

532 - SYSTEMS FOR DATA SCIENCE
Milestone Goals - 8th April, 2024
Topic: Prediction of Critical Temperatures of Superconductors
Done by - Muskan Kothari, Ujjwal Gupta

Milestones goals:

1. Setting up the spark streaming pipeline to read the data from the file in mini-batches - **COMPLETED**
2. Utilizing data pre-processing by setting up RDDs from the data and performing transformations. - **IN PROGRESS (65%)**
3. Utilize regression models provided in Pyspark ML library to compute predictions on large-scale data with ~10,000 records. - **PENDING**

Justification: Data preprocessing phase is still in progress (2nd milestone goal), with some more analysis to be done on the feature data (numerical), hence the 3rd goal has not been started. As such, currently we are not seeing any blocker or major issues.