Machine Learning in Spark

Analytics Vidhya

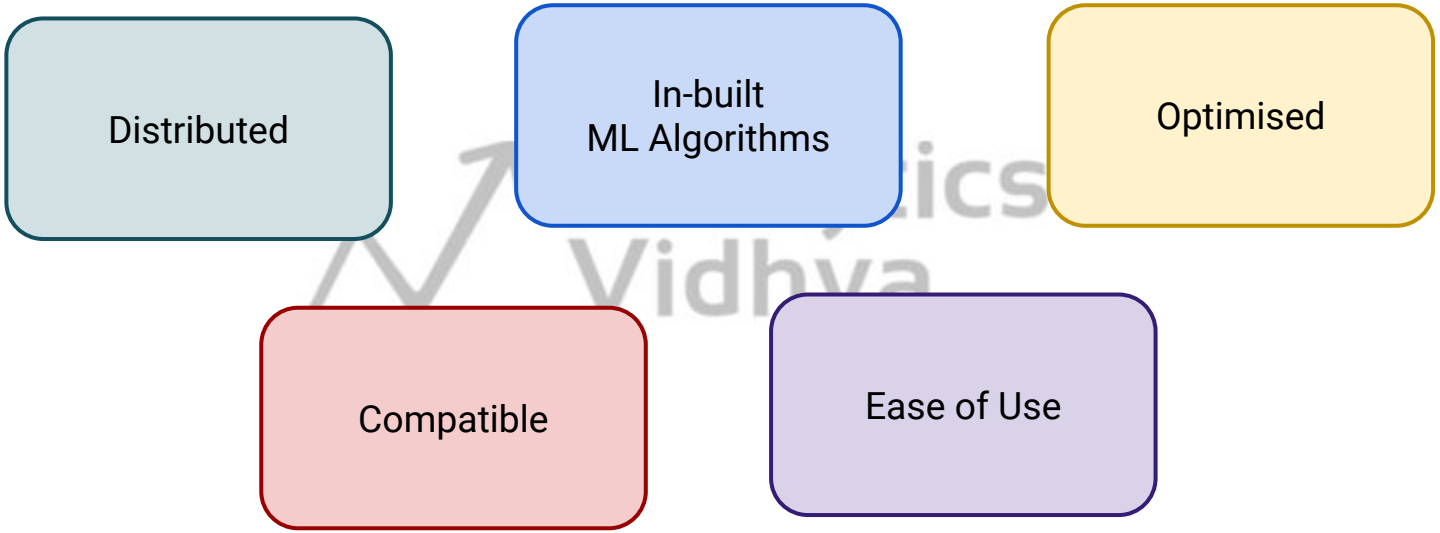
Analytics Vidhya

vs

Analytics Vidhya

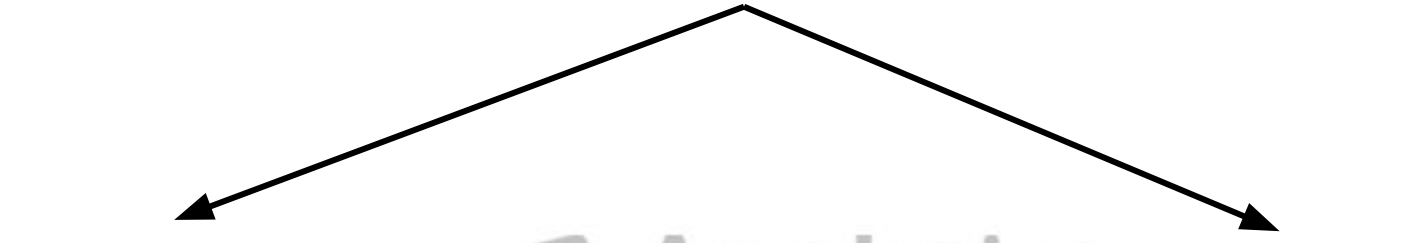
Analytics Vidhya

Why ML in Spark?



