

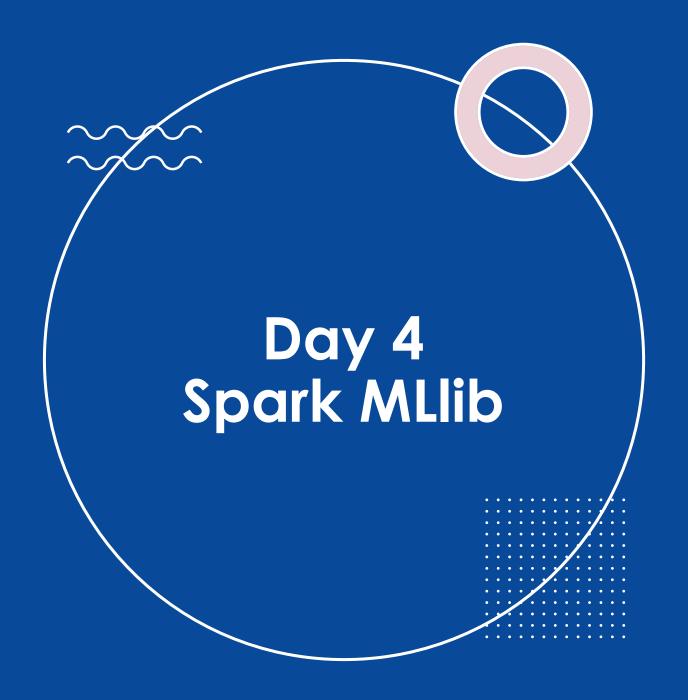


**FDP Kultivations** 

Spark MLlib Workshop



- Day 1
  - Intro Big data Scenario
  - Spark Overview
  - o Rdd's processing with hands on
- Day 2
  - Intro SQL and Big Data motivation
  - Spark Dataframe and SQL API
  - o Hand's on
- Day 3
  - o Full Exercises Day 1 and Day 2
- Day 4
  - Intro to Spark MLlib (Plus)



- What is Machine Learning
- Machine Learning Lifecycle
- Spark MLlibs Algorithms
- Use Cases for Spark MLlibs
- Basket Analysis Lab
- Customer Churn Lab

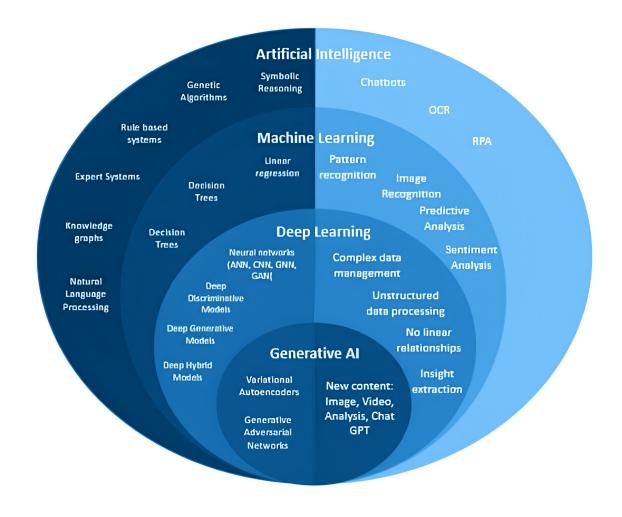


# What is Machine Learning?

While many ML algorithms can be deterministic, some exhibit non-deterministic behavior due to inherent randomness.



Machine Learning is a branch of AI that uses algorithms to learn from data and make predictions or decisions without explicit programming.

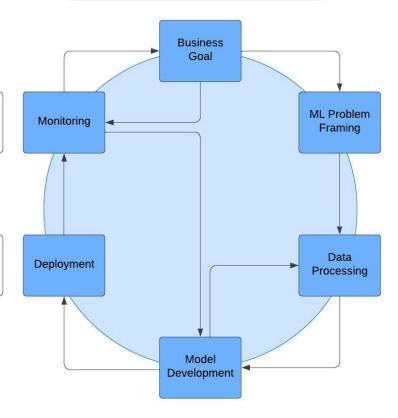


### **Machine Learning Lifecycle**

An organization considering ML should have a clear idea of the problem, and the business value to be gained by solving that problem. You must be able to measure business value against and success criteria.

Ensures your model is maintaining a desired level of performance through early detection and mitigation.

After a model is trained, tuned, evaluated and validated, you can deploy the model into production. You can then make predictions and inferences against the model.



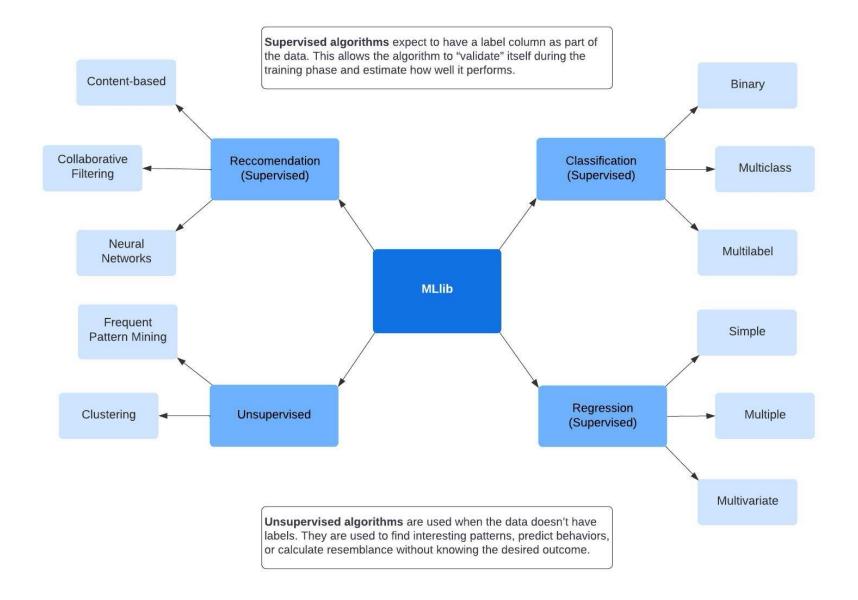
The business problem is framed as a machine learning problem. Determining what to predict and how performance and error metrics must be optimized is a key step in this phase.

Collecting data, preparing data, and feature engineering that is the process of creating, transforming, extracting, and selecting variables from data.

Model building, training, tuning, and evaluation.
Includes creating a CI/CD pipeline that automates
the build, train and release to staging and
production environments.

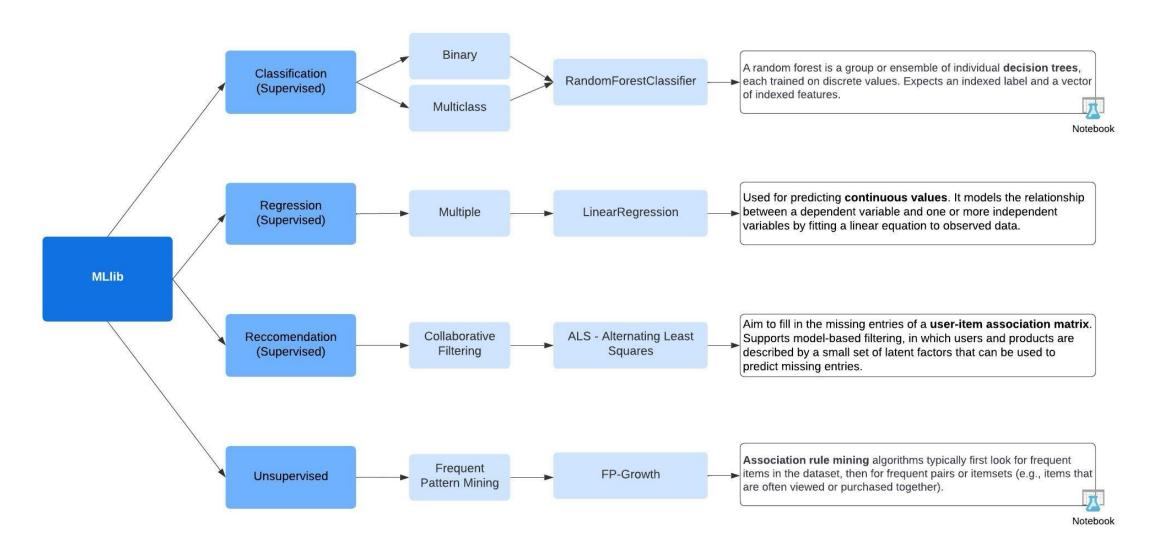


## Spark MLlib Algorithms





#### **Use Cases for Spark MLlib**







# Do you have a quick question?



Data Engineer - Team Friends From Work





