

Zadatak za završni ispit

Zadatak se sastoji u dizajniranju i implementaciji sistema za detekciju izuzetaka (outlier) i noviteta (novelty) u domenu tekstualnih dokumenata.

Algoritam treba da bude zasnovan na odgovarajućoj modifikaciji ideje izložene u odjeljku **3.9 Methods for High Degrees of Similarity** iz knjige *Mining of Massive Datasets* autora *J. Leskovec, A. Rajaraman i J. D. Ullman*.

Potrebno je implementirati scenarije za treniranje i testiranje sistema.

Potrebno je uporediti performanse sistema 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.

Najviše dva studenta mogu da rade u grupi. Projekat se usmeno brani u terminima završnog ispita.