

Izveštaj

Prepoznavanje emocija u govoru

Aleksandra Zarija EE49/2014 aleksandra.zarija@gmail.com

Milica Milošević EE47/2015 mima.milosevic96@gmail.com

I UVOD

Prepoznavanje emocija u govoru je pored prepoznavanja facijalnih ekspresija jedno od obećavajućih modela automatskog prepoznavanja ljudskih emocija. Iako emotivno stanje ne menja lingvistički sadržaj, predstavlja značajan faktor u ljudskoj komunikaciji, jer daje mogućnost povratne informacije. Tokom proteklih godina, prepoznavanje emocija pronalazi primenu u razvoju sigurnosnih sistema, u psihijatrijskoj medicini, ali i u razvoju video igrice, a u budućnosti i u komunikaciji čovek-mašina.

U ovom radu su predstavljena dva pristupa problemu prepoznavanja emocija u govoru, kako bismo pronašli potencijalni standard za klasifikaciju emocija u tehničkom smislu. Posmatrani skup emocija je diskretan i sadrži ljutnju, strah, tugu, radost i neutralno stanje koje predstavlja odsustvo ostalih emocija. Prvi pristup podrazumeva obučavanje modela na više osoba, a testiranje na osobi čiji podaci nisu viđeni do sad, dok se drugi obučava i testira na jednoj istoj osobi.

II BAZA PODATAKA

Baza podataka se zasniva na audio snimcima 6 osoba od kojih su 3 muškog i 3 ženskog pola. Za svaku od 6 osoba postoje audio snimci svake od 5 emocija: ljutnja, strah, tuga, radost i neutralno stanje koje predstavlja odsustvo ostalih emocija. Audio snimci se razlikuju po dužini trajanja (kraće rečenice u trajanju od 2 sekunde i duže u trajanju od 4 sekunde).

Pomoću programa *Praat*, izvukli smo obeležja audio snimaka koje dalje obrađujemo. Od tako dobijenih podataka su izračunate srednja vrednost, standardna devijacija, varijansa, interkvartilni opseg, dinamički opseg i trend promene piča. Dalja razmatranja se vrše nad gore navedenim parametrima u formi matrice.

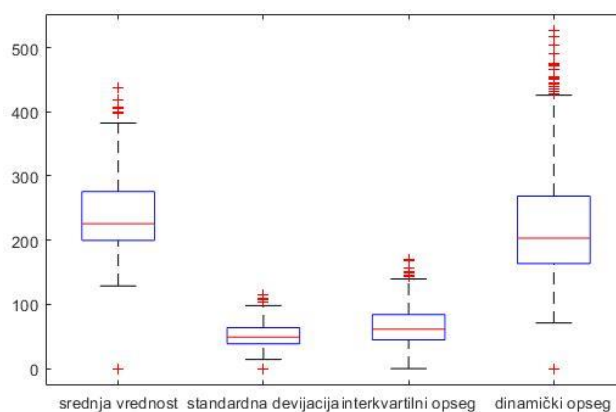
III ANALIZA PODATAKA

Podaci na osnovu kojih je vršena klasifikacija su srednja vrednost, standardna devijacija, varijansa, interkvartilni opseg i dinamički opseg. Oni se razlikuju u zavisnosti od emocije. Pošto je varijansa kvadratni stepen standardne devijacije, informacija o varijansi je sadržana u standardnoj devijaciji, pa je varijansa izostavljena sa boxplotova.

