



Chapter 1

An Exploration of Python Libraries in Machine Learning Models for Data Science

Jawahar Sundaram

 <https://orcid.org/0000-0002-8101-8725>
CHRIST University (Deemed), India


Sujith Jayaprakash

 <https://orcid.org/0000-0003-1933-6922>
BlueCrest University College, Ghana

K. Gowri

Sri Ramakrishna College of Arts and Science, India

Harishchander Anandaram

 <https://orcid.org/0000-0003-2993-5304>
Amrita Vishwa Vidyapeetham, India


S. Devaraju

 <https://orcid.org/0000-0003-3116-4772>
VIT Bhopal University, India


C. Manivasagan

Rathnavel Subramaniam College of Arts and Science, India

S. Gokuldev

 <https://orcid.org/0000-0001-8393-4674>
Rathinam College of Arts and Science, India

M. Thenmozhi

 <https://orcid.org/0009-0002-0846-2325>
Sri Eshwar College of Engineering, India

ABSTRACT

Python libraries are used in this chapter to create data science models. Data science is the construction of models that can predict and act on data, which is a subset of machine learning. Data science is an essential component of a number of fields because of the exponential growth of data. Python is a popular programming language for implementing machine learning models. The chapter discusses machine

DOI: 10.4018/978-1-6684-8696-2.ch001

learning's role in data science, Python's role in this field, as well as how Python can be utilized. A breast cancer dataset is used as a data source for building machine learning models using Python libraries. Pandas, numpy, matplotlib, seaborn, scikit-learn, and tensorflow are some Python libraries discussed in this chapter, in addition to data preprocessing methods. A number of machine learning models for breast cancer treatment are discussed using this dataset and Python libraries. A discussion of machine learning's future in data science is provided at the conclusion of the chapter. Python libraries for machine learning are very useful for data scientists and researchers in general.

1. INTRODUCTION

The process of data mining (DM) involves the preparation of data from different sources, such as databases, text files, streams, as well as the modeling of that data using a variety of techniques, depending on the goal that one is trying to achieve, such as classification, clustering, regression, association rule mining. The use of machine learning (ML) techniques in DM enables the discovery of new knowledge in the organization. Data preparation is a part of data analysis which includes preprocessing and manipulating the data as part of the analysis process. There are many aspects involved with data preprocessing, such as cleaning, integrating, transforming and reducing raw data to make it more suitable for analysis, and there are other aspects involved in data wrangling, which is a process of taking the preprocessed data and changing its format so that it can be easily modelled.

Machine learning has grown rapidly in the past few years, and today there are many types and subtypes of machine learning. In the field of machine learning, you are studying what makes computers capable of learning on their own without the need to be explicitly programmed. Using this method, it is possible to solve problems that cannot be solved numerically. Machine learning models can either be classified, grouped, or regressed depending on the purpose they are intended to serve. A linear regression model is used to understand the relationship between the inputs and the outputs of a project's numerical values.

There are many classification models that can be used to identify a particular post's sentiment. An individual's review can be classified as either positive or negative based on the words used in it. Using these models, it is possible to classify emails as spam or not based on their contents. Using a clustering model, we are able to find objects with characteristics that are similar to each other. ML algorithms in many different parts of the world are used in interesting and interesting ways.

29 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/an-exploration-of-python-libraries-in-machine-learning-models-for-data-science/330569

Related Content

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

Machine Learning and Emotions: The Hidden Language in Your Voice

Jesús Heriberto Orduño-Osuna, María E. Raygoza L., Roxana Jiménez-Sánchez, Guillermo M. Limón-Molina and Fabian N. Murrieta-Rico (2024). *Machine and Deep Learning Techniques for Emotion Detection* (pp. 1-23).

www.irma-international.org/chapter/machine-learning-and-emotions/347289

Using Open-Source Software for Business, Urban, and Other Applications of Deep Neural Networks, Machine Learning, and Data Analytics Tools

Richard S. Segall and Vidhya Sankarasubbu (2022). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-28).

www.irma-international.org/article/using-open-source-software-for-business-urban-and-other-applications-of-deep-neural-networks-machine-learning-and-data-analytics-tools/307905

A Review on Time Series Motif Discovery Techniques and Application to ECG Signal Classification: ECG Signal Classification Using Time Series Motif Discovery Techniques

Ramanujam Elangovan and Padmavathi S. (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 39-56).

www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-and-application-to-ecg-signal-classification/238127

Measuring Throughput and Latency of Machine Learning Techniques for Intrusion Detection

Winfred Yaokumah, Charity Y. M. Baidoo and Ebenezer Owusu (2024). *Handbook of Research on AI and ML for Intelligent Machines and Systems* (pp. 27-49).

www.irma-international.org/chapter/measuring-throughput-and-latency-of-machine-learning-techniques-for-intrusion-detection/334467