

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Filip Jurec
Mariela Markovinović
Marta Novak
Vedrana Raguž
Ivana Žeger

**Analiza političkih preferencija
studentata metodama strojnog
učenja**

Zagreb, 2020.

Ovaj rad izrađen je na Zavodu za radiokomunikacije pod vodstvom doc. dr. sc. Darija Bojanjca i predan je na natječaj za dodjelu Rektorove nagrade u akademskoj godini 2019./2020.

SADRŽAJ

1. Uvod	1
2. Priprema anketnih pitanja i analiza provedenih anketa	3
2.1. Kandidati	3
2.2. Fokus grupa	6
2.3. Analiza ankete	7
3. Metode rada	12
3.1. Analiza glavnih komponenti	12
3.1.1. Implementacija analize glavnih komponenti pomoću program- skog jezika Python	16
3.2. Slučajne šume	20
3.2.1. Važnost značajki	21
4. Političke preferencije studentske populacije	23
4.1. Analiza prve ankete	23
4.2. Analiza druge ankete	38
4.3. Važnost značajki	48
5. Predviđanje rezultata drugog kruga izbora	52
6. Zaključak	57
Literatura	59
Popis tablica	62
Popis slika	64

1. Uvod

Odraz pojma „politika“ duboko je ukorijenjen u ljudsku svakodnevicu. Osnovna općenita definicija pojma najčešće podrazumijeva cjelokupno teorijsko i praktično djelovanje usmjereno unapređenju, vođenju i organiziranju države. Doticaj s tim višeznačnim terminom ima svaki pojedinac koji u nekom društvenom odnosu izražava svoje interese. Politika je smatrana umijećem upravljanja u različitim oblicima društvenog djelovanja i društvenih institucija preko kojih ljudi nastoje ostvariti svoje društvene težnje. Politika formalno označava i sam institucionalni poredak države, ali i aktivnost donošenja odluka u svim sferama društvenog života [1].

Kompleksna priroda politike pogoduje činjenici da se utjecaj njenog funkcioniranja tiče pojedinaca, ali i društva kao cjeline. Ako odluke donosi izabrana vlast, politika pogađa društvo u cjelini. Predsjednički izbori u Republici Hrvatskoj primjer su aktivnosti koja utječe na sve državljane. Predsjednik Republike Hrvatske predstavlja i zastupa Republiku Hrvatsku u zemlji i inozemstvu; brine se za redovito i usklađeno djelovanje te za stabilnost državne vlasti; odgovara za obranu neovisnosti i teritorijalne cjelovitosti Republike Hrvatske [2].

Predsjednički izbori 22. prosinca 2019. godine poslužili su kao inspiracija za osmišljavanje vrlo interdisciplinarnog projekta. Početna zamisao bila je definirati i analizirati slijed potrebnih radnji pomoću kojih bi od mišljenja ljudi iz okoline bilo moguće predvidjeti pobjednika izbora. Na taj način bi se teorijska i praktična znanja stečena tijekom obrazovanja mogla primijeniti u stvarnom svijetu. Promišljanje o navedenoj zamisli dovelo je do zaključka kako bi za takvu djelatnost bila potrebna velika količina ispitanika čija bi se mišljenja preoblikovala u podatke koje je moguće analizirati. Opisivanjem modela koji traže uzorke u velikoj količini podataka, objašnjavaju pozadinu tih podataka, odnosno izvlačenjem znanja iz podataka bavi se strojno učenje. Strojno učenje iznimno je interdisciplinarno područje računarske znanosti; spoj statistike, umjetne inteligencije, teorije vjerojatnosti, teorije informacije, baza

podataka i mnogih drugih domena. Odlučeno je u svrhu analize odrediti i koristiti prigodne metode statistike i strojnog učenja.

Početni korak projekta bila je analiza samih kandidata i njihovih programa kako bi se izvukle najbitnije karakteristike potrebne za daljnji rad. U suradnji s Filozofskim fakultetom Sveučilišta u Zagrebu organizirana je fokus grupa čiji je cilj bio okupiti mali, ali reprezentativan broj mladih sudionika čija će diskusija ukazati na smjer daljnjeg istraživanja. Analizom razgovora napravljena je sistematizacija stavova. Temeljem rezultata izrađena je anketa čija bi rješenja daljnjom analizom trebala poslužiti za željenu predikciju. Tvrdnje ankete bile su podijeljene u nekoliko tematskih cjelina kako bi se stavovi ispitanika analizirali iz različitih kutova. Ankete za prvi i drugi izborni krug objavljene su u velikim internetskim grupama te je prikupljen značajan broj rješenja. Nužno je bilo otkriti odgovarajući model strojnog učenja koji će pomoći u daljnjem radu. Nadalje, bilo je potrebno utvrditi način interpretacije odgovora pojedinog ispitanika, što konkretni odgovori otkrivaju o političkom opredjeljenju ispitanika te povezanost političkog opredjeljenja i samog izbora kandidata. Na kraju, bilo je potrebno utvrditi što velika količina takvih podataka govori o konačnom izboru dijela populacije koji je ispunio anketu te u kojoj se mjeri računalna predikcija i stvarnost poklapaju.