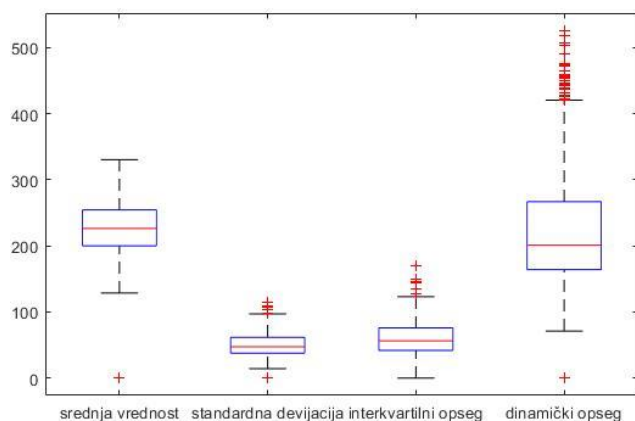
Ako se posmatraju boxplotovi srednjih vrednosti, uočava se da ljutnja i radost imaju šire interkvartilne opsege u odnosu na ostale emocije. Takođe kod neutralnog stanja je prisustvo *outlier*-a minimalno.

Dinamički opsezi standardne devijacije su kod svih emocija uski, a medijane se nalaze oko 50. Samo kod radosti nema prisustva *outlier*-a.

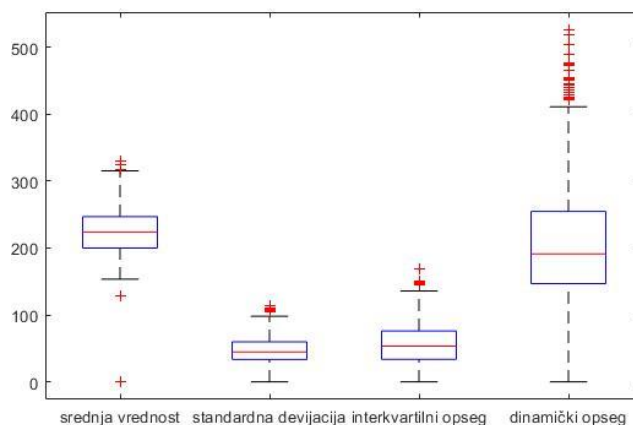
Interkvartilni opseg dinamičkog opsega je samo kod radosti oko 120, a dinamički opseg je širi u odnosu na ostale emocije i samim tim ima i manje *outlier*-a.



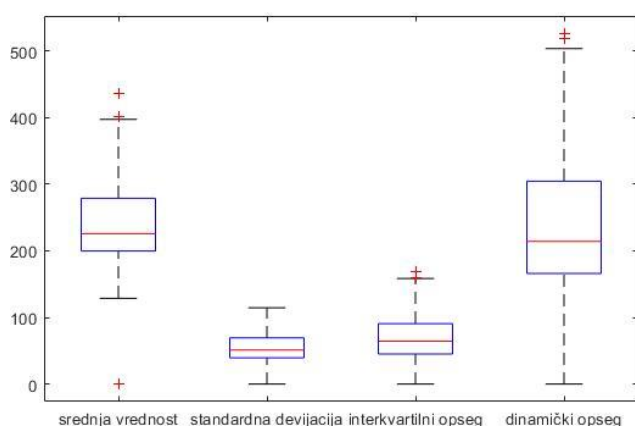
Slika 1. Boxplotovi podataka za ljutnju



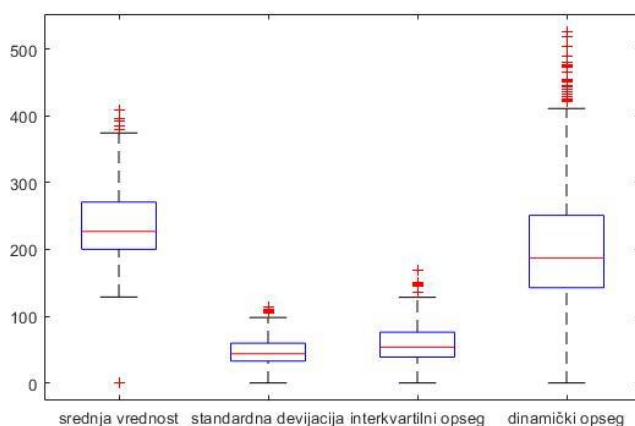
Slika 2. Boxplotovi podataka za neutralno stanje



Slika 5. Boxplotovi podataka za tugu



Slika 3. Boxplotovi podataka za radost



Slika 4. Boxplotovi podataka za strah

Takođe se posmatra promena piča tj da li raste, opada ili stagnira u zavisnosti od emocije. Za dati skup podataka trend promene piča stagnira kada su u pitanju ljutnja i radost, dok opada kod neutralnog stanja, straha i tuge.

