Tools for Data Science

▼ Week 3

Libraries for Data Science:

- Libraries usually contain built-in modules providing different functionalities that can be used directly
- You can use data visualization methods to communicate with others and display meaningful results of an analysis
- For machine learning, the Scikit-learn library contains tools for statistical modeling including regression, classification, clustering and so on
- TensorFlow is a low-level framework used in large scale production of deep learning models
- Apache Spark is a general-purpose cluster-computing framework allowing you to process data using compute clusters

Application Programming Interfaces (APIs):

- An application programming interface (API) allows communication between two pieces of software
- API is the part of the library you see while the library contains all the components of the program
- REST APIs allow you to communicate through the internet and take advantage of resources like storage, data, artificially intelligent algorithms, and much more

Data Sets:

- Open data is fundamental to Data Science
- Community Data License Agreement makes it easier to share open data
- Open data sets might not meet enterprise requirements, due to the impact they might have on the business

Sharing Enterprise Data - Data Asset eXchange:

- The IBM Data Asset eXchange (DAX) site contains high-quality open data sets
- DAX open data sets include tutorial notebooks that provide basic and advanced walk throughs for developers
- DAX and MAX are available on the IBM Developer website
- You can get, run, and preview data sets and notebooks on DAX
- DAX notebooks are opened in Watson Studio

Machine Learning Models - Learning from Models to Make Predictions:

- Machine learning (ML) uses algorithms – also known as “models” – to identify patterns in the data
- The process by which the model learns data patterns is called “model training”
- Types of ML are Supervised, Unsupervised, and Reinforcement
- Supervised learning comprises two types of models, regression and classification
- Deep learning refers to a general set of models and techniques that loosely emulate the way the human brain solves a wide range of problems

The Model Asset eXchange:

- The Model Asset eXchange is a free open source repository for ready-to-use and customizable deep learning microservices
- To reduce time to value, consider taking advantage of pre-trained models for certain types of problems
- MAX model-serving microservices are built and distributed on GitHub as open source Docker images
- Red Hat OpenShift is a Kubernetes platform used to automate deployment, scaling, and management of microservices
- Ml-exchange.org has multiple predefined models
- The CodePen tool lets users edit front-end languages

Summary:

- Libraries usually contain built-in modules that provide different functionalities.
- You can use data visualization methods to communicate with others and display meaningful results of an analysis.
- For machine learning, the Scikit-learn library contains tools for statistical modeling, including regression, classification, clustering, and so on.
- Large-scale production of deep-learning models use TensorFlow, a low-level framework.
- Apache Spark is a general-purpose cluster-computing framework that allows you to process data using compute clusters.
- An application programming interface (API) allows communication between two pieces of software.
- API is the part of the library you see while the library contains all the components of the program.
- REST APIs allow you to communicate through the internet and take advantage of resources like storage, data, artificially intelligent algorithms, and much more.
- Open data is fundamental to Data Science.
- Community Data License Agreement makes it easier to share open data.
- The IBM Data Asset eXchange (DAX) site contains high-quality open data sets.

- DAX open data sets include tutorial notebooks that provide basic and advanced walk-throughs for developers.
- DAX notebooks open in Watson Studio.
- Machine learning (ML) uses algorithms – also known as “models” – to identify patterns in the data.
- Types of ML are Supervised, Unsupervised, and Reinforcement.
- Supervised learning comprises two types of models, regression and classification.
- Deep learning refers to a general set of models and techniques that loosely emulate the way the human brain solves a wide range of problems.
- The Model Asset eXchange is a free, open-source repository for ready-to-use and customizable deep-learning microservices.
- MAX model-serving microservices are built and distributed on GitHub as open-source Docker images.
- You can use Red Hat OpenShift, a Kubernetes platform, to automate deployment, scaling, and management of microservices.
- Ml-exchange.org has multiple predefined models.

Reading: **GettingStartedwithMAX&DAX.pdf**