Top Five Machine Learning Libraries in Python: A Comparative Analysis

Mothe Rajesh and M. Sheshikala

Abstract Nowadays machine learning (ML) is used in all sorts of fields like health care, retail, travel, finance, social media, etc. ML system is used to learn from input data to construct a suitable model by continuously estimating, optimizing, and tuning parameters of the model. To attain the stated, Python programming language is one of the most flexible languages, and it does contain special libraries for ML applications, namely SciKit-Learn, TensorFlow, PyTorch, Keras, Theano, etc., which is great for linear algebra and getting to know kernel methods of machine learning. The Python programming language is great to use when working with ML algorithms and has easy syntax relatively. When taking the deep-dive into ML, choosing a framework can be daunting. The most common concern is to understand which of these libraries has the most momentum in ML system modeling and development. The major objective of this paper is to provide extensive knowledge on various Python libraries and different ML algorithms in comparison with meet multiple application requirements. This paper also reviewed various ML algorithms and application domains.

Keywords Machine Learning · Libraries · Python

1 Introduction

Machine learning (ML) is the most popular technology in today's world. The ML-domain is very immense, and it is developing quickly, being constantly apportioned and sub-parcelled relentlessly into various sub-fortes and types [1]. AI is the domain of study that enables PCs to learn without being unequivocally modified and tackles issues that can't be addressed by mathematical. Among the various kinds of ML assignments, a pivotal qualification is drawn among regulated and unaided learning: Supervised machine learning, the program is "prepared" on a predefined set of "preparing models", which then, at that point, work with its capacity to arrive at an

M. Rajesh (✉) · M. Sheshikala
SR University, Warangal, Telangana, India
e-mail: mraj1210@gmail.com

V. Bhateja et al. (eds.), *Intelligent System Design*, Lecture Notes in Networks and Systems 494, https://doi.org/10.1007/978-981-19-4863-3_55

exact resolution when given new information. Solo machine learning is a program that gives a lot of information and should discover examples and connections in that.

These days, the strength of an organization is estimated by the measure of information it has. These organizations examine the information and concentrate valuable data. For example, E-Marts continues proposing items dependent on your buying patterns and Facebook and Twitter consistently recommend companions and posts in which you may be intrigued. Information in crude structure resembles raw petroleum, we need to filter raw petroleum to make petroleum and diesel. Additionally, machine learning is helpful in handling the information to get valuable experiences.

Machine learning algorithms are being applied in heaps of spots curiously. It's turning out to be progressively universal with an ever-increasing number of utilization's in places where we can't consider, in some application of medical field, academic and Data Center Optimization. In the medical field, machine learning assumes a fundamental part and is as a rule progressively applied to clinical image division, image enlistment, multi-modal picture combination, PC supported conclusion, image directed care will be taken, picture comment, and picture information base recovery, where disappointment could be lethal [2]. In academics, educators need to plan showing materials, physically grade understudies' schoolwork, and give criticism to the understudies on their learning progress. Understudies, then again, regularly amazingly troublesome "one-size-fits-all" training measure that isn't customized to their capacities, needs, and training setting [3].

In ongoing advances, ML furnish new freedoms to handle difficulties in instruction framework by gathering and dissecting the understudy's information and produce when they cooperate with a learning framework. Those large number of racks of murmuring workers utilize tremendous measures of energy; together, all current server farms utilize generally 2% of the world's power, and whenever left unchecked, this energy request could develop as quickly as Internet use. So, making server farms run as productively as conceivable is an exceptionally serious deal. ML is appropriate for the DC climate given the intricacy of plant activities and the plenitude of existing observing information and a portion of the undertakings dealt with by machine learning DC Plant Configuration Optimization [4].

2 Overview of Machine Learning Libraries

Python is becoming famous step by step and has begun to supplant numerous well-known dialects in the business. The effortlessness of Python has drawn in numerous designers to assemble libraries for machine learning and data science, due to this load of libraries, Python is practically well known as R for data science. It is an engaging decision for an algorithmic turn of events and exploratory information examination [5, 6]. The overview of top machine learning libraries in Python is shown in Fig. 1.

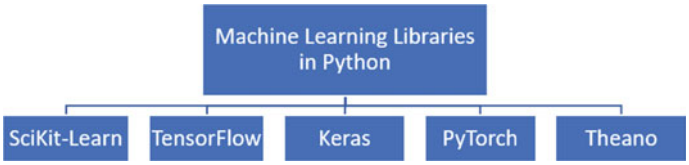

Fig. 1 Overview of top machine learning libraries in Python

2.1 Scikit-Learn

Scikit-learn [7] is a famous AI library in Python. It is a library coordinating a wide range of best-in-class AI algorithms for medium-scale directed and solo issues. Accentuation is put on convenience, execution, documentation, and API consistency. It has insignificant conditions and is circulated under the worked-on empowering its utilization in both scholastic and business settings. Scikit-learn accompanies the help of different calculations, for example, Classification, Regression, Clustering, Dimensionality Reduction, Model Selection, and Pre-handling.

