ACC: ACC 3

ACC Bucket: Data Science & Analytics - 3

Course: Applied Data Science

OVERALL COURSE OBJECTIVES: The aim of these courses is to empower students to master advanced data science by understanding and applying deep learning models, exploring and analyzing big data, and leveraging various high-level tools to handle real-life examples across several domains, ultimately contributing to the broader field of Artificial Intelligence.

LEARNING OUTCOMES: On successful completion of the course the students shall be able to:

1. Understand and apply deep learning models to various disciplines like Natural Language Processing, Computer Vision, and Time Series Analysis using Keras and TensorFlow.
2. Implement Anomaly Detection, Time Series Forecasting, and Image Recognition on real-life examples from IoT, Financial Market Data, and Literature.
3. Scale artificial brains using tools like Kubernetes, Apache Spark and GPUs.
4. Build a big data ecosystem using tools and methods from the courses in the specialization.
5. Analyze and explore a big data set simulating user-generated data, and conduct exploratory analysis using tools such as Splunk and Open Office.
6. Prepare, analyze, and report big data problems using advanced tools like KNIME, Spark's MLLib and Gephi.

COURSE CONTENT:

Module 1: Applied Machine Learning in Python [31 Hours]

This course will introduce the learner to applied machine learning, focusing more on the techniques and methods than on the statistics behind these methods. The course will start with a discussion of how machine learning is different than descriptive statistics, and introduce the scikit learn toolkit through a tutorial. The issue of dimensionality of data will be discussed, and the task of clustering data, as well as evaluating those clusters, will be tackled. Supervised approaches for creating predictive models will be described, and learners will be able to apply the scikit learn predictive modelling methods while understanding process issues related to data generalizability (e.g. cross validation, overfitting). The course will end with a look at more advanced techniques, such as building ensembles, and practical limitations of predictive models. By the end of this course, students will be able to identify the difference between a supervised (classification) and unsupervised (clustering) technique, identify which technique they need to apply for a particular dataset and need, engineer features to meet that need, and write python code to carry out an analysis.

Sub-Topic
Fundamentals of Machine Learning - Intro to SciKit Learn
Supervised Machine Learning - Part 1 & 2

Formative Assessments:
4 graded quizzes and 10 programming assignments/labs.

Module 2: Big Data - Capstone Project [21 Hours]

The Capstone Project for Big Data is a culminating project in which learners build a big data ecosystem using tools and methods from previous courses. The project involves analyzing a simulated big data set from an imaginary game, "Catch the Pink Flamingo". This five-week project walks learners through various big data science steps such as acquisition, exploration, preparation, analysis, and reporting. Initially, the course introduces the data set and guides through exploratory analysis using tools like Splunk and Open Office. Later, it delves into advanced tools like KNIME, Spark's MLLib, and Gephi for complex big data problems. In the final week, learners are taught to create engaging reports and presentations. A collaboration with Splunk allows top project learners to present to Splunk and interact with their recruiters and engineering leadership.

Sub-Topic
Acquiring, Exploring, and Preparing the Data
Clustering with Spark
Data Classification with KNIME
Graph Analytics of Simulated Chat Data With Neo4j
Reporting and Presenting Your Work
Simulating Big Data for an Online Game

Formative Assessments:
1 quizzes and 5 peer-review assignments.

ASSESSMENT:

For summative assessments, Coursera will provide question banks for which exams can be conducted on the Coursera platform or the faculty will create their own assessments.

Note: If a Course or Specialization becomes unavailable prior to the end of the Term, Coursera may replace such Course or Specialization with a reasonable alternative Course or Specialization.