**ABSTRACT**

Now-a-days, people face various diseases due to the environmental condition and their living habits. So the prediction of disease at earlier stage becomes important task. But the accurate prediction on the basis of symptoms becomes too difficult for doctor. The correct prediction of disease is the most challenging task. To overcome this problem data mining plays an important role to predict the disease. Medical science has large amount of data growth per year. Due to increase amount of data growth in medical and healthcare field the accurate analysis on medical data which has been benefits from early patient care. With the help of disease data, data mining finds hidden pattern information in the huge amount of medical data. We proposed general disease prediction based on symptoms of the patient. We have prepared models for 4 diseases currently which are calculated on the basis of medical record of hundreds of patients. The models are based on SVM and Logistic regression

algorithms.

The past existing AI models for medical care examination are centered around one sickness for each

investigation. Like one examination if for diabetes investigation, one for malignancy examination, one

for skin infections like that. There is no regular framework where one investigation can perform more

than one infection expectation. In our proposed system, we unify multiple diseases under a single user

interface where you can perform predictions on Heart diseases, breast cancer, and diabetes. In this

work, we are using the machine learning classification algorithms like LogisticRegression, Support

Vector Machine (SVM), K-Nearest-Neighbors (KNN) to perform the prediction of multiple diseases.

## **LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
| **Term** | **Meaning** |
| ML | Machine Learning |
| SVM | Support Vector Machine |
| KNN | K Nearest Neighbour |
| LR | Logistic regression |

**LIST OF FIGURES**

* + - Fig. 4.1 - Diabetes prediction
    - Fig 4.2 - Diabetes prediction
    - Fig 4.3 - Parkinsons prediction

**Table of Contents**

* Abstract
* List of Abbreviations
* List of Figures

|  |  |
| --- | --- |
| **Chapter** | **Title** |
| I | Introduction |
| II | Literature Review |
| III | Problem Identification |
| IV | Methodology |

|  |
| --- |
|  |

**Chapter - I**

**INTRODUCTION**

Machine Learning is the domain that uses past data for predicting. Machine Learning is the understanding of

computer system under which the Machine Learning model learn from data and experience. The

machine learning algorithm has two phases:

1) Training 2) Testing.

To predict the disease from a patient’s symptoms and from the history of the patient, machine learning technology is struggling from past decades. Healthcare issues can be solved efficiently by using Machine Learning Technology. We are applying complete machine learning concepts to keep the track of patient’s health.

ML model allows us to build models to get quickly cleaned and processed data and deliver results faster. By

using this system doctors will make good decisions related to patient diagnoses and according to that, good

treatment will be given to the patient, which increases improvement in patient healthcare services. To

introduce machine learning in the medical field, healthcare is the prime example. To improve the accuracy of

large data, the existing work will be done on unstructured or textual data.

Diabetes, and heart disease are for the most part driving reasons for death in the present society. Heart disease is a term that refers to a group of illnesses that affect your heart. Arrhythmias (issues with heart rhythm), coronary artery disease, and congenital heart defects are all diseases that fall under the category of heart illness (the defects of the heart you are born with).The term "heart disease" is often used instead of the term "cardiovascular disease."The cardiovascular disease normally indicates heart attack, angina (heart pain), or stroke, also conditions that affect your rhythm valves or muscles of your heart also referred to as heart diseases.An article published in “Science Direct” states that in India the annual deaths are nearly 11 million, of which 28 percent of deaths are due to cardiovascular

disease.

