

Introduction to the Tutorial

...

A Little bit about myself

- Soumendra Prasad Dhanee
 - Head of Data Science and Senior Data Scientist in startups
 - Until very recently
 - soumendra@gmail.com
 - <https://github.com/soumendra>
 - <https://twitter.com/dataBiryani>
-

Outline

What we'll be learning in this tutorial

- Spark and an introduction to its EcoSystem
 - Using Spark with Python, PySpark and Jupyter Notebooks
 - Basic and Advanced Spark Data Structure
 - Machine Learning with Spark
 - Machine Learning Applications
-

Evaluation Scheme

- 10% for Class Participation
 - Attendance and punctuality
 - Doing homework and setups
 - 30% for group hands-on project
 - 60% for final written exam
-