Tabela 1. Trend promene piča po emocijama

	pič
ljutnja	→
neutralno stanje	↘
radost	→
strah	↘
tuga	↘

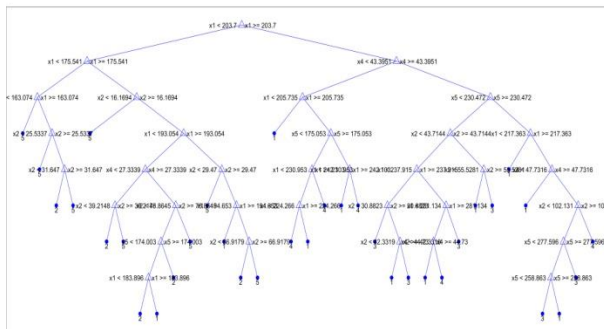
IV OPIS KLASIFIKATORA

KNN klasifikacija se bazira na metodi najbližih suseda i predstavlja najintuitivniji metod za klasifikaciju neobeleženih primeraka na osnovu sličnosti sa primerima iz trening skupa. Jedino što ovaj klasifikator zahteva jeste celobrojni broj suseda k , skup labeliranih primeraka (trening skup) i metriku za merenje bliskosti dva vektora obeležja.

Izuzetno je jednostavan za implementaciju, veoma je prilagodljiv zbog korišćenja lokalnih informacija i neosetljiv je na složenost stvarnih raspodela po pojedinim klasama. Međutim, zahteva obima izračunavanja, zbog čega zauzima velik skladišni prostor, a sa smanjenjem trening skupa performanse mu značajno opadaju.

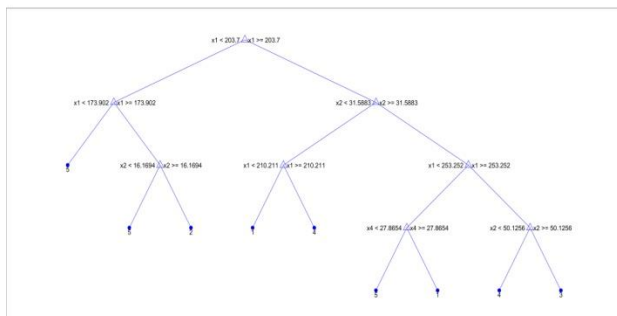
Stablo odlučivanja je intuitivan algoritam koji se koristi kao vizuelno i analitičko oruđe za klasifikaciju. Ovaj klasifikator podrazumeva deljenje prostora predikcije na više zasebnih, nepreklapajućih regiona. Pošto se ovakva podela prostora grafički može predstaviti kao drvo, analogija se prenosi i na ostale termine. Svaki unutrašnji čvor stabla predstavlja mesto odluke o klasifikaciji, a svaka grana ishod testa. Staze od korena do lista predstavljaju pravila klasifikacije.

Pored toga što su vrlo slični ljudskoj logici, prednost ovih algoritama je što se lako kombinuju sa drugim tehnikama. Nažalost, proračuni mogu postati veoma kompleksni naročito ako su vrednosti u trening skupu nepouzdanе ili ako je mnogo ishoda povezano.



