Contribution ID: **524** Type: **Sectional talk**

Impact of anonymization level on the resilience of dataset clusters in big data

Thursday 10 July 2025 16:45 (15 minutes)

Personal data anonymization is an important step in dataset preprocessing, especially when dealing with sensitive information. However, the impact of this process on the quality of clustering remains poorly understood. The presented study analyzes the impact of different anonymization techniques affect the clustering results. The experimental part of the work is based on the application of ISODATA, maximin distance (Maximin) and hierarchical clustering algorithms to different datasets. The results obtained demonstrate that, for a limited number of features, depersonalization contributes to a clearer separation of the resulting clusters and while preserving the overall data structure and its trend. These findings indicate a future problem with the risks of personal data de-identification.

Author: Mr ДИК, Александр

Co-authors: BOGDANOV, Alexander (St. Petersburg University St. Petersburg, Russia); Mr SAVKOV, Egor (Consern Avrora scientific and production association jsc); KIYAMOV, JASUR; SHCHEGOLEVA, Nadezhda (St. Petersburg University St. Petersburg, Russia); Dr ДИК, Геннадий

Presenter: Mr ДИК, Александр

Session Classification: Round Table on the Areas of Work of the SPbSU-JINR Joint Scientific and Educational Laboratory