

AI VIET NAM – COURSE 2023

Big Data Exercises

(Spark, PySpark and its Application)

Ngày 6 tháng 8 năm 2023

Probability Exercise - Rubric		
Câu	Kiến Thức	Đánh Giá
Câu 1-10	- Kiến thức về big data: hadoop, spark, pyspark	- Hiểu được cấu trúc cơ bản của Big Data và các công cụ cần thiết cho việc xử lý Big Data
Bài thực hành 1	- Cách thức cài lưu trữ và truy xuất dữ liệu big data với google cloud storage và google drive	- Biết cấu hình cài lưu trữ và truy xuất dữ liệu big data với google cloud storage và google drive
Bài thực hành 2	- Kiến thức cơ bản pyspark, spark session, spark SQL	- Biết cách sử dụng spark SQL để truy vấn dữ liệu lớn.
Bài thực hành 3	- Kiến thức về các thư viện máy học cần thiết xử lý big data với pyspark	- Biết cách sử dụng thư viện máy học trong pyspark để giải quyết bài toán Linear Regression.