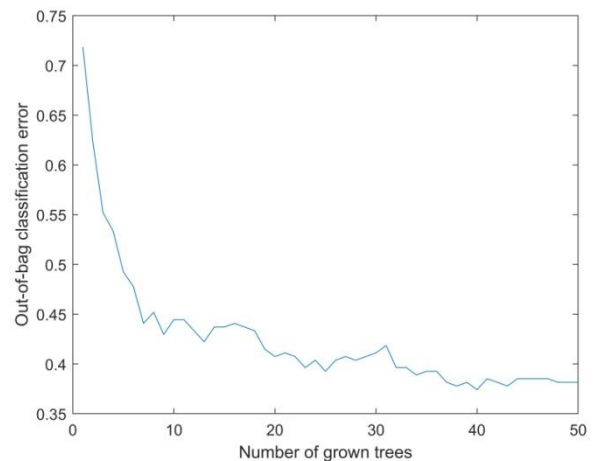
Slika 6. Stablo odlučivanja bez ograničenja broja čvorova

Ovakav algoritam se lako preobuči, jer teži da po klasama rasporedi sve podatke iz trening skupa. Ovaj problem se može rešiti regulisanjem broja čvorova što donosi višestruke prednosti. Na taj način ne samo da se smanji kompleksnost stabla, već se mogu očekivati i bolji rezultati na testu.



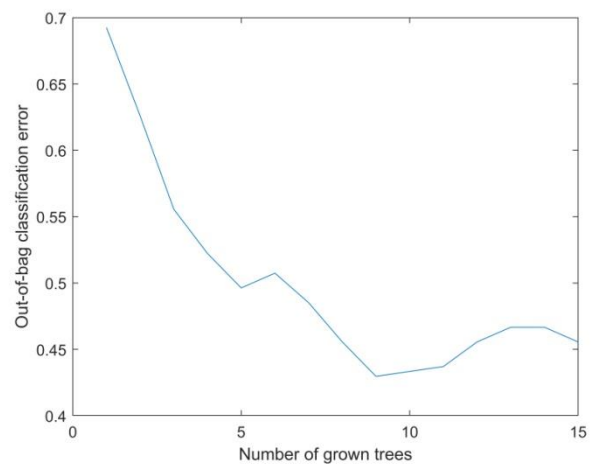
Slika 7. Stablo odlučivanja ograničeno na maksimalno 11 čvorova

Problem preobučavanja kod stabla odlučivanja se može rešiti i na drugi način. Uzimanjem uzastopnih uvezanih (*bootstrapped*) uzoraka iz jedinstvene baze podataka, pravi se n različitih baza na kojima se model trenira. Predikcije dobijene za svako drvo se pamte i na kraju se odluka o klasifikaciji donosi na osnovu većine rezultata. Otuda je ovaj algoritam dobio ime TreeBagging algoritam. Agregacijom stabala odlučivanja na ovaj način se postižu bolji rezultati u odnosu na klasifikator koji koristi samo jedno stablo odlučivanja, ali samo do neke mere. Naime, veliki broj stabala će davati bolje rezultate samo do određene granice nakon koje rezultati počinju da stagniraju, tj daljim povećanjem broja stabala se ne dobijaju ništa bolji rezultati.



Slika 8. Prikaz *TreeBagger* algoritma za 50 stabala

U tom slučaju je potrebno ograničiti broj stabala na onaj broj na kojem se na grafiku primećuje da funkcija prestaje da opada da bi sprečili bespotrebno povećavanje složenosti.



Slika 9. Prikaz *TreeBagger* algoritma za 15 stabala

V ANALIZA REZULTATA

Slede matrice konfuzije za sva 3 klasifikatora koje odgovaraju prvom pristupu u kojem se obučavanje vrši nad 5 osoba iz baze, a testira na jednoj. Da se uočiti da sva 3 klasifikatora imaju problem sa prepoznavanjem neutralnog stanja. Iako svi klasifikatori mešaju ljutnju i radost, taj procenat je kod kNN najviši sa preko 30%. Kod kNN klasifikatora i *TreeBagger*-a, strah je u više od 23% slučajeva svrstan pod neutralno stanje, dok je pri korišćenju stabla obluka čak više puta klasifikovan kao tuga nego ispravno. Najlošije predviđanje tuge se primećuje kod kNN klasifikatora gde je u 27% slučajeva klasifikovana kao neutralno stanje.

