

# Mining Data Streams

Topics References

# The Stream Data Model

- Chapter 4
- 4.1 The Stream Data Model
- Page 131 (Check Contents Page)
- Book: Anand Rajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, 2nd Edition, Cambridge University Press, 2014, ISBN: 9781107077232.
- <https://www.youtube.com/watch?v=SRTSVxUnsNI&t=15s>

# Sampling Data in a Stream

- Page: 135, Book
- Types of Sampling
  - Fixed proportion,
  - Fixed size,
  - Biased reservoir
- References
  - <https://www.youtube.com/watch?v=895jWrdNA5I>
  - <https://www.youtube.com/watch?v=MbvZijOrF3s>

# Filtering Streams

- Bloom Filters
- Probability of not getting false positives
- Page: 138
- <https://www.youtube.com/watch?v=qBTdukbzc78&t=328s>
- <https://www.youtube.com/watch?v=0P2mkqkrfck>

# Counting Distinct Elements in a Stream

- Page 142
- Flajolet-Martin Algorithm:
- References:
- <https://www.youtube.com/watch?v=A5DOq-SGt5A&t=495s>

# Moments

- Page 145

# Decaying Windows

- Page 157

# A Stream-Clustering Algorithm

- Page 283