2. Priprema anketnih pitanja i analiza provedenih anketa

Predsjednički izbori održavaju se svakih pet godina od 1992. godine, kada je izabran prvi predsjednik Republike Hrvatske, Franjo Tuđman. Pravo biti biran za predsjednika Republike Hrvatske imaju hrvatski državljani s navršenih 18 godina. Političke stranke i birači predlažu kandidate na temelju pravovaljano prikupljenih 10 000 potpisa birača. Svaki birač može svojim potpisom podržati samo jednog kandidata [3]. Vlada Republike Hrvatske raspisuje izbore koji se obavljaju u roku od najmanje 30, a najviše 60 dana prije isteka mandata trenutnom predsjedniku Republike [4]. Predsjednik je izabran ako osvoji preko 50% glasova birača koji su glasovali. Ako niti jedan kandidat ne osvoji dovoljan broj glasova, izbor se ponavlja za 14 dana te u njemu sudjeluju dva kandidata s najviše osvojenih glasova u prvome krugu.

Izbori za predsjednika održani 22. prosinca 2019. godine bili su sedmi predsjednički izbori u Republici Hrvatskoj. U prvome krugu sudjelovalo je jedanaest kandidata. To su: Nedjeljko Babić, Anto Đapić, Kolinda Grabar-Kitarović, Dario Juričan, Mislav Kolakušić, Dejan Kovač, Zoran Milanović, Dalija Orešković, Katarina Peović, Ivan Pernar i Miroslav Škoro. U tablici 2.1 prikazane su stranke, broj prikupljenih potpisa za kandidaturu, broj osvojenih glasova u prvome krugu te broj glasova prikazan u obliku postotka. Ankete koje su provodile brojne novine također su ukazivale na sličan rezultat. Uvijek bi dominirala tri kandidata: Zoran Milanović, Kolinda Grabar-Kitarović i Miroslav Škoro.

2.1. Kandidati

Pri stvaranju ankete, fokus je bio na kandidatima koji su prema prije provedenim anketama imali više od 5% glasova (Zoran Milanović, Kolinda Grabar-Kitarović, Miroslav Škoro, Mislav Kolakušić). Prvi korak bio je proučiti program i stavove kandidata.

Tablica 2.1: Kandidati [5]

	Stranka	Broj potpisa	Broj glasova	Postotak
Nedjeljko Babić	HSSČKŠ	14 000	3 014	0,15%
Anto Đapić	DESNO	13 000	4 001	0,21%
Kolinda Grabar-Kitarović	HDZ	230 000	507 628	26,65%
Dario Juričan	Neovisni kandidat	13 000	87 883	4,61%
Mislav Kolakušić	Neovisni kandidat	15 000	111 916	5,87%
Dejan Kovač	HSLŠ	16 000	18 107	0,95%
Zoran Milanović	SDP	78 000	562 783	29,55%
Dalija Orešković	Neovisni kandidat	11 000	55 163	2,89%
Katarina Peović	RF	14 000	21 387	1,12%
Ivan Pernar	SIP	15 000	44 057	2,31%
Miroslav Škoro	Neovisni kandidat	80 000	465 704	24,45%

Kolinda Grabar-Kitarović

Kolinda Grabar-Kitarović postala je prva predsjednica Republike Hrvatske pobijedivši dotadašnjeg predsjednika Ivu Josipovića. Svoj program za predsjedničke izbore predstavila je pod nazivom „Hrvatska zna i mora bolje“. U programu se zalaže za Hrvate u Bosni i Hercegovini kao jednakopravan narod. Smatra da bi preduvjet Srbiji za članstvo u Europskoj uniji trebalo biti razrješavanje pitanja nestalih u Domovinskom ratu. Također, ostat će glas protiv korupcije, rješavat će problem blokiranih, zagovarati jednake plaće, prava žena i majki. Predstavljajući program rekla je da obitelj treba biti jezgra hrvatskoga društva. Smatra da je nužno porezno rasterećenje te fleksibilizacija radnog vremena. Rečenicom „Samo mi sami možemo potpuno osiguravati svoj mir.“, naglasila je nužnost osnaživanja te ulaganja u hrvatsku vojsku i policiju.

Mislav Kolakušić

Svojim programom, Mislav Kolakušić donosi najviše promjena i obećanja. Želi povećanje ovlasti predsjednika, po uzoru na Francusku. Prema njegovu mišljenju, narod bira predsjednika te mu daju povjerenje što znači da on treba imati moć u svojim

rukama. Smatra da Hrvatskoj treba „reset“; drastične promjene, stvaranje gospodarstva te otpuštanje minimalno 20% zaposlenika državne uprave. Obećaje da će se boriti protiv korupcije te po uzoru na Rumunjsku osnovati antikorupcijsku agenciju i državnu poreznu agenciju koje bi provodile kontrolu protiv korupcije. Smatra da su promjene u pravosuđu neophodne, primjerice, povećanje plaća sucima te uvođenje pravednog izbora sudaca na način da tri bira predsjednik, tri građani i tri sabor. U programu navodi da želi uvođenje elektroničkog glasovanja za izbore i referendume. Smatra da je pogrešno što se stranke smiju financirati iz državnog proračuna. Želi uvesti financiranje stranke isključivo donacijama svojih članova.

Zoran Milanović

Bivši premijer svoju je kampanju podijelio na nekoliko bitnih stavki. Pod sloganom „Hrvatska jednakih“ zalaže se za Hrvatsku u kojoj svi imaju jednaku priliku i koja bi trebala štiti slabije. Smatra da predsjedniku ne trebaju veće ovlasti, već karakter. Borit će se za društvenu pravdu, jednakost i slobodu, prava slabijih i demokratske procese u politici. Sloganom „Hrvatska slobode“ naglašava da je Hrvatska zemlja u kojoj svatko ima priliku uspjeti, a posao predsjednika je štiti kako slobode građana tako i slobode medija. Treći dio programa pod nazivom „Moderna Hrvatska sa stavom“ zalaže se za Hrvatsku koja pažljivo bira partnere u svijetu te bira ono što je najbolje za nju, a to je moderna, progresivna i napredna Europa. Svjestan je ovlasti koje predsjednik ima te smatra da ne daje nerealna obećanja koja ne može ispuniti.