Tabela 2. Matrica konfuzije dobijena primenom kNN klasifikatora

kNN klasifikator		Predviđene vrednosti				
		ljutnja	neutralno	radost	strah	tuga
Prave vr.	ljutnja	164	56	110	20	10
	neutralno	77	54	88	68	73
	radost	125	87	120	19	9
	strah	59	90	21	164	26
	tuga	45	98	17	48	152

Tabela 3. Matrica konfuzije dobijena primenom stabla odlučivanja

Stablo odlučivanja		Predviđene vrednosti				
		ljutnja	neutralno	radost	strah	tuga
Prave vr.	ljutnja	163	48	84	23	42
	neutralno	74	21	31	78	156
	radost	106	82	112	21	39
	strah	75	46	25	106	108
	tuga	30	53	22	50	205

Tabela 4. Matrica konfuzije dobijena primenom *Tree bagger* algoritma

Tree Bagger		Predviđene vrednosti				
		ljutnja	neutralno	radost	strah	tuga
Prave vr.	ljutnja	144	55	116	33	12
	neutralno	54	39	96	80	91
	radost	98	80	153	12	17
	strah	52	83	20	175	30
	tuga	19	69	33	50	189

Slede matrice konfuzije za sva 3 klasifikatora koje odgovaraju drugom pristupu u kojem se obučavanje vrši nad jednom osobom iz baze i testira na istoj. U ovom slučaju se dobijaju mnogo bolji rezultati. Neutralno stanje koje je u prvom pristupu imalo nizak stepen prepoznavanja, na ovaj način je ispravno klasifikovano u više od 80% slučajeva kod sva 3 klasifikatora. Ipak klasifikatori i dalje teško prave distinkciju između ljutnje i radosti i to čak sa istim procentom od 30%. Takođe, kNN klasifikator je ispravno svrstao radost u samo 50% slučajeva. Razlika između tuge i neutralnog stanja je veće, procenat pogrešne klasifikacije je spušten na ispod 15%.

Tabela 5. Matrica konfuzije dobijena primenom kNN klasifikatora

kNN klasifikator		Predviđene vrednosti				
		ljutnja	neutralno	radost	strah	tuga
Prave vr.	ljutnja	123	3	44	9	1
	neutralno	8	150	2	6	14
	radost	61	8	93	17	1
	strah	23	13	14	123	7
	tuga	6	28	0	10	136

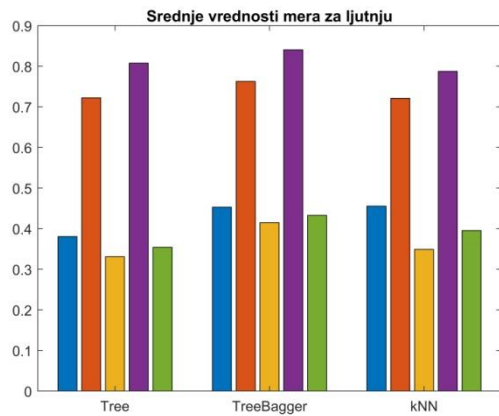
Tabela 6. Matrica konfuzije dobijena primenom stabla odlučivanja

Stablo odlučivanja		Predviđene vrednosti				
		ljutnja	neutralno	radost	strah	tuga
Prave vr.	ljutnja	104	6	58	10	2
	neutralno	5	143	2	6	24
	radost	52	5	102	19	2
	strah	14	10	19	123	14
	tuga	2	21	2	17	138

