**DATA MINING**

**ISE**

**Topic: Exclusive Python frameworks for AI & ML application**

**Name: Saurabh Nagre**

**PRN: 2019BTECS00080**

**BATCH: B7**

**Introduction:**

In the 1950's, a young mathematician Alan Turing explored the possibility of the existence of artificial intelligence mathematically. Framework serves as a foundation, so you're not starting entirely from scratch. In AI ML, we need to build an algorithmic model. These frameworks help us to work on boiler plate code.We can build classification, regression, clustering, speech recognition using these frameworks.

The number of ML algorithms, as well as their different software implementations, is quite large. Many software tools for DM using ML techniques have been in development for the past 25 years (Jovic et al. 2014). Their common goal is to facilitate the complicated data analysis process and to propose integrated environments on top of standard programming languages. These tools are designed for various purposes: as analytics platforms, predictive systems, recommender systems, processors (for images, sound or language). A number of them are oriented to fast processing and streaming of large-scale data, while others are specialized in implementing ML algorithms including NNs and DL. Again, it is important to emphasize that there is no single tool suitable for every problem and often a combination of them is needed to succeed.

**Significance:**

We want our work to get simplified, that is what frameworks and libraries do for AI-ML developers. Python is popular for vast library support. We need to avoid coding duplicate code, and to focus on our particular project, that’s where python AI ML frameworks help out.

**Literature Survey:**  Dr. V. Hanuman Kumar, Sr. Assoc. Prof., NHCE, Bangalore-560 103

He wrote a journal title as “Python Libraries, Development Frameworks and Algorithms for Machine Learning Applications”. In this he explained most of python frameworks and libraries of different categories including visualization, statistical, ML algorithms, Deep learning and Neural Networks. He also explained the use cases of AI and ML and some of the algorithms in it.

**Theory:**

1. TensorFlow

TensorFlow is an open-source software library for numerical computation using data flow graphs (TensorFlow 2018). TensorFlow was created and is maintained by the Google Brain team within Google’s Machine Intelligence research organization for ML and DL. It is currently released under the Apache 2.0 open-source license. TensorFlow is designed for large-scale distributed training and inference. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The distributed TensorFlow architecture contains distributed master and worker services with kernel implementations. These include 200 standard operations, including mathematical, array manipulation, control flow, and state management operations written in C++. TensorFlow was designed for use both in research, development and production systems. It can run on single CPU systems, GPUs, mobile devices and large-scale distributed systems of hundreds of nodes.

2. Pytorch:

PyTorch is a Python library for GPU-accelerated DL (PyTorch 2018). The library is a Python interface of the same optimized C libraries that Torch uses. It has been developed by Facebook’s AI research group since 2016. PyTorch supports tensor computation with strong GPU acceleration, and DNNs built on a tape-based autograd system. It has become popular by allowing complex architectures to be built easily. Typically, changing the way a network behaves means to start from scratch. PyTorch uses a technique called reverse-mode auto-differentiation, which allows to change the way a network behaves with small effort (i.e. dynamic computational graph or DCG).

3. Apache Spark ML

SparkMLlib/ML containsMachine Learning algorithms such as classification, regression, clustering or collaborative filtering; featurization tools for feature extraction, transformation, dimensionality reduction and selection; pipeline tools for constructing, evaluating and tuning ML pipelines; and persistence utilities for saving and loading algorithms, models and pipelines. It also contains tools for linear algebra, statistics and data handling. With the exception of the distributed data parallel model, MLlib can be easily used together with stream data as well. For this purpose, MLlib offers few basic ML algorithms for stream data such as streaming linear regression or streaming k-means. For a larger class of ML algorithms, one has to let models learn offline and then apply the model on streaming data online.

**Conclusion:**

We have discussed various python libraries and development frameworks used to build up a Machine Learning system.

**References:**

* <https://www.tensorflow.org/overview>
* [https://pytorch.org/](https://pytorch.org/docs/stable/index.html)
* <https://spark.apache.org/docs/latest/ml-guide.html>
* <https://towardsdatascience.com/top-8-ai-and-machine-learning-frameworks-for-beginners-4e3996fc27f>
* <https://towardsdatascience.com/your-first-apache-spark-ml-model-d2bb82b599dd>