Miroslav Škoro

Miroslav Škoro, pod sloganom „Sad il' nikada!“ predstavio se kao kandidat iz naroda. Smatra da ovo nije Hrvatska za koju su se branitelji borili. Domovinski rat polazna je točka modernizacije Hrvatske. Želi da se riješi pitanje nestalih te da se vrati arhivi iz Srbije. U kampanji se zalaže za promjene ovlasti predsjednika, povećanje odgovornosti, mogućnost raspisivanja referenduma, sazivanja vlade i predlaganje kandidata za suce Ustavnog suda. Kritizira dvije vladajuće stranke koje se izmjenjuju te se zalaže za promjenu izbornog zakona. Želi da Hrvatska teži ravnopravnosti u Europskoj uniji te da traži potporu u obrani zajedničke granice. Posebice, zalaže se za monetarni suverenitet, kunu, a ne euro. Obećaje zalaganje za manjine, mogućnost elektronskog glasovanja te ulaganje u modernizaciju vojske.

2.2. Fokus grupa

Metoda fokus grupa jest kvalitativni oblik istraživanja koji uključuje grupnu diskusiju o nekoj zadanoj temi. Osnovni je cilj fokus grupe potaknuti dubinsku diskusiju kojom će se istražiti vrijednosti ili stavovi ispitanika prema nekom problemu ili temi, odnosno razumjeti i objasniti značenja, vjerovanja i kulturu koja utječe na osjećaje, stavove i ponašanja individua [6]. Metoda fokus grupe nametnula se kao optimalna metoda u svrhu kreiranja ankete te otkrivanja stavova mladih. U fokus grupama obično sudjeluje 6-12 sudionika, odnosno 8+/-2. U tome je slučaju grupa dovoljno mala da omogući svakom sudioniku da iznese svoj stav, a u isto vrijeme dovoljno velika da se razvije grupna dinamika [6]. Grupne diskusije obično traju između 90 i 120 minuta, a sudionike u raspravi vodi osposobljeni moderator. Rasprava se gradi od općega prema posebnome te postupno usmjerava na relevantno pitanje [7]. Diskusija se provodi prema prije pripremljenom nizu tema. U slučaju ovoga rada, u fokus grupi sudjelovalo je sedam osoba (studenata). Osobe su bile birane na način da se pokriju različiti politički stavovi i razmišljanja kod mladih ljudi. Pitanja su bila otvorenog tipa te se tako omogućila fleksibilnost prilikom odgovaranja. Moderator fokus grupe bila je izv. prof. dr. sc. Renata Geld s Filozofskog fakulteta.

Sudionici su iznosili mišljenje o osnovnim značajkama kandidata za koje će glasati, idealnom kandidatu za kojeg bi glasali, važnosti predsjednika i predsjedničkih izbora, općenitom izvršavanju građanske dužnosti, izvorima informacija, utjecaju okoline na političke stavove, kalkulaciji oko glasanja, programima kandidata, itd.

Analiza rezultata razgovora fokus grupe nije linearnog tipa, već se analitički procesi preklapaju kroz pet faza: upoznavanje, identifikaciju tematskog okvira, indeksiranje, mapiranje i interpretaciju. Proces je krenuo razmjenom i analizom detaljnih bilješki i transkribiranog audiozapisa istraživača - članova projektnog tima. Identifikacija tematskog okvira podrazumijevala je razvijanje kategorija i formiranje deskriptivnih izjava. Indeksiranje je uključivalo sažimanje izjava, sortiranje i usporedbe. Mapiranjem su se izjave iz originalnog konteksta stavile u prilagođeni, novi tematski sadržaj. Interpretacijom su doneseni zaključci na temelju pojedinog pitanja koji su kasnije korišteni pri formiranju tvrdnji za anketu. U sažetku, zaključeno je da ne postoji idealan predsjednički kandidat. Analizom su utvrđeni stavovi i razmišljanja mladih o temi predsjedničkih izbora. Spoznala su se pitanja koja su mladima bitna, kao i načini na koje se mladi informiraju o kandidatima.

2.3. Analiza ankete

Prilikom analize predsjedničkih izbora koristile su se dvije ankete. Prva je analizirala stavove glasača o pojedinim temama te pitanje za koga će glasovati u prvome krugu izbora. Druga anketa sadržavala je ista pitanja uz dodatna pitanja o izboru glasača u prvome krugu izbora te drugome krugu. Pitanja ankete nastala su kao rezultat iskazanih stavova mladih tijekom diskusije u fokus grupi te samih programa kandidata. Obje ankete formirane su u elektroničkom obliku preko Google obrazaca te su bile objavljene u različitim Facebook grupama kako bi ih ispunio što veći broj ljudi.

Stavovi glasača vrednovali su se odgovorima od 1 do 5:

- U potpunosti se ne slažem (1).
- Ne slažem se (2).
- Niti se slažem, niti se ne slažem (3).
- Slažem se (4).
- U potpunosti se slažem (5).

Anketna pitanja bila su podijeljena u nekoliko tematskih cjelina:

1. Stavovi o ulozi predsjednika:

- (a) Potrebno je povećati ovlasti predsjednika Republike Hrvatske.
- (b) Predsjednik može biti osoba bez prethodnog političkog iskustva.
- (c) Predsjedniku je u radu nužna potpora političkih stranaka.
- (d) Povećanje životnog standarda građana jedna je od osnovnih zadaća predsjednika.

