IC-PBL+ Project. Streaming Data Processing

Younghoon Kim (nongaussian@hanyang.ac.kr)

Data Streaming Assumptions

- Dataset is learned one example at a time
- Each unique example in the dataset is seen only once
- The ordering of data in the stream is arbitrary
- The model being trained can be evaluated at any point within the data stream

Project Description

- Given
  - Dataset: https://figshare.com/ndownloader/files/31727663
  - Twitter streaming dataset on COVID-19 (2021 Nov)
  - Snowflake IDs of posted messages
- Find
  - Temporal patterns
- Using
  - Opensource projects for processing streaming data
  - Kafka, spark streaming, flink

Goals

- Focus on the tutorial of how to use and build a streaming data processing system using Kafka or Spark Streaming
- Presentation, report and demo
- Team activities
  - Nov 29
- Presentation
  - Dec 6
- Report deadline
  - Dec 13