**Chapter - II**

**Literature Review**

This section describes the study of previously proposed models for predicting the diseases which are

related to our proposed work. Several studies have been made for detecting various diseases. They have

applied various data mining techniques for efficiently predicting a variety of diseases.[7]G Naveen

Kishore and few other authors proposed the work named Prediction Of Diabetes Using Machine

Learning Classification Techniques proposed. In this work, various classification algorithms like SVM,

Logistic Regression, Decision Tree, KNN, Random Forest are utilized on the 769 instances of the Pima

dataset which contain features like Pregnancies, Blood pressure, body mass index, etc. They have

reported the highest accuracy as 74.4 %for the classification algorithm Random Forest and the lowest

accuracy in this work is attained by the KNN reported as 71.3%. [8]The work “Understanding the

lifestyle of people to identify the reasons for Diabetes using data mining” proposed by Gavin Pinto,

Radhika Desai, and Sunil Jangid discussed reducing the risk of diabetes disease using data mining

techniques and also discussed diabetes sub-classification. The authors used Naïve Bayes and SVM

classification algorithms on the dataset collected by a survey using google forms and reported the

accuracy of 64.92 for SVM and 60.44 for Naïve Bayes.

[9]In the work presented by M.Marimuthu, S.DeivaRani, Gayatri. R described the cardio diseases in a

detailed manner and also applied the classification algorithms like SVM, Decision Tree, Naïve Bayes,

K-Nearest Neighbors on the Framingham dataset from Kaggle. The authors compared various machine

learning algorithms for the forecast of the risk of heart disease. The highest reported accuracy in this

work is 83.60% for the KNN classification algorithm. [10]In the work proposed by Purushottam, Richa

Sharma and Dr. Kanak Saxena discuss cardiovascular sickness by using the implementation of

Knowledge Extraction based on Evolutionary Learning (java programming technique for making the

development model for data mining issues). The highest reported accuracy in this work is 86.7%.

M. Chinna Rao, K. Ramesh, and G. Subbalakshmi presented a decision support system for

heart disease prediction utilising the Nave Bayes classification method, which discussed the extraction

of hidden information heart disease dataset that can address complex queries.Amandeep Kaur and

Jyothi Arora presented a study that covered the examination of algorithms such as KNN, SVM, ANN,

and Decision Tree on the heart disease dataset and plotted the accuracies graph.[14]Noreen Fatima

proposed work on the Cancer forecast the data mining techniques and machine learning techniques that

can predict cancer effectively on the large health records and described the study previous existing

models.

Ch. Shravya, K.Pravallika, Shaik Subhani presented the work on Breast cancer prediction using

Supervised machine learning techniques on the dataset and also analyzed the results with

(PCA)principal component analysis and also used the dimensionality reduction and explained in a wellmannered way.Nikitha Rane, Jean Sunny presented work on the classification of Cancer using

machine learning concepts and their major discussion point is detecting cancer in very early stages so

that a lot of lives can be saved.Dilip Singh Sisodia ,Deepti Sisodia predicted diabetes using

classification techniques and reported an accuracy of around 76% on the Pima dataset.Dr. J.Ajayan, Dr.B.Santhosh Kumar ,T.Daniya have predicted the occurrence of Breast cancer using KNN

algorithm with accuracy value of 83.33%.Mümine KAYA KELEŞ predicted cancer usingRandom

Forest algorithms and reported an accuracy value of 92.20%.

**Chapter - III**

**Problem Identification**

Many people don’t go to doctors as they think it is a small problem but small problem can be a symptoms of a big harmful disease. We use this model to predict the diseases.

Predictive analytics in healthcare is a difficult endeavor, but it can eventually assist practitioners in making timely decisions regarding patients' health and treatment based on massive data.

Diseases like diabetes, and heart related diseases are causing many deaths globally but most of these deaths are due to the lack of timely check-ups of the diseases.

Therefore, early recognition and diagnosis of these diseases can save a lot of lives. This work is all about predicting diseases that are harmful using machine learning classification algorithms.

**Chapter – IV**

**Methodology**

Currently the model is trained for 4 diseases :-

* Diabetes
* Heart
* Parkinsons
* Migraine

We are looking forward to add more of the diseases

50% of our work is completed .