2. Stavovi o stanju u Hrvatskoj:

- (a) Hrvatski državni aparat je prevelik.
- (b) Potrebno je smanjenje broja radnih mjesta u javnom i državnom sektoru za minimalno 20%.
- (c) Hrvatskoj je potrebna promjena izbornog zakona.

3. Karakteristike kandidata:

- (a) Predsjednika biram po principu manjeg zla.

- (b) Osobni dojam o kandidatu mi je bitniji od njegovog programa.
- (c) Predsjednik mora imati autoritarnu crtu.

4. Osobni stavovi:

- (a) Domovinski rat je ishodišna točka moderne Hrvatske.
- (b) Potrebno je povećati ulaganja u Oružane snage Republike Hrvatske.
- (c) Euro treba zamijeniti kunu kao sredstvo plaćanja u Hrvatskoj.
- (d) Predsjednik mora biti „kandidat iz naroda“.
- (e) Predsjednik Republike Hrvatske mora biti vjernik.
- (f) U Hrvatskoj postoje političari koji preuzimaju odgovornost za vlastite poteze.
- (g) Prihvatanje rodne ideologije i svih ostalih društvenih sloboda nužno je da bi Hrvatska postala moderna zajednica.
- (h) Pobačaj je pravo žene na izbor.
- (i) Cijepljenje djeteta treba biti slobodan izbor roditelja.
- (j) Marihuanu valja legalizirati.

5. Za koga ćete glasovati na izborima? Ponuđeni odgovori su:

- Nedjeljko Babić
- Anto Đapić
- Kolinda Grabar-Kitarović
- Dario Juričan
- Mislav Kolakušić
- Dejan Kovač
- Zoran Milanović
- Dalija Orešković
- Katarina Peović
- Ivan Pernar
- Miroslav Škoro
- Ne izlazim
- Križam listić.

Druga anketa imala je dodatna pitanja:

1. Za koga ste glasovali u drugom krugu predsjedničkih izbora 2015. godine?
 - Ivo Josipović
 - Kolinda Grabar-Kitarović
 - Nisam imao/la pravo glasa
 - Nisam glasovao/la
2. Za koga ćete glasovati na predstojećim parlamentarnim izborima?
 - Ispitanici su mogli unositi sami odgovore.
3. Za koga ćete glasovati na izborima?
 - Zoran Milanović
 - Kolinda Grabar-Kitarović
 - Ne izlazim na izbore
 - Križam listić

Prva anketa sastojala se od 24, a druga od 27 pitanja. Za ispunjavanje svake od anketa bilo je potrebno nekoliko minuta. Prvu anketu ispunila su 1672 ispitanika, a drugu njih 799. U uzorku su prevladavali studenti; 85,53% u prvoj anketi i 85,98% u drugoj. Prosječna dob ispitanika bila je 22,85 godine, a prevladavale su osobe ženskog spola. U anketi su sudjelovale osobe koje imaju minimalno 18 godina, odnosno mogućnost glasovanja na predsjedničkim izborima. Najmlađa osoba imala je 18 godina, a najstarija 88 godina.

Prilikom analize vodilo se računa o statističkoj značajnosti podataka. Statistička značajnost predstavlja tvrdnju da je mala vjerojatnost da su rezultati dobiveni iz podataka testiranjem ili eksperimentiranjem slučajni, odnosno da postoji određena pravilnost u podacima [8]. Jedan od problema statističke značajnosti je činjenica da se istraživanja vrše na uzorku populacije, ali ne na cijeloj populaciji. Veličina uzorka mora biti dovoljno velika. Uzorak mora biti reprezentativan i nepristran, tj. kvalitetno predstavljati populaciju. Potrebno je unaprijed odrediti vjerojatnost pogreške uzorkovanja koja postoji u bilo kojem testu koji ne uključuje ukupnu populaciju.

Zbog činjenice da među ispitanicima dominiraju mladi, nadalje u radu analizirani su samo odgovori ispitanika koji su u vrijeme ispunjavanja ankete imali između 18 i 28 godina. Nakon odgovarajuće prilagodbe skupa podataka, prva anketa sadržavala je 1551, a druga 727 odgovora.

Tablica 2.2: Osnovne informacije o uzorku

	Anketa 1	Anketa 2
Spol	53,8% ženski spol	60,25% ženski spol
	46,2% muški spol	39,75% muški spol
Dob	21,28 godina	21,47 godina
Studenti	92,01%	93,26%

