

# Framework for Data Analytics

Tushar B. Kute,  
<http://tusharkute.com>



# Framework

- Framework in software terminology is a collection of individual software components available in code form and ready to run (what we call as libraries) that can be run independently or together to achieve a complicated task on any machine.
- The important part is ready to run, which means you do not have to put any effort in reinventing the wheel, it is already done for you, and you just have to learn to customize them to your liking to create your application-specific software, suiting your business needs.

# Top Frameworks

- TensorFlow
- Scikit-learn
- Keras
- Pandas
- Spark MLlib
- PyTorch
- Matplotlib
- Numpy
- Seaborn
- Theano

# Tensorflow

- TensorFlow is an end-to-end Machine Learning platform featuring comprehensive, flexible framework of tools and libraries along with community resources, helping you build Machine Learning powered applications easily.
- It also makes it easy to integrate data like inputting SQL tables, graphs, images all together.
- TensorFlow was first created by Google Brain Team and to this day remains open-source.

# Scikit-learn

- Scikit-learn is an open-source Machine Learning library for use in Python programming language, featuring various classification, clustering and regression algorithms.
- It is designed to interoperate with numerical and scientific libraries like NumPy and SciPy, both developed and used in Python.

# Keras

- Keras is a popular open-source software library that is capable of running atop other libraries like TensorFlow, Theano and CNTK.
- With a lot of data, you can dabble in Deep Learning and AI over this framework.

# Pandas

- A data manipulation and analysis language written in python and for python offer data structures and operations for manipulating Numpy based tables and time series.
- It is used to normalize incomplete and messy data with features of shaping, slicing, dicing and merging datasets.

# Spark MLlib

- A library with an extensive support for Java, Scala, Python and R, this framework can be used on Hadoop, Apache Mesos, Kubernetes, over cloud services dealing with multiple data sources.



# PyTorch

- A Facebook developed framework, PyTorch is an AI-specific framework for Deep Learning.
- The PyTorch library allows dynamic updates of graphs allowing on the fly changes to the architecture.

# Matplotlib

- Based on MATLAB, Matplotlib is a plotting library for Python, with extensive support for rich visualization and dynamic charts.
- It is a numerical extension of the Numpy library to generate stunning graphs and plots.
- The default visualization library in every data science project in Python, Matplotlib helps you create interactive visualizations including histograms, 3Dplots, scatterplots, image plots, bar charts and many more

# Numpy

- Numpy, an open-source library, brings in the computational power of C to Python, with powerful data structures for number-crunching applications like Quantum Computing, Statistical computing, signal processing, image processing, graphs and networks, astronomy processes, cognitive psychology and more.

# Seaborn

- An open-source Python library, Seaborn is a visualization package based on Matplotlib.
- You get to work with high-level interfaces for producing rich and attractive statistical graphs.

# Theano

- Similar to Numpy, Theano is for numerical computation and is best at manipulating and evaluating mathematical expressions.
- Theano ensures that computations are expressed efficiently on either CPU or GPU architectures.

# Thank you

*This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License*



@mitu\_skillologies



/mITuSkillologies



@mitu\_group



/company/mitu-  
skillologies



MITUSkillologies

## Web Resources

<https://mitu.co.in>

<http://tusharkute.com>

[contact@mitu.co.in](mailto:contact@mitu.co.in)

[tushar@tusharkute.com](mailto:tushar@tusharkute.com)