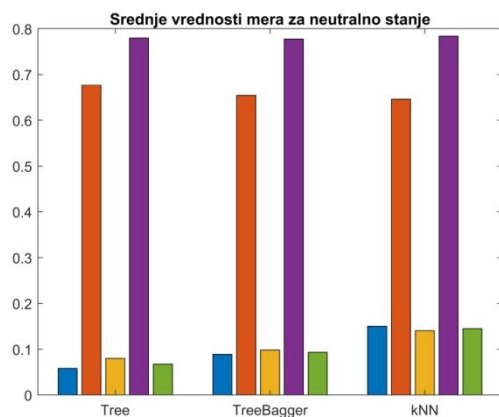
Tabela 7. Matrica konfuzije dobijena primenom *Tree bagger* algoritma

Tree Bagger		Predviđene vrednosti				
		ljutnja	neutralno	radost	strah	tuga
Prave vr.	ljutnja	108	5	54	11	2
	neutralno	6	147	1	4	22
	radost	51	5	103	20	1
	strah	12	12	21	126	9
	tuga	2	21	0	8	149

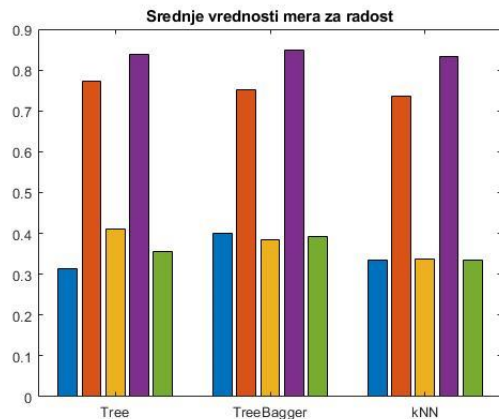
Pored matrica konfuzije, rezultate posmatramo i pomoću srednjih vrednosti mera validacije. Slede slike na kojima se nalazi uporedni prikaz preciznosti (plavo), tačnosti (narandžasto), odziva (žuto), specifičnosti (ljubičasto) i F1 norme (zeleno) za sva 3 klasifikatora.