Tablica 2.3: Distribucija odgovora na pitanja prve ankete

		1	2	3	4	5
A1N1	Potrebno je povećati ovlasti predsjednika Republike Hrvatske.	196 (12,6%)	221 (14,2%)	396 (25,5%)	403 (26%)	335 (21,6%)
A1N2	Predsjednik može biti osoba bez prethodnog političkog iskustva.	424 (27,3%)	385 (24,8%)	301 (19,4%)	253 (16,3%)	188 (12,1%)
A1N3	Predsjedniku je u radu nužna potpora političkih stranaka.	311 (20,1%)	317 (20,4%)	387 (25%)	388 (25%)	148 (9,5%)
A1N4	Povećanje životnog standarda građana jedna je od osnovnih zadaća predsjednika.	89 (5,7%)	156 (10,1%)	267 (17,2%)	433 (27,9%)	606 (39,1%)
A1N5	Hrvatski državni aparat je prevelik.	14 (0,9%)	42 (2,7%)	294 (19%)	324 (20,9%)	877 (56,5%)
A1N6	Potrebno je smanjenje broja radnih mjesta u javnom i državnom sektoru za minimalno 20%.	69 (4,4%)	132 (8,5%)	319 (20,6%)	372 (24%)	659 (42,5%)
A1N7	Hrvatskoj je potrebna promjena izbornog zakona.	36 (2,3%)	121 (7,8%)	368 (23,7%)	387 (25%)	639 (41,2%)
A1N8	Predsjednika biram po principu manjeg zla.	207 (13,3%)	175 (11,3%)	224 (14,4%)	428 (27,6%)	517 (33,3%)
A1N9	Osobni dojam o kandidatu mi je bitniji od njegovog programa.	255 (16,4%)	448 (28,9%)	457 (29,5%)	268 (17,3%)	123 (7,9%)
A1N10	Predsjednik mora imati autoritarnu crtu.	116 (7,5%)	209 (13,5%)	347 (22,4%)	582 (37,5%)	297 (19,1%)
A1N11	Domovinski rat je ishodišna točka moderne Hrvatske.	374 (24,1%)	260 (16,8%)	417 (26,9%)	251 (16,2%)	249 (16,1%)
A1N12	Potrebno je povećati ulaganja u Oružane snage Republike Hrvatske.	476 (30,7%)	420 (27,1%)	379 (24,4%)	183 (11,8%)	93 (6%)
A1N13	Euro treba zamijeniti kunu kao sredstvo plaćanja u Hrvatskoj.	621 (40%)	270 (17,4%)	322 (20,8%)	172 (11,1%)	166 (10,7%)
A1N14	Predsjednik mora biti „kandidat iz naroda“.	233 (15%)	285 (18,4%)	448 (28,9%)	343 (22,1%)	242 (15,6%)
A1N15	Predsjednik Republike Hrvatske mora biti vjernik.	950 (61,3%)	180 (11,6%)	204 (13,2%)	104 (6,7%)	113 (7,3%)
A1N16	U Hrvatskoj postoje političari koji preuzimaju odgovornost za vlastite poteze.	361 (23,3%)	532 (34,3%)	338 (21,8%)	216 (13,9%)	104 (6,7%)
A1N17	Prihvatanje rodne ideologije i svih ostalih društvenih sloboda je nužno da bi Hrvatska postala moderna zajednica.	481 (31%)	320 (20,6%)	334 (21,5%)	152 (9,8%)	264 (17%)
A1N18	Pobačaj je pravo žene na izbor.	876 (56,5%)	193 (12,4%)	176 (11,3%)	91 (5,9%)	215 (13,9%)
A1N19	Cijepljenje djeteta treba biti slobodan izbor roditelja.	794 (51,2%)	294 (19%)	173 (11,2%)	114 (7,4%)	176 (11,3%)
A1N20	Marihuani valja legalizirati.	493 (31,8%)	326 (21%)	362 (23,3%)	139 (9%)	231 (14,9%)

Tablica 2.4: Distribucija odgovora na pitanja druge ankete

		1	2	3	4	5
A2N1	Potrebno je povećati ovlasti predsjednika Republike Hrvatske.	122 (16,8%)	131 (18%)	181 (24,9%)	176 (24,2%)	117 (16,1%)
A2N2	Predsjednik može biti osoba bez prethodnog političkog iskustva.	261 (35,9%)	167 (23%)	134 (18,4%)	106 (14,6%)	59 (8,1%)
A2N3	Predsjedniku je u radu nužna potpora političkih stranaka.	143 (19,7%)	119 (16,4%)	203 (27,9%)	184 (25,3%)	78 (10,7%)
A2N4	Povećanje životnog standarda građana jedna je od osnovnih zadaća predsjednika.	68 (9,4%)	93 (12,8%)	141 (19,4%)	198 (27,2%)	227 (31,2%)
A2N5	Hrvatski državni aparat je prevelik.	6 (0,8%)	19 (2,6%)	139 (19,1%)	153 (21%)	410 (56,4%)
A2N6	Potrebno je smanjenje broja radnih mjesta u javnom i državnom sektoru za minimalno 20%.	29 (4%)	44 (6,1%)	186 (25,6%)	179 (24,6%)	289 (39,8%)
A2N7	Hrvatskoj je potrebna promjena izbornog zakona.	22 (3%)	52 (7,2%)	151 (20,8%)	181 (24,9%)	321 (44,2%)
A2N8	Predsjednika biram po principu manjeg zla.	110 (15,1%)	64 (8,8%)	115 (15,8%)	203 (27,9%)	235 (32,3%)
A2N9	Osobni dojam o kandidatu mi je bitniji od njegovog programa.	123 (16,9%)	198 (27,2%)	226 (31,1%)	120 (16,5%)	60 (8,3%)
A2N10	Predsjednik mora imati autoritarnu crtu.	32 (4,4%)	49 (6,7%)	166 (22,8%)	299 (41,1%)	181 (24,9%)
A2N11	Domovinski rat je ishodišna točka moderne Hrvatske.	184 (25,3%)	128 (17,6%)	194 (26,7%)	115 (15,8%)	106 (14,6%)
A2N12	Potrebno je povećati ulaganja u Oružane snage Republike Hrvatske.	212 (29,2%)	201 (27,6%)	178 (24,5%)	86 (11,8%)	50 (6,9%)
A2N13	Euro treba zamijeniti kunu kao sredstvo plaćanja u Hrvatskoj.	250 (34,4%)	137 (18,8%)	165 (22,7%)	94 (12,9%)	81 (11,1%)
A2N14	Predsjednik mora biti „kandidat iz naroda“.	103 (14,2%)	147 (20,2%)	215 (29,6%)	150 (20,6%)	112 (15,4%)
A2N15	Predsjednik Republike Hrvatske mora biti vjernik.	460 (63,3%)	73 (10%)	95 (13,1%)	49 (6,7%)	50 (6,9%)
A2N16	U Hrvatskoj postoje političari koji preuzimaju odgovornost za vlastite poteze.	141 (19,4%)	242 (33,3%)	175 (24,1%)	116 (16%)	53 (7,3%)
A1N17	Prihvatanje rodne ideologije i svih ostalih društvenih sloboda je nužno da bi Hrvatska postala moderna zajednica.	248 (34,1%)	152 (20,9%)	146 (20,1%)	85 (11,7%)	96 (13,2%)
A2N18	Pobačaj je pravo žene na izbor.	404 (55,6%)	89 (12,2%)	82 (11,3%)	44 (6,1%)	108 (14,9%)
A2N19	Cijepljenje djeteta treba biti slobodan izbor roditelja.	388 (53,4%)	106 (14,6%)	90 (12,4%)	61 (8,4%)	82 (11,3%)
A2N20	Marihuani valja legalizirati.	226 (31,1%)	147 (20,2%)	192 (26,4%)	52 (7,2%)	110 (15,1%)