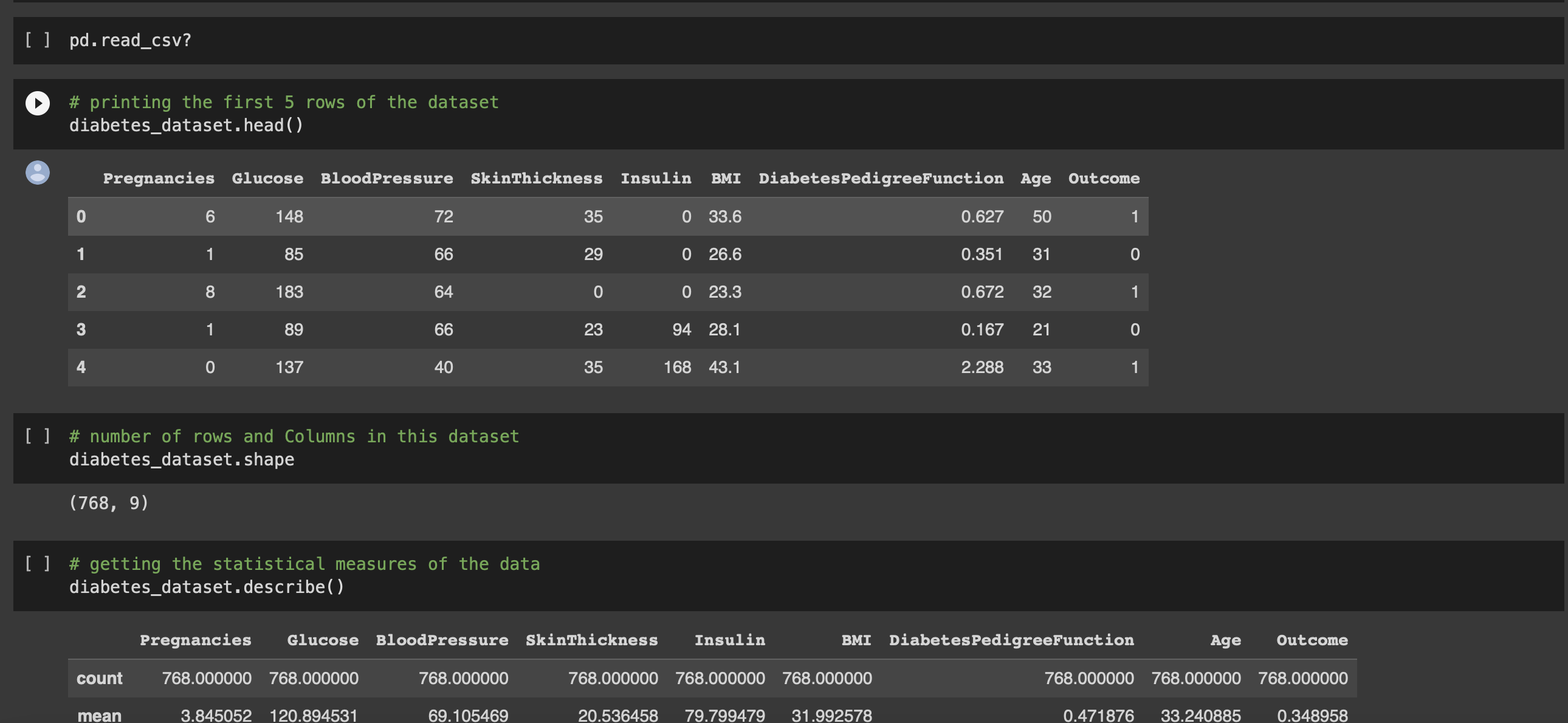


Image 4.1(Diabetes prediction)