Analytics Vidhya

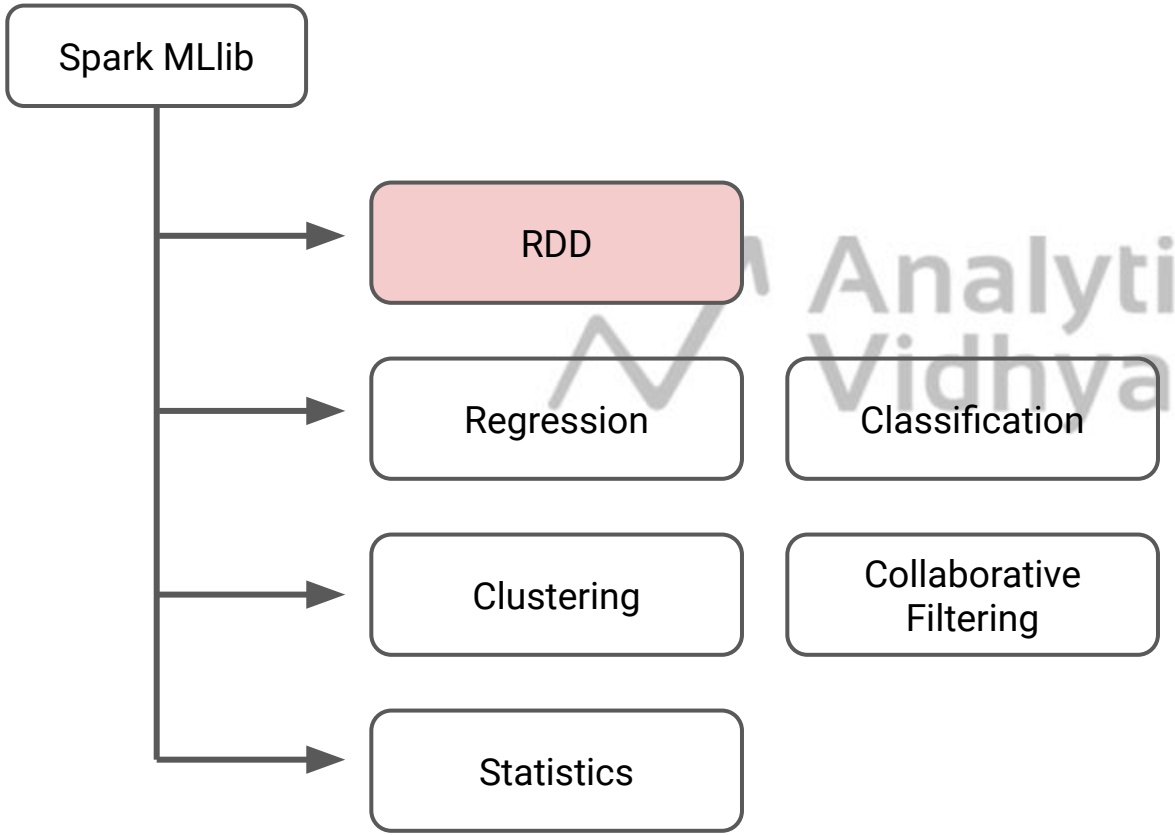
ML Libraries in Spark



Analytics Vidhya

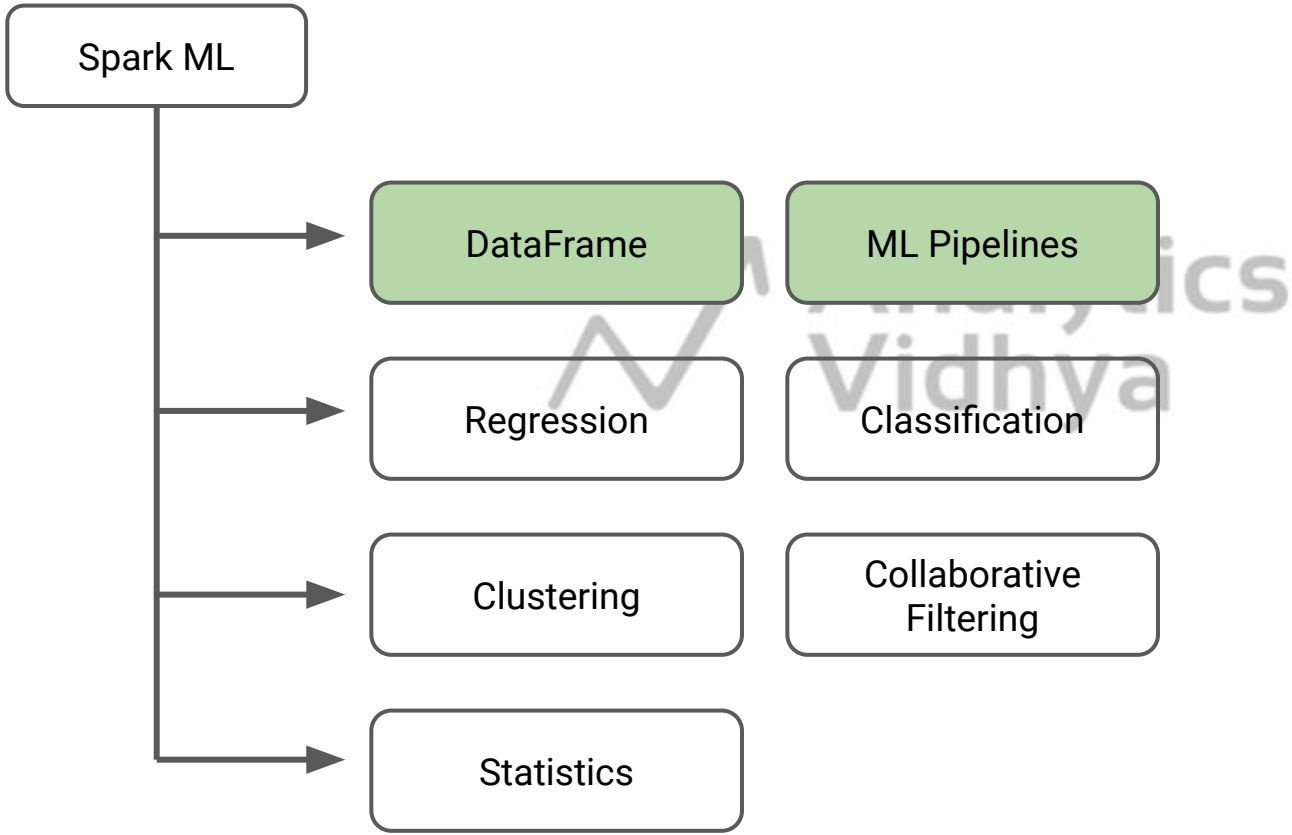
Analytics Vidhya

MLlib - RDD based



Analytics Vidhya

Spark ML - Dataframe based



Analytics Vidhya

Why Spark ML?

It is more user-friendly.

It provides ML pipelines and new features will be created for this only.

The RDD based API is not deprecated but is in maintenance mode.

The RDD based API will not add any new features.

Analytics Vidhya

Analytics Vidhya

Thank You!!

Analytics
Vidhya