Tablica 2.5: Odgovori na pitanje „Za koga ćete glasovati (ste glasovali) u prvome krugu predsjedničkih izbora?“ iz prve i druge ankete

Kandidati	Broj glasova u prvoj anketi	Broj glasova u drugoj anketi
Nedjeljko Babić	2 (0,1%)	0 (0%)
Anto Đapić	3 (0,2%)	3 (0,4%)
Kolinda Grabar-Kitarović	98 (6,3%)	44 (6,1%)
Dario Juričan	225 (14,5%)	77 (10,6%)
Mislav Kolakušić	166 (10,7%)	51 (7%)
Dejan Kovač	59 (3,8%)	19 (2,6%)
Zoran Milanović	246 (15,9%)	158 (21,7%)
Dalija Orešković	115 (7,4%)	41 (5,4%)
Katarina Peović	50 (3,2%)	19 (2,6%)
Ivan Pernar	59 (3,8%)	22 (3%)
Miroslav Škoro	270 (17,4%)	126 (17,3%)
Ne izlazim na izbore	136 (8,8%)	144 (19,8%)
Križam listić	122 (7,9%)	23 (3,2%)

Tablica 2.6: Odgovori na pitanje „Za koga ste glasovali na predsjedničkim izborima 2015. godine?“

Odgovori	Broj glasova
Ivo Josipović	137 (18,8%)
Kolinda Grabar-Kitarović	86 (11,8%)
Nisam glasovao/la	116 (16%)
Nisam imao/la pravo glasa	388 (53,4%)

Tablica 2.7: Odgovori na pitanje „Za koga ćete glasovati na predsjedničkim izborima?“

Odgovori	Broj glasova
Kolinda Grabar-Kitarović	138 (19%)
Zoran Milanović	380 (52,3%)
Ne izlazim na izbore	80 (17,7%)
Križam listić	129 (17,7%)

3. Metode rada

3.1. Analiza glavnih komponenti

Na kompleksnost analize i obrade velikih skupova podataka uvelike utječe broj značajki. Kako bi se skupovi s većim brojem značajki uspješno interpretirali, potrebno ih je svesti na manji broj faktora korištenjem metoda korelacije. Pri tome se nastoji sačuvati što više dostupnih podataka. Iako postoje mnoge metode razvijene u tu svrhu, analiza glavnih komponenti (engl. *Principal Component Analysis, (PCA)*) je jedna od češće korištenih [9].

Analiza glavnih komponenti, također poznata kao Karhunen–Loève ili Hotellingova transformacija, neparametarska je metoda čijim se procesom iz velikih skupova podataka izdvajaju ključne informacije [10]. Osmislio ju je Karl Pearson, a kasnije neovisno razvio statističar Harold Hotelling [11]. Koristi se u brojnim područjima analize – od neuroznanosti do računalne grafike [10]. Cilj ove analize je smanjiti dimenzionalnost skupova podataka s većim brojem značajki kao i minimizirati gubitak informacija [9]. To se postiže stvaranjem novih varijabli koje međusobno nisu u korelaciji. Novonastale varijable ne definiraju se unaprijed nego ovise o izvornom skupu podataka. Može ih se definirati kao linearne funkcije varijabli izvornog skupa podataka. One se u daljnjoj analizi smatraju glavnim komponentama ili faktorima. Dakle, faktori su uglavnom nezavisni podskupovi koji se sastoje od međusobno povezanih varijabli, odnosno značajki. Njihovo interpretiranje ovisi o značenju kombinacija promatranih varijabli koje su u visokoj korelaciji s faktorom [12]. Interpretacija faktora uspješnija je u slučaju velike korelacije između promatranih varijabli i faktora, dok pritom te iste varijable nisu u korelaciji s preostalim faktorima.

Analiza glavnih komponenti primjenjuje se putem odabira značajki, pripreme korelacijske matrice, ekstrakcije skupa faktora iz korelacijske matrice, određivanja broja faktora, rotiranja faktora u svrhe povećanja interpretabilnosti te interpretacije rezultata

[12]. Ovom se metodom analizira cjelokupna varijanca u promatranim varijablama. Varijanca je mjera koja se definira kao srednja vrijednost odstupanja podatka od aritmetičke sredine svih podataka iz skupa [11].

