

Chapter 46

Machine Learning in Python: Diabetes Prediction Using Machine Learning


Astha Baranwal

VIT University, India

Bhagyashree R. Bagwe

VIT University, India

Vanitha M

 <https://orcid.org/0000-0001-7726-2621>

VIT University, India

ABSTRACT

Diabetes is a disease of the modern world. The modern lifestyle has led to unhealthy eating habits causing type 2 diabetes. Machine learning has gained a lot of popularity in the recent days. It has applications in various fields and has proven to be increasingly effective in the medical field. The purpose of this chapter is to predict the diabetes outcome of a person based on other factors or attributes. Various machine learning algorithms like logistic regression (LR), tuned and not tuned random forest (RF), and multilayer perceptron (MLP) have been used as classifiers for diabetes prediction. This chapter also presents a comparative study of these algorithms based on various performance metrics like accuracy, sensitivity, specificity, and F1 score.

INTRODUCTION

Diabetes is a disease which happens when the glucose level of the blood becomes high, which eventually leads to other health problems such as heart diseases, kidney disease etc. Several data mining projects have used algorithms to predict diabetes in a patient. Though, in most of these projects, nothing is mentioned about the dangers of diabetes in women post-pregnancies. While data mining has been successfully applied to various fields in human society, such as weather prognosis, market analysis, engineering

DOI: 10.4018/978-1-6684-6291-1.ch046

Machine Learning in Python

diagnosis, and customer relationship management, the application in disease prediction and medical data analysis still has room for improvement in accuracy.

Machine learning relates closely to Artificial Intelligence (AI) and makes software applications predict outcomes through statistical analysis. The algorithms used allow for reaching an optimal accuracy rate in predicting the output from the input data. Machine learning follows similar processes used in data mining and predictive modeling. They recognize patterns through the data entered and then adjust the actions of the program accordingly.

Machine learning algorithms are categorized as supervised learning and unsupervised learning. Supervised learning requires input data and the desired output data to build a training model. The training model is built by a data analyst or a data scientist. A feedback is then furnished concerning the accuracy of the model and other performance metrics during algorithm training. Revising is done as needed. Once the training phase is completed, the model can predict outcomes for new data. Classification is one of the many data mining tasks. Classification comes under supervised learning which implies that the machine learns through examples in Classification. In classification, every instance from the dataset is classified into a target value. Classification can either be binary or multi-label. Sometimes, one particular instance can also have multiple classes known as multi-class classification. Classification algorithms are majorly used for prediction and come under the category of predictive learning.

Unsupervised learning is used to draw inferences from the input data which do not have any labeled responses. This data is not categorized, labeled or classified into classes. Clustering analysis, one of the most common unsupervised learning method, is used to find hidden patterns in data or to form groups based on the input data.

While machine learning models have been around for decades, they have gained a new momentum with the rise of AI. Deep learning models are now used in most of the advanced AI applications. If these models are implemented for medical uses, they could be revolutionary for the society. Diagnosis of diseases like diabetes would be easier than ever. Machine learning in medical diagnosis applications fall under three classes: Pathology, Oncology and Chatbots. Pathology deals with the diagnosis of diseases with the help of machine learning models created with the data of diagnostic measurements of the patients. Oncology uses deep learning models to determine cancerous tissues in patients. Chatbots designed using AI and machine learning techniques can identify patterns in the symptoms of the patients and suggest a potential diagnosis or it can recommend further courses of action. This chapter falls under the pathological uses of machine learning as the model created will give diagnosis of whether a patient is diabetic or not.

This chapter focuses on implementing machine learning algorithms on the diabetes dataset in python. Python is a great language to support machine learning. It was created by Guido van Rossum. It is powerful, multipurpose, and simple and has an easy to use syntax. The length of a python code is generally relatively short. It is not overly strict. It is a fun language to work with because it lets us focus on the problem rather than the syntax. Python is a general purpose language with applications in a wide range of fields like web development, mathematical computing, graphical user interfaces, etc. These wide range of applications python most suitable for implementing machine learning algorithms. It is possible to implement some machine learning algorithms in python and later deploy the web service for it. Or it is also possible to create a graphical user interface for the deployed web service. This all becomes possible due to the diverse nature of the language. Various python libraries like, NumPy, SciPy, Matplotlib, Keras, Pandas, TensorFlow, and etc support machine learning in python. Scikit-learn library supports almost all the major machine learning classification, clustering and regression models. The Flask python web

25 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/machine-learning-in-python/307489

Related Content

Features Selection Study for Breast Cancer Diagnosis Using Thermographic Images, Genetic Algorithms, and Particle Swarm Optimization

Amanda Lays Rodrigues da Silva, Máira Araújo de Santana, Clarisse Lins de Lima, José Filipe Silva de Andrade, Thifany Ketuli Silva de Souza, Maria Beatriz Jacinto de Almeida, Washington Wagner Azevedo da Silva, Rita de Cássia Fernandes de Lima and Wellington Pinheiro dos Santos (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-18).

www.irma-international.org/article/features-selection-study-for-breast-cancer-diagnosis-using-thermographic-images-genetic-algorithms-and-particle-swarm-optimization/277431

Multi-Objective Optimization of EDM Process on AISI P-20 Tool Steel Using Multi-Criteria Decision-Making Technique

Souvick Chakraborty (2021). *Machine Learning Applications in Non-Conventional Machining Processes* (pp. 33-44).

www.irma-international.org/chapter/multi-objective-optimization-of-edm-process-on-aisi-p-20-tool-steel-using-multi-criteria-decision-making-technique/271496

Comparative Analysis of Various Soft Computing Technique-Based Automatic Licence Plate Recognition Systems

Nitin Sharma, Pawan Kumar Dahiya and B. R. Marwah (2021). *Handbook of Research on Machine Learning Techniques for Pattern Recognition and Information Security* (pp. 18-37).

www.irma-international.org/chapter/comparative-analysis-of-various-soft-computing-technique-based-automatic-licence-plate-recognition-systems/279902

Intelligent Prediction Techniques for Chronic Kidney Disease Data Analysis

Shanmugarajeshwari V. and Ilayaraja M. (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 19-37).

www.irma-international.org/article/intelligent-prediction-techniques-for-chronic-kidney-disease-data-analysis/277432

Analysis and Implications of Adopting AI and Machine Learning in Marketing, Servicing, and Communications Technology

Priyal J. Borole (2024). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-11).

www.irma-international.org/article/analysis-and-implications-of-adopting-ai-and-machine-learning-in-marketing-servicing-and-communications-technology/338379