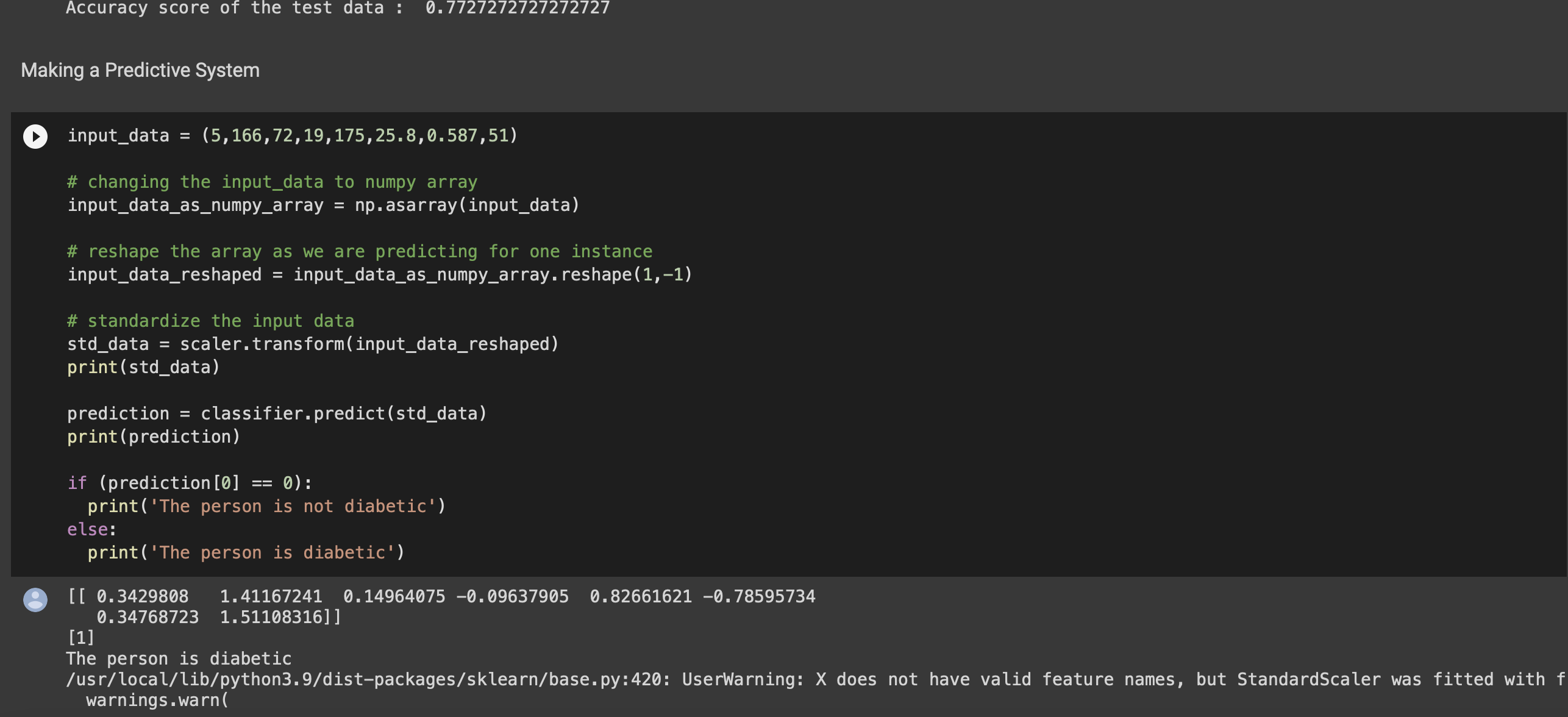
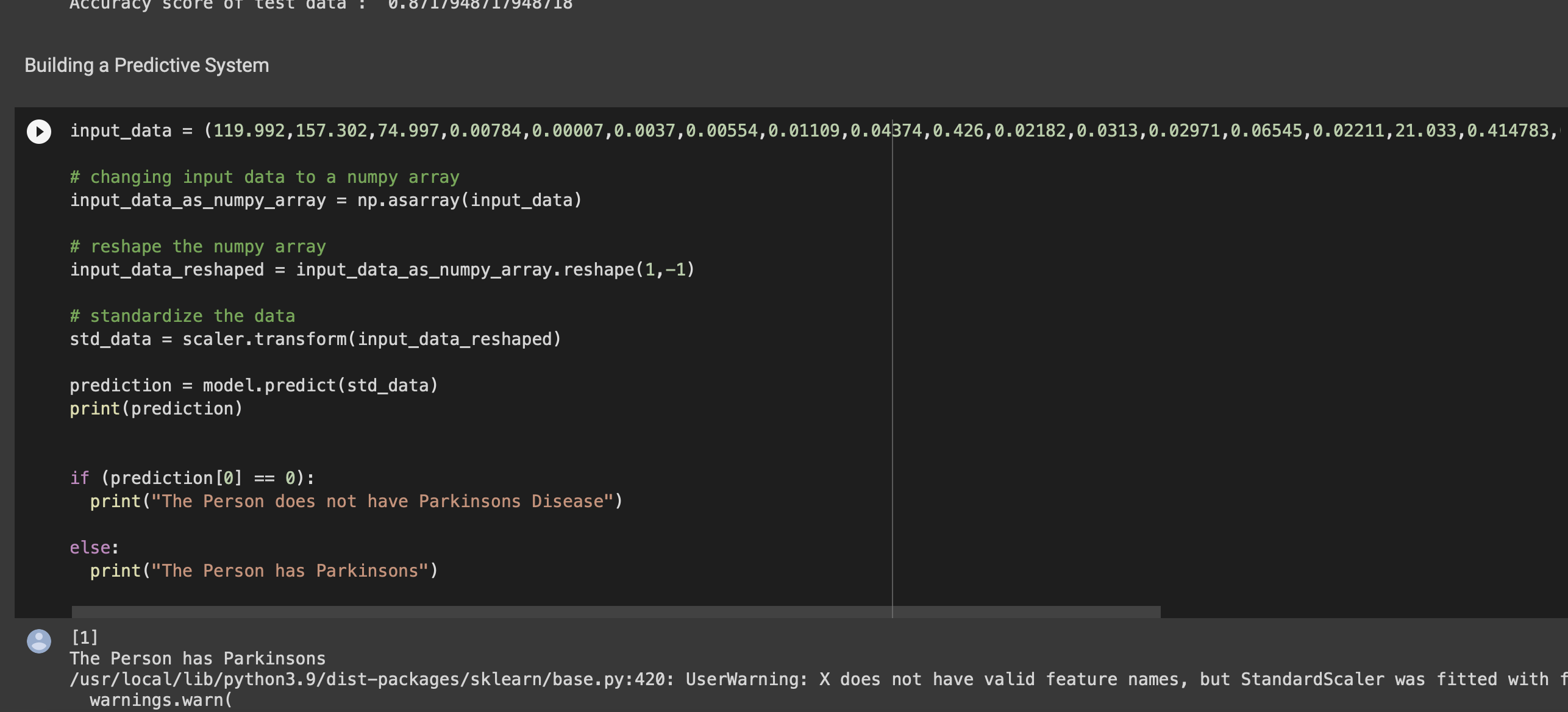


Image 4.3(parkinson prediction)

References :

* + 1. SEDGEWIK, R. Algorithms in C : Fundamentals, data structures, sorting, searching. Third Edition. 2007. ISBN 0-201-31452-5.
    2. KNUTH, D. The Art of Computer Programming: Fundamental Algorithms. Third Edition. 2004. ISBN 0-201-89683-4.
    3. KNUTH, D. The Art of Computer Programming: Sorting and Searching. Second Edition. 2004. ISBN 0-201-89685-0.
    4. GeeksforGeeks. Available from: hhttps://[www.geeksforgeeks.org/i.](http://www.geeksforgeeks.org/i)
    5. Stackoverflow. Available from: hhttps://stackoverflow.com/i.
    6. Java documentation. Available from: hhttps://docs.oracle.com/javase/8/i.