Primjena analize glavnih komponenti sastoji se od dvaju osnovnih nedostataka. Za razliku od metoda poput regresijske analize, diskriminantne funkcijske analize ili logističke regresije – analiza glavnih komponenti nema vanjski kriterij pomoću kojeg se provodi testiranje rješenja. Stoga se može zaključiti kako zbog nepostojanja opisanog vanjskog kriterija nije moguće provjeriti ispravnost konačnog rješenja. Dodatno, pri koraku rotiranja faktora postoji mogućnost beskonačnog broja rotacija. Rotacije odgovaraju jednakom iznosu varijance izvornih podataka s različito definiranim faktorima [12]. Konačan odabir rotacija ovisi isključivo o osobnoj procjeni interpretabilnosti. Može se zaključiti kako provedbom ove analize ne postoji mogućnost određivanja jedinstvenog rješenja, već ih je beskonačan broj.

Analiza glavnih komponenta jedna je od metoda faktorske analize. Faktorska analiza naziv je za skup metoda čiji je cilj smanjenje broja manifestnih varijabli među kojima postoji povezanost ili korelacija. Njenom primjenom dobije se uvid u strukturu međusobne povezanosti značajki [13]. Faktorska analiza dijeli se na dva osnovna tipa: eksploratornu i konfirmatornu. Svrha eksploratorne analize opis je i sumiranje podataka putem zajedničkog grupiranja varijabli koje su u međusobnoj korelaciji. Konfirmatorna faktorska analiza mnogo je sofisticiranija tehnika koja se koristi u naprednim stadijima istraživačkog procesa za testiranje hipoteze o latentnim procesima [12]. Jednostavnije rečeno, eksploratorna analiza povezana je s razvijanjem, a konfirmatorna s testiranjem hipoteze.

Temeljni pojmovi vezani za razumijevanje i provedbu analize glavnih komponenta uključuju:

- opažanu korelacijsku matricu, odnosno korelacijsku matricu promatranih varijabli,
- reproduciranu korelacijsku matricu, odnosno korelacijsku matricu glavnih komponenti,
- rezidualnu korelacijsku matricu koja opisuje razliku između dviju matrica iznad i
- rotaciju faktora, odnosno postupak povećanja interpretabilnosti rješenja bez

mijenjanja matematičkih svojstava u osnovi rješenja [12].

Rotacija faktora proces je povećanja interpretabilnosti rješenja bez mijenjanja matematičkih svojstava u osnovi rješenja [12]. Ona može biti ortogonalna ili kosa. U slučaju da su svi faktori međusobno nepovezani – rotacija je ortogonalna. Suprotno tome, rotacija je kosa ako su faktori u korelaciji.

Prvi faktor ili glavna komponenta je linearna kombinacija promatranih varijabli kojom se omogućuje maksimalno razdvajanje subjekta zajedno s maksimiziranjem varijance komponentnih vrijednosti [12]. Druga glavna komponenta nastaje iz rezidualnih korelacija. Svaka sljedeća komponenta nastaje ekstrahiranjem maksimalne varijabilnosti iz rezidualnih korelacija. U metodi analize glavnih komponenti prva komponenta ekstrahira najviše varijance, a posljednja najmanje. Također, svaka je komponenta ortogonalna na prethodno ekstrahiranu komponentu. Korištenje opisanih glavnih komponenti uvelike olakšava tumačenje rezultata u drugim analizama [12].

Poput svih tipova analize podataka, prilikom analize glavnih komponenti javljaju se određena ograničenja. Osjetljiva je na veličine korelacije i slučajeve među kojima postoji veliki skok, odnosno razlika. Nedostatak podataka također uzrokuje probleme. Dodatno, koeficijenti korelacija često su manje pouzdani kada se utvrđuju iz manjeg broja uzoraka [12]. S obzirom na to da je veza između parova varijabli linearna, uspješnost analize glavnih komponenti opada u slučaju degradacije linearnosti.

Tablicom 3.1 prikazane su informacije o osnovnim matricama koje se koriste pri analizi glavnih komponenti. Svaka matrica opisana je nazivom, oznakom i veličinom. Veličine matrica prikazane su putem oznaka:

- p , koja predstavlja broj varijabli i
- m , koja se odnosi na broj glavnih komponenti.

Tablica 3.1: Matrice korištene primjenom analize glavnih komponenti

Naziv	Oznaka	Red matrice
Korelacijska matrica	R	$p \times p$
Matrica svojstvenih vrijednosti	L	$m \times m$
Matrica svojstvenih vektora	V	$p \times m$
Matrica uzoraka / opterećenja	A	$p \times m$

Za ispravnu pripremu korelacijske matrice potrebno je poznavati pojmove pozitivne dijagonale i dijagonaliziranja matrica. Pozitivna dijagonala uključuje elemente matrice od njenog gornjeg lijevog do donjeg desnog kuta. Dijagonaliziranje je proces transformacije gdje se u novonastaloj matrici elementi pozitivne dijagonale sastoje od brojeva, dok se preostala mjesta sastoje izričito od nula. U analizi glavnih komponenti, brojevi na pozitivnoj dijagonali predstavljaju varijancu iz korelacijske matrice. Ona se postiže jednadžbom

$$\mathbf{L} = \mathbf{V}'\mathbf{R}\mathbf{V}. \quad (3.1)$$

Matrica svojstvenih vektora, \mathbf{V} , kvadratna je matrica za koju vrijedi da su vektori matrice linearno nezavisni. Matrica svojstvenih vrijednosti, \mathbf{L} , dijagonalna je matrica za koju vrijedi da je zbroj svojstvenih vrijednosti jednak ukupnoj varijanci standardiziranih varijabli te da se svojstvene vrijednosti nižu od najveće prema najmanjoj [13]. Matrica svojstvenih vektora množi se s transponiranom matricom. Rezultat je matrica identiteta čiji su elementi pozitivne dijagonale jedinice. Preostala mjesta sastoje se od nula. Dakle, matrica identiteta postiže se jednadžbom

