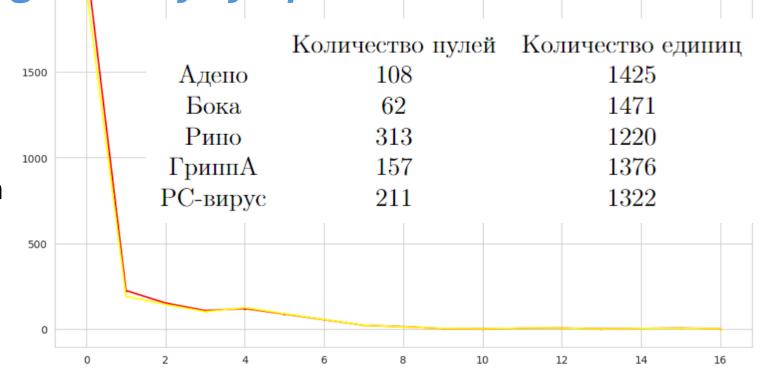# Machine learning and data mining methods for problems of virus type recognition by symptoms

Motivation: Assessing the statistical significance of symptoms and calculating prior and posterior probabilities of the presence of each type of virus

Problems:

- Class imbalance
- Presence of non-binary features and weak connections

	Количество нулей	Количество единиц
Адено	108	1425
Бока	62	1471
Рино	313	1220
ГриппА	157	1376
РС-вирус	211	1322

Solutions:

- Application of statistical tests
- Usage of ensemble machine learning methods