(k,P)-Anonymity

KAPRA Algorithm Implementation

- Alex Di Stefano (S4638131)
- Christian Stingone (S4644983)

Contents

- k and P Anonymization Levels
- KAPRA Algorithm
- Dataset
- Analysis

k and P Anonymization Levels

k-requirement: Each anonymization envelope appears at least k times.

P-requirement: Consider any k-group G of time series having the identical anonymization envelope, there are at least P-1 other time series in G having the same QI pattern representation as PR[r]

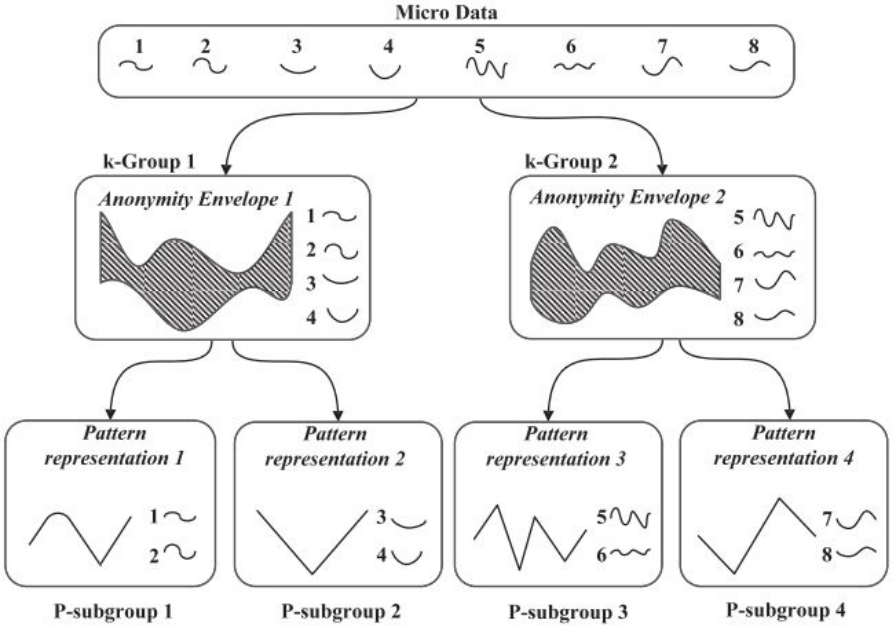
k and P Anonymization Levels

(k,P)-Anonymity is based on two level:

- **First level:** k-anonymity is required for time series in the entire database. That means the records in the published database can be grouped by the quasi-identifier attribute values, and each group should contain at least k records.

- **Second level:** P-anonymity is required for the pattern representations (PRs) associated with each record in a same group. Specifically, each group can be divided into subgroups, each of which contains at least P records having identical PRs.

k and P Anonymization Levels



The k-groups and P-subgroups of (k,P)-Anonymity

KAPRA Algorithm

- Create-tree phase with entire dataset
  - Initialization
  - Node Splitting
- Recycle bad-leaves phase
- Group formation phase
  - Top-Down Preprocessing
  - Group Formation
  - Group Post Processing

Dataset

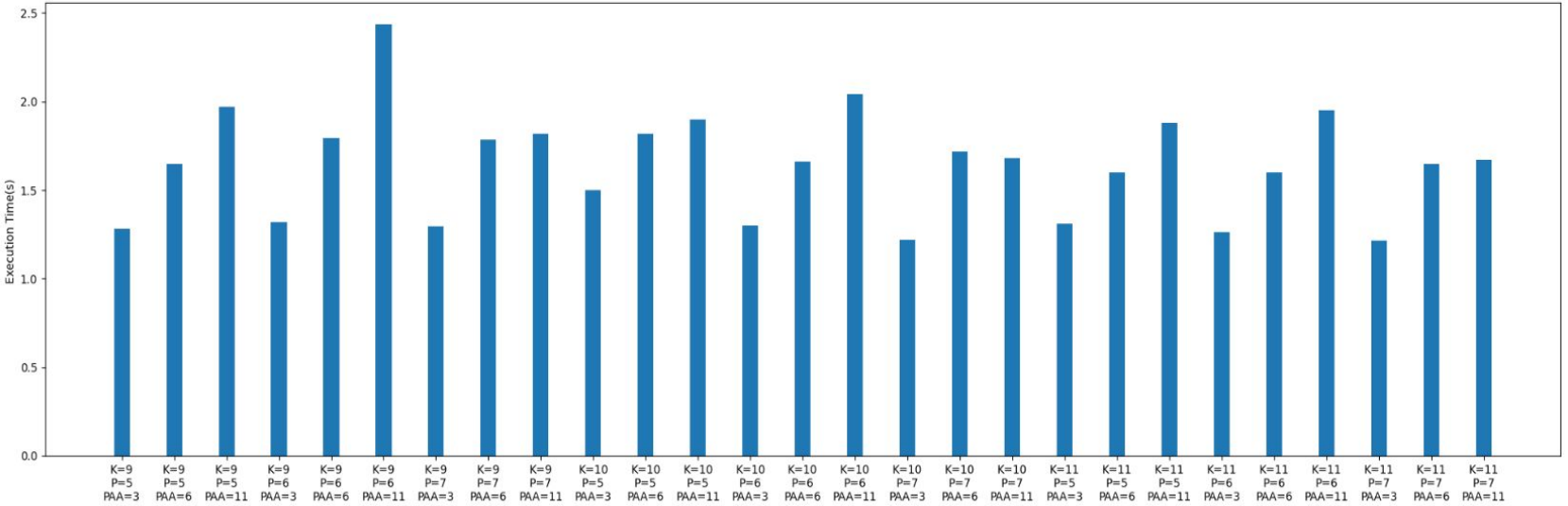
StackOverflow Questions Count Time Series: consist of count of various questions of specific libraries for each month. Used for time analysis and Instant Value Loss analysis.

