**Zadatak za završni ispit**

Zadatak se sastoji u dizajniranju i implementaciji algoritma za detekciju izuzetaka (outlier) i/ili noviteta (novelty) u skupu tekstualnih dokumenata.

Algoritam treba da bude zasnovan na odgovarajućoj modifikaciji ideje izložene u:

1. Odjeljku ***3.9 Methods for High Degrees of Similarity*** iz knjige *Mining of Massive Datasets* autora *J. Leskovec, A. Rajaraman i J. D. Ullman*.
2. Odjeljku ***4.3 Filtering Streams*** iz knjige *Mining of Massive Datasets* autora *J. Leskovec, A. Rajaraman i J. D. Ullman*.

Uporediti performanse predloženih algoritama sa algoritmima izloženim u materijalu <https://scikit-learn.org/stable/modules/outlier_detection.html>.

Jezik za implementaciju je Python 3.

Pored implementacije, neophodno je pripremiti seminarski rad koji treba da sadrži definiciju problema koji se rješava, detaljan opis vašeg algoritma i opis spovedenih eksperimenata zajedno sa diskusijom o ostvarenim rezultatima.

Studenti sa parnim brojem indeksa rade temu pod rednim brojem 1., studenti sa neparnim brojem indeksa rade temu sa rednim brojem 2.

Rješenje se šalje na mejl analitickaobradapodataka@gmail.com najkasnije do 15. juna do 12 sati. Usmena odbrana projekta biće u terminima završnog ispita.