Slika 10. Srednje vrednosti mera za ljudnju za prvi pristup

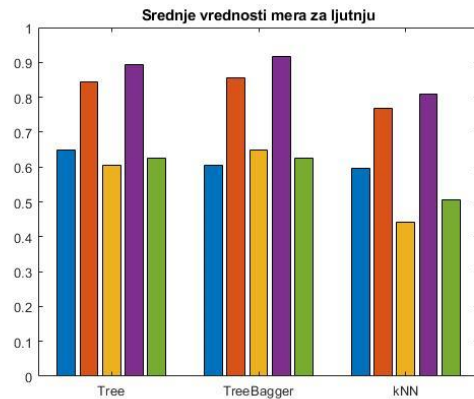


Slika 11. Srednje vrednosti mera za neutralno stanje za prvi pristup

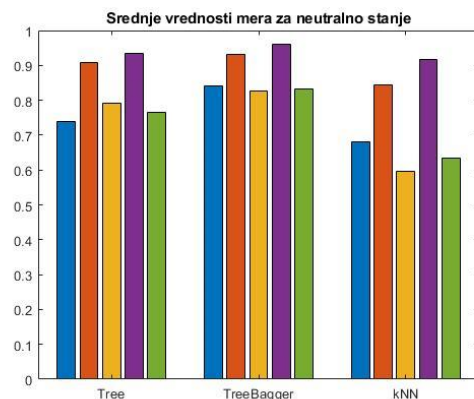


Slika 12. Srednje vrednosti mera za radost za prvi pristup

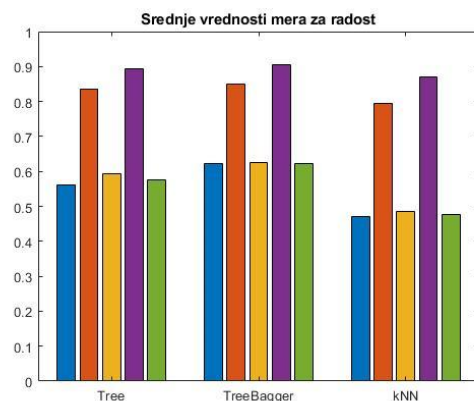
Radi lakšeg poređenja rezultata, slede prikazi mera za drugi pristup. Primećujemo značajan porast preciznosti i odziva, a samim tim i F1 norme u drugom slučaju. Specifičnost je u svim slučajevima visoka sa preko 80%. Pобољшanje rezultata se najbolje primećuje na primeru neutralnog stanja



Slika 13. Srednje vrednosti mera za ljudnju za drugi pristup



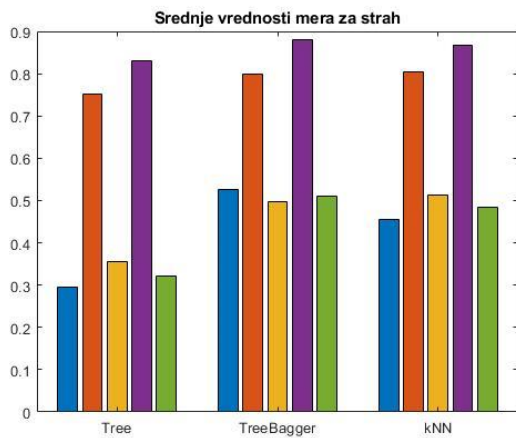
Slika 14. Srednje vrednosti mera za neutralno stanje za drugi pristup



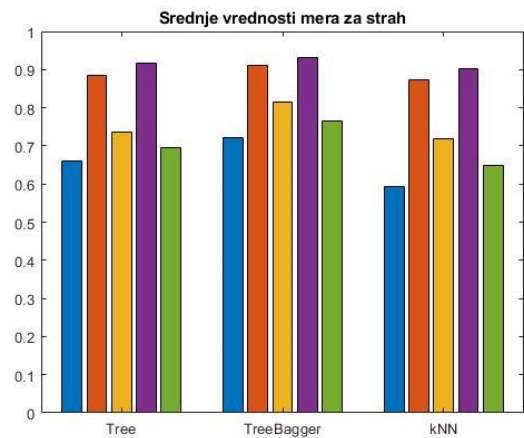
Slika 15. Srednje vrednosti mera za radost za drugi pristup

Za skoro sve emocije najbolje rezultate daje *Tree Bagger* klasifikator, sa izuzetkom slučaja klasifikacije neutralnog stanja u prvom pristupu gde kNN ima bolje mere.

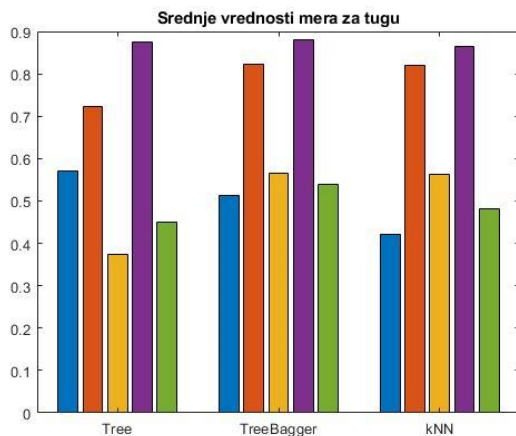
Drugi pristup je omogućio da se stopa prepoznavanja skoro duplira u procentima, što je u slučaju *TreeBagger* algoritma rezultiralo u skoku sa 40% za prvi pristup, na 73% za drugi pristup.



