

بسمه تعالی

پروژه های این درس:

1- <https://spark.apache.org/docs/latest/ml-classification-regression.html>

مطالب بخش ۱ به الگوریتم های مختلف پرداخته که تمامی این الگوریتم ها را بررسی کنید.

2- https://github.com/tirthajyoti/Spark-with-Python/tree/master/Python-and-Spark-for-Big-Data-master/Spark_for_Machine_Learning

مطالب بخش ۲ کد سورس نوشته شده را تحلیل و بررسی نمایید. پس از تحلیل کد مطالب را جمع بندی نمایید.

spark streaming

این بخش هم مربوط به پردازش جریان داده است که بصورت اختیاری هست.

3- <https://github.com/harunurrashid97/Machine-learning-with-PySpark/blob/master/First%20PySpark%20ml%20model/First%20ML%20models%20using%20PySpark.ipynb>

در بخش ۳ پروژه عملی مطابق نمونه که از کتابخانه pandas استفاده شده برای یک دیتاست مشخص و الگوریتم های مختلف دقت مدل ها را بدست آورید.

جهت مطالعه و بررسی:

1- <https://towardsdatascience.com/first-time-machine-learning-model-with-pyspark-3684cf406f54>

2- <https://github.com/harunurrashid97/Machine-learning-with-PySpark/blob/master/First%20PySpark%20ml%20model/First%20ML%20models%20using%20PySpark.ipynb>

3- <https://towardsdatascience.com/machine-learning-with-pyspark-and-mlib-solving-a-binary-classification-problem-96396065d2aa>

