Implementing Predictive Analytics with Spark in Azure HDInsight

Course Syllabus

Prerequisites

Students taking this course should have the following prerequisite knowledge and skills:

- Familiarity with Hadoop clusters in HDInsight
- Familiarity with database concepts and basic SQL query syntax.
- Familiarity with basic programming constructs (for example, variables, loops, conditional logic). Experience with Scala or Python in Spark is preferable.
- A willingness to learn actively and persevere when troubleshooting technical problems.

Course Outline

The course consists of the following modules:

- Module 1: Introduction to Data Science with Spark
- Module 2: Machine Learning with Spark
- Module 3: Real-time Machine Learning with Spark
- Module 4: Final Exam

The course should take an estimated 3-4 hours per week for four weeks, including lecture, further reading, hands-on labs, and assessments. A further week is allowed for assignment submission.

Grading Policy

This course includes coursework, some of which is graded. Each module in the course includes an ungraded lab (which is designed to give you hands-on practice with the technologies and techniques taught in the module), and a graded assessment, in which you must answer all questions. Additionally, at the end of the course you must complete a final exam.

The module assessments account for 50% of the total grading for the course, and the final exam accounts for the remaining 50%. You must achieve an overall score of 70% or more to pass this course.

In the module assessments, you have two attempts at each question, and the correct answers can be displayed after you submit your final answer. In the final exam, you are restricted to one attempt per question, and to protect the integrity of the exam, the correct answers are not displayed.

All module assessments and the final exam must all be completed before the course deadline. However, we recommend that you complete each module assessment in the same week as the associated module was taken.

Discussion

We encourage all students to submit questions, observations, and comments in the **Discussion** section. If you have any issues while working on the course, check here first – your fellow students may have already found a resolution!

Please remember that the discussion forum is open to all students and staff, and while we love to see passionate engagement, abusive or inflammatory behavior will not be tolerated.

While the course staff will monitor the discussions and respond appropriately where possible, due to the volume of students attending this course, it will not be possible for the course staff to answer every question individually. You should still post questions however, because in many cases, your fellow students may be able to help.