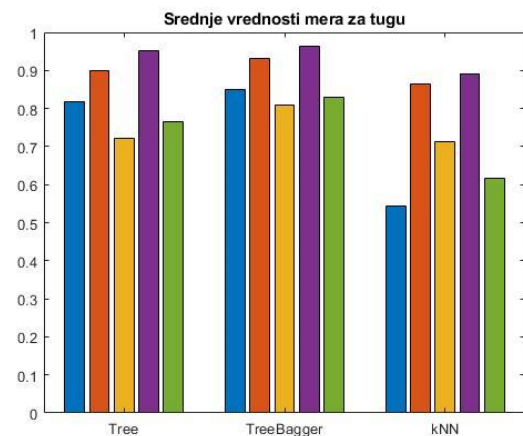
Slika 16. Srednje vrednosti mera za strah za prvi pristup



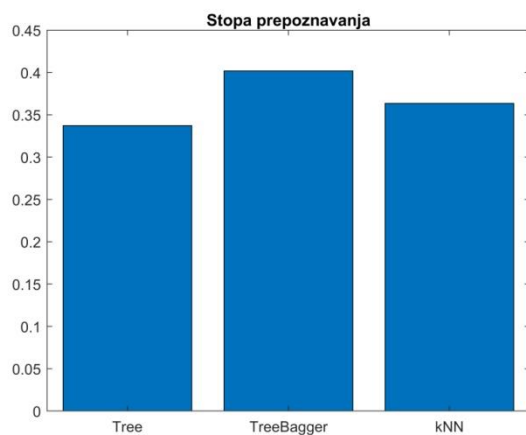
Slika 19. Srednje vrednosti mera za strah za drugi pristup



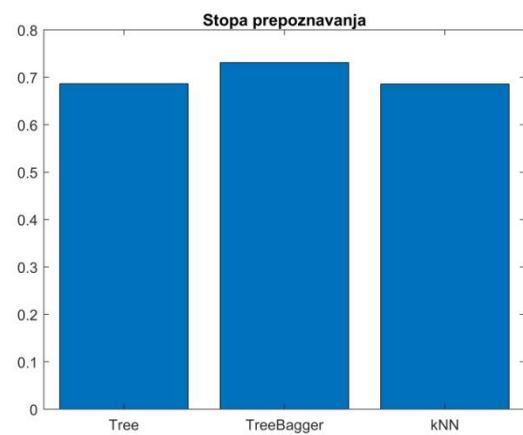
Slika 17. Srednje vrednosti mera za tugu za prvi pristup



Slika 20. Srednje vrednosti mera za tugu za drugi pristup



Slika 18. Stopa prepoznavanja emocija za prvi pristup



Slika 21. Stopa prepoznavanja emocija za drugi pristup

VI ZAKLJUČAK

Posmatranjem mera uspešnosti testa, primećuje se da su prilikom primene pristupa tokom kojeg se testira na nepoznatoj osobi preciznost i odziv, a samim tim i F1 norma u proseku 30% niži nego kada su trening i test skup vezani za istu osobu. Pri tome najuspešniji je *TreeBagger* algoritam sa stopom prepoznavanja od 40% za prvi i 73% za drugi pristup.

Kada su u pitanju matrice konfuzije, vidi se da oba pristupa mešaju ljutnju i radost. Problem sa klasifikacijom radosti je izraženiji u drugom slučaju. U prvom slučaju postoji problem sa klasifikovanjem neutralnog stanja, što je u drugom pristupu regulisano, a stopa prepoznavanja neutralnog stanja je sa 10% skočila na 80%. U prvom pristupu takođe klasifikatori loše raspoznaju tugu, strah što je u drugom slučaju minimizovano.

U oba slučaja klasifikatori mešaju emocije na očekivan način ako se u obzir uzmu sličnosti ustanovljene između ljutnje i radosti tj tuge, straha i neutralnog stanja. Ipak drugi pristup daje bolje rezultate, koji se eventualno mogu dodatno poboljšati povećanjem seta podataka.