$$\mathbf{I} = \mathbf{V}'\mathbf{V}. \quad (3.2)$$

Promjenom zapisa jednadžbe (3.1) dobije se formula za korelacijsku matricu, odnosno

$$\mathbf{R} = \mathbf{V}\mathbf{L}\mathbf{V}'. \quad (3.3)$$

Dakle, korelacijska matrica postiže se množenjem matrica svojstvenih vrijednosti i odgovarajućih svojstvenih vektora [12]. Dodatno, matrica svojstvenih vrijednosti zapisuje se pomoću drugog korijena. Taj postupak rezultira jednadžbom

$$\mathbf{R} = \mathbf{V}\sqrt{\mathbf{L}}\sqrt{\mathbf{L}}\mathbf{V}'. \quad (3.4)$$

Matrica uzoraka, \mathbf{A} , računa se pomoću jednadžbe

$$\mathbf{A} = \mathbf{V}\sqrt{\mathbf{L}}. \quad (3.5)$$

Tada se $\sqrt{\mathbf{L}}\mathbf{V}'$ označava s \mathbf{A}' . Uvrštavanjem novonastalih oznaka dobije se dodatna jednadžba za korelacijsku matricu

$$\mathbf{R} = \mathbf{A}\mathbf{A}'. \quad (3.6)$$

Shodno tome, korelacijska matrica može se dobiti pomoću informacija o matrici uzoraka/opterećenja, odnosno množenjem matrice uzoraka s njom transponiranom matricom. Obje matrice zapravo su kombinacije svojstvenih vektora i drugog korijena

svojtvenih vrijednosti [12].

Pravi uvid u strukturu međusobnih odnosa manifestnih varijabli često nije moguć samo putem glavnih komponenata [13]. Tada je potrebno koristiti rotacije, odnosno transformacije glavnih komponenata u svrhe postizanja jednostavne faktorske strukture. Pri tome se povećava interpretabilnost, odnosno maksimiziraju visoke i minimiziraju niske korelacije. Metoda koja se za to najčešće koristi jest *varimax*. Njen je cilj maksimizirati varijancu opterećenja faktora povećanjem visokih opterećenja i snižavanjem niskih [12]. Za to služi transformacijska matrica Λ , odnosno matrica sinusa i kosinusa kuta ψ . Prikaz opisane matrice Λ vidljiv je u jednadžbi

$$\Lambda = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix}. \quad (3.7)$$

Nerotirana matrica opterećenja množi se s transformacijskom matricom čime se dobije rotirana matrica opterećenja [12]. Jednadžba opisanog postupka je

$$A_{\text{nerotirano}}\Lambda = A_{\text{rotirano}}. \quad (3.8)$$

3.1.1. Implementacija analize glavnih komponenti pomoću programskog jezika Python

Analiza glavnih komponenti služi za prepoznavanje uzoraka u skupu podataka kao i predočenje tih podataka na način da se istaknu njihove sličnosti i razlike [14]. Njenim procesom se iz skupa varijabli na temelju nereducirane korelacijske matrice izračunavaju dimenzije koje su međusobno linearno nezavisne te se nazivaju glavnim komponentama. Analiza glavnih komponenti jedna je od jednostavnih metoda za redukciju podataka čijim procesom nastaje novi koordinatni sustav sa smanjenim brojem dimenzija u odnosu na izvorni skup podataka. Poput mnogih drugih multivarijatnih metoda, ova analiza nije bila široko prihvaćena niti korištena do pojave računala [11]. Analiza glavnih komponenti izvrstan je alat za analizu skupa s velikim brojem podataka. Primarne ciljeve analize glavnih komponenti može se podijeliti u:

- odabir najbitnijih informacija iz skupa podataka,
- kompresiju veličine skupa podataka uz zadržavanje informacija koje se smatraju bitnim,
- pojednostavljenje opisa skupa podataka,
- analizu strukture varijabli i

- reduciranje podataka smanjenjem broja dimenzija bez većeg gubitka informacija [11].

Dva su načina smanjenja broja dimenzija – selekcija i ekstrakcija značajki. Za uspješnu selekciju potreban je početni skup značajki iz kojega se odabire željeni podskup. Za razliku od selekcije, pri ekstrakciji je potrebno dobiti informacije iz skupa značajki pomoću kojih se omogućuje konstrukcija novog potprostora značajki.

Za uspješnu implementaciju opisane analize koristi se programski jezik Python, a pri smanjenju dimenzija rabi se metoda ekstrakcije značajki. Koraci implementacije ove analize se mogu podijeliti na:

1. skaliranje početnog skupa podataka,
2. konstrukciju kovarijacijske matrice,
3. pronalazak svojstvenih parova dekompozicijom kovarijacijske matrice,
4. sortiranje svojstvenih vrijednosti,
5. smanjenje dimenzionalnosti novog potprostora značajki,
6. konstrukciju projekcijske matrice W i
7. pretvorbu početnog d -dimenzionalnog skupa podataka u k -dimenzionalni potprostor značajki.

Prvi korak obuhvaća standardiziranje podataka iz početnog skupa kako bi svi bili jednako skalirani. Pri tome se koristi metoda *StandardScaler* dostupna u Python biblioteci *scikit-learn*