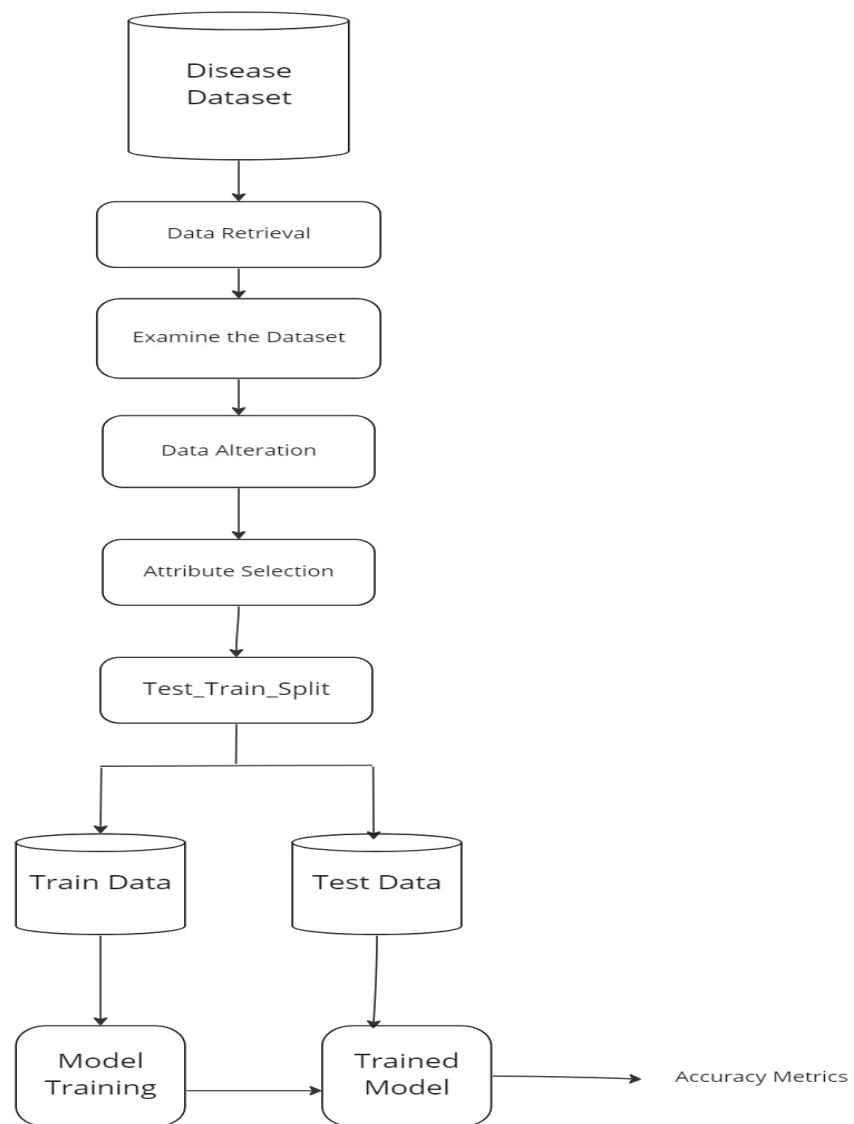
Total Rows : 4962   Total Diseases : 42

Fungal infection
Allergy
GERD
Chronic cholestasis
Drug Reaction
Peptic ulcer diseae
AIDS
Diabetes
Gastroenteritis
Bronchial Asthma
Hypertension
Migraine
Cervical spondylosis
Paralysis (brain hemorrhage)
Jaundice
Malaria
Chicken pox
Dengue
Typhoid
hepatitis A
Hepatitis B

Hepatitis C
Hepatitis D
Hepatitis E
Alcoholic hepatitis
Tuberculosis
Common Cold
Pneumonia
Dimorphic hemmorhoids(piles)
Heart attack
Varicose veins
Hypothyroidism
Hyperthyroidism
Hypoglycemia
Osteoarthristis
Arthritis
(vertigo) Paroymsal Positional Vertigo
Acne
Urinary tract infection
Psoriasis
Impetigo

Architectures planned to be used:

- Decision Tree
- Random Forest
- K Nearest Neighbours
- Naïve Bayes Algorithm

Flow of Project:

```
                    ┌─────────────┐
                    │   Disease   │
                    │   Dataset   │
                    └─────────────┘
                           │
                    ┌─────────────┐
                    │Data Retrieval│
                    └─────────────┘
                           │
                    ┌──────────────────┐
                    │Examine the Dataset│
                    └──────────────────┘
                           │
                    ┌─────────────┐
                    │Data Alteration│
                    └─────────────┘
                           │
                    ┌──────────────────┐
                    │Attribute Selection│
                    └──────────────────┘
                           │
                    ┌──────────────────┐
                    │ Test_Train_Split │
                    └──────────────────┘
                      │            │
              ┌────────────┐   ┌────────────┐
              │ Train Data │   │ Test Data  │
              └────────────┘   └────────────┘
                    │                │
              ┌────────────┐   ┌────────────┐
              │   Model    │──▶│  Trained   │──▶ Accuracy Metrics
              │  Training  │   │   Model    │
              └────────────┘   └────────────┘
```

miro