

[Inspiratron.org](https://inspiratron.org) - [Natural language processing, machine learning and cybersecurity](#)

Klasifikacija teksta i analiza sentimenta rečenica na srpskom jeziku

by Nikola Milošević - Sunday, January 27, 2013

<https://inspiratron.org/blog/2013/01/27/klasifikacija-teksta-i-analiza-sentimenta-recenica-na-srpskom-jeziku/>

Klasifikacija teksta i analiza sentimenta predstavljaju jedan od problema oblasti koja se naziva procesuiranje prirodnog jezika. Ova oblast spaja nekoliko nauka poput računarstva, veštačke inteligencije i lingvistike. Problem koji rešava klasifikacija teksta je klasifikacija nepoznatih dokumenata u korisnički preddefinisane klase. Analiza sentimenta je specijalan slučaj klasifikacije teksta u kome postoje dve klase - pozitivna i negativna. Na ovaj način ovo predstavlja binarnu klasifikaciju. Međutim postojanje 2 klase ne olakšava posao, jer se pri ovoj klasifikaciji radi sa ljudskim emocijama i osećanjima. Kako problem klasifikacije teksta, tako i problem analize sentimenta se rešavaju uz pomoć algoritama supervizovanog mašinskog učenja. Supervizovano mašinsko učenje je učenje gde se algoritmu pokazuju prvo rešeni primeri, na osnovu kojih on uči kako da rešava problem. Prilično slično je odvojenju deteta u školu. Postoji takođe nesupervizovano učenje, gde se algoritam pušta da radi odmah na novim podacima. Međutim klasifikacija teksta se ne može rešiti pomoću takvog algoritma.

Danas stanje sa procesuiranjem srpskog jezika je prilično loše. Jako malo ljudi se bavi rešavanjem problema iz ove oblasti, pa je tako na nedavnoj konferenciji u hotelu Hajat povodom obeležavanja svetskog dana jezičkih tehnologija rečeno da srpskom jeziku pretilo virtuelno izumiranje (http://www.b92.net/tehnopolis/aktuelno.php?yyyy=2012&mm=10&nav_id=655860)

Ovim radom želeo sam malo da doprinesem oblasti i srpskom jeziku u virtuelnom svetu. Tako u radu je opisan rad na trenutno verovatno najpreciznijem stemmer-u za srpski jezik, kao i klasifikatoru sentimenta, koji radi uz pomoć Naive Bayes algoritma. Pored Naive Bayes algoritma, klasifikator sentimenta koristi još neke obrade, poput obrada negacija, stop termine, takođe koristi Laplace-ovo poravnanje

(Laplace smoothing). Pomoću Laplasovog poravnanja se rešava problem pojave novih reši u novim dokumentima, koje nisu bile poznate u trening setu i određivanje njihovog sentimenta.

Rad u celini možete preuzeti i pročitati sa sledećeg linka:

[Analiza sentimenta rešenica na srpskom jeziku \(master rad\) - Nikola Milošević](#)

All rights reserved and copyrighted by inspiratron.org and Nikola Milosevic