Scikit-learn enjoys a few benefits like, it has a go-to bundle that comprises a multitude of strategies for executing the standard calculations of AI. It has a basic and predictable interface that helps fit and change the model over any dataset. It is the most appropriate library for making pipelines that assist with building a quick model. It is additionally the awesome solid arrangement of machine learning models. On the other hand has a few disservices additionally like, it isn't fit for utilizing downright information to calculations, and it is intensely subject to the SciPy stack.

2.2 TensorFlow

TensorFlow [8] (TF) is created by Google Brain and a team of Google, it has decent documentation, a lot of functionalities alongside the rudiments and it is feasible to make code entirely customizable. Since it is composed as a low-level library, it is a bit more earnestly to dominate. TensorBoard is a representation instrument that accompanies every one of the standard establishments of TF. It permits clients to screen their models, boundaries, misfortunes, and substantially more. A portion of the fundamental regions where TensorFlow sparkles are Handling profound neural organizations, NLP, Abstraction capacities, Image, Text, and Speech acknowledgment, Effortless cooperation of thoughts and code. The center undertaking of TensorFlow is to fabricate profound learning models. Figure 2 illustrates the data flow processing mechanism in TensorFlow.

TensorFlow enjoys many benefits, some resemble assists us with carrying out support learning. We can straight away imagine machine learning models utilizing TensorBoard, an instrument in the TensorFlow library. We can convey the models constructed utilizing TensorFlow on CPUs just as GPUs. On the other hand, it

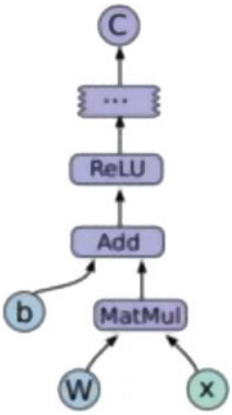

Fig. 2 Data flow processing in TensorFlow

has a few weaknesses like, it runs significantly more slow in contrast with those CPUs/GPUs that are utilizing different structures. The computational charts in TensorFlow are moderate when executed.

2.3 Keras

Keras [9] is based on and works on top of TensorFlow programming in it, which is like the way on a more significant level. The expense for that is harder customization of code. Notably, customization and tweaking of code are a lot simpler when coding at a low level. Keras includes a few of the structure squares and apparatuses fundamental for making a neural organization, for example, Neural layers, Activation and cost capacities, Objectives, Batch standardization, Dropout, Pooling.

Keras enjoys a few benefits like, it is the awesome examination work and product prototyping. The Keras system is compact. It permits a simple portrayal of neural organizations. It is exceptionally productive for representation and demonstrating. On other hand, it has a few impediments like it is delayed as it requires a computational chart before carrying out an activity.

2.4 PyTorch

PyTorch [10] (PT) is created as well as utilized by Facebook. This was grown later than TensorFlow, however, its local area is developing quickly. PyTorch runs its code in a more procedural style, while in TensorFlow, one first necessity to plan the entire model and afterward run it inside a session. Along these lines, it is a lot simpler to troubleshoot code in PyTorch. It has more "pythonic" codes, it is simpler to learn and simpler to use for speedy prototyping. PyTorch and Keras additionally have great documentation. A portion of the crucial provisions that put PyTorch

aside from TensorFlow is Tensor registering with the capacity for sped-up handling through GP units, easy to study, utilize and coordinate with the remainder of the Python environment, support for neural organizations based on a tape-based auto diff framework. The different modules PyTorch accompanies that help makes and train neural organizations are Tensors-torch.Tensor, Optimizers-torch.optim module, Neural Networks-nn module, and Autograd.

PyTorch enjoys a few benefits like its system is famous for its speed of execution. It is equipped for taking care of amazing charts. It additionally incorporates different Python articles and libraries. On the other hand, it has a few weaknesses like the local area for PyTorch isn't broad, and it slacks to give content to questions. In contrast with other Python structures, PyTorch has lesser elements as far as giving representations and application investigating.

2.5 Theano

Theano [11] library is a Python interface for the upgrading compiler. After enhancement and accumulation, the capacities become accessible as regular Python capacities, yet have superior. Vector, grid, and tensor activities are upheld and productively paralleled on accessible equipment. A portion of the elements that make Theano a hearty library for doing logical estimations. It assists for GPUs to do more in uncompromising calculations contrasted with CPUs, strong joining relation with NumPy, faster and stable assessments of even the trickiest of factors, and ability to make custom C code for your numerical activities. Theano enjoys a few benefits like, it upholds GPUs that assist applications with performing complex calculations proficiently, it is straightforward execute Theano as a result of its combination with NumPy, there is an immense local area of engineers utilizing Theano. On another hand, it has a few burdens like, it is slower in the back end, there are different issues in Theano's low-level API, it gives a ton of back-end blunders. Also, the Theano library has a precarious expectation to absorb information.

The overall comparison of machine learning libraries is done in the next section, here in Table 1, all the basic information of the libraries is mentioned like developed by whom, launched in which year, written in which languages, and well-known applications.

3 Comparative Analysis of Machine Learning Libraries

Table 2 gives the libraries information taken from GitHub. It is a popular online hosting service for version control [12]. Here the information is taken based on parameters like the number of stars, forks, contributors, and activity on the library repository [13]. If we look into the table, all the libraries have a good number of stars, which means all the libraries are performing better and users giving good feedback.

Table 1 Machine learning libraries basic information

Library	Developed by	Launched year	Written in	Well-known applications
Scikit-Learn	David Cournapeau	2007	Python, C, C++, and Cyhton	Spotify, Inria
TensorFlow	Google Brain and team of Google	2015	Python, CUDA, and C++	Lib
PyTorch	Facebook's AI Research Lab	2016	Python, CUDA, and C++	Apple and Samsung Electronics
Keras	François Chollet	2015	Python	Uber and Netflix
Theano	MILA, University of Montreal	2007	Python, CUDA	Zetaops and Vuclip

Table 2 GitHub comparative information about machine learning libraries

Library	Stars	Forked	Contributors	Activity
Scikit-Learn	33,337	16,358	1253	38 (94)
TensorFlow	120,548	72,009	1835	193 (1889)
PyTorch	24,782	5877	933	151 (912)
Keras	38,197	14,585	774	21 (52)
Theano	2033	475	46	3 (17)

In the list of specified libraries, TensorFlow has more contributors, which implies that it has more popularity among the all other machine learning libraries [14–17]. Theano is becoming famous day by day in machine learning applications.

4 Conclusions

There are many more libraries in the ML world, but these are the most popular and widely used. ML is a huge world and the most promising tech right now. No matter the programming language or the area a developer is working in, learning to work with libraries is important. Doing so helps in de-complexing the things and to cut the tedious effort.

References

1. https://www.toptal.com/machine-learning/machine-learningtheory-an-introductory-prime
2. There A, Jeon M, Sethi IK, Xu B (2017) Machine learning theory and applications for healthcare. J Healthc Eng 2017. Article ID 5263570

3. Dilhara M, Ketkar A, Dig D (2021) Understanding software-2.0: a study of machine learning library usage and evolution. ACM Trans Softw Eng Methodol (TOSEM) 30(4):1–42
4. Gao J (2014) Machine learning applications for data center optimization
5. Dubois PF (ed) (2007) Python: batteries included, volume 9 of computing in science & engineering. IEEE/AIP
6. Milmann KJ, Avaizis M (eds) (2011) Scientific Python, volume 11 of computing in science & engineering. IEEE/AIP
7. Pedregosa F et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
8. Abadi M et al (2016) TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI'16)
9. Chollet F (2019) Keras. https://github.com/fchollet/keras
10. PyTorch (2019) https://pytorch.org/
11. Theano Development Team (2016) Theano: a Python framework for fast computation of mathematical expressions. eprint: arXiv:1605.02688
12. GitHub (2019) https://github.com/
13. Sheshikala M, Kothandaraman D, Vijaya Prakash R, Roopa G (2019) Natural language processing and machine learning classifier used for detecting the author of the sentence. Int J Recent Technol Eng 8(3):936–939
14. Ravi Kumar R, Babu Reddy M, Praveen P (2019) An evaluation of feature selection algorithms in machine learning. Int J Sci Technol Res 8(12):2071–2074
15. Kumar RR, Reddy MB, Praveen P (2019) Text classification performance analysis on machine learning. Int J Adv Sci Technol 28(20):691–697
16. Kollem S, Reddy KRL, Rao DS (2019) A review of image denoising and segmentation methods based on medical images. Int J Mach Learn Comput 9(3):288–295
17. Tizpaz-Niari S, Černý P, Trivedi A (2020) Detecting and understanding real-world differential performance bugs in machine learning libraries. In: Proceedings of the 29th ACM SIGSOFT international symposium on software testing and analysis. Association for Computing Machinery, New York, NY, pp 189–199. https://doi.org/10.1145/3395363.3404540