MLTollsStackOverflow:

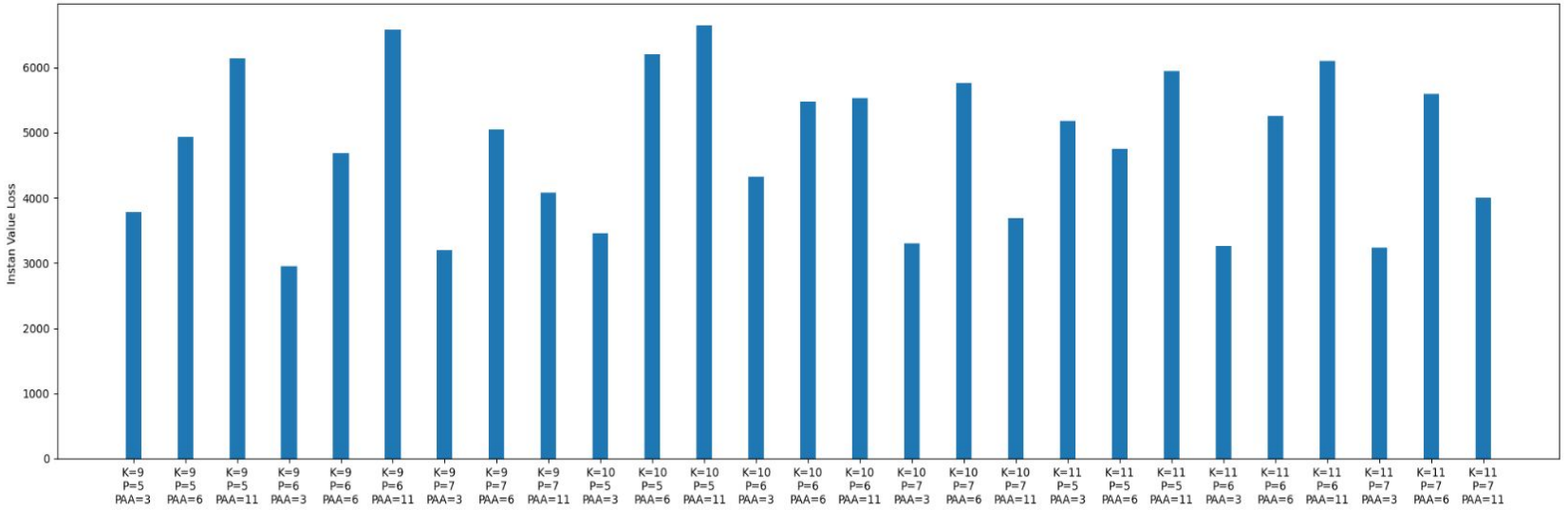
- 82 columns
- 133 rows

https://www.kaggle.com/datasets/aishu200023/stackindex

Analysis - Execution Time



Analysis - Instant Value Loss



Results of Analysis

Best solutions:

- Execution Time: **K=11, P=7, PAA=3**
- Instant Value Loss: **K=9, P=6, PAA=3**

As almost always, the best case is about halfway between the two metrics and looking carefully at the two graphs and comparing them we can see that in this case **K=10, P=7, PAA=3** there is the best compromise for this test run.

Thanks for the attention

View full project here.

- Alex Di Stefano (S4638131)
- Christian Stingone (S4644983)