Learning Curve for PA4:

Understanding MapReduce and Apache Spark:

- Studied the fundamentals of MapReduce and its applications in distributed systems.

- Explored Apache Spark’s capabilities for data processing, focusing on DataFrame operations and transformations.

Week 3: Developing the MapReduce Logic:

- Implemented the filtering logic to identify incorrect inferences.

- Grouped and aggregated data to compute error counts per producer.

Key Learnings

- Gained proficiency in Apache Spark and its API for distributed data processing.

- Learned to integrate MapReduce concepts into practical implementations.

- Developed debugging skills for distributed systems, especially in Spark.

- Maintain detailed documentation of data transformations for traceability.

This implementation exemplifies the iterative learning process in distributed computing, with Apache Spark serving as a robust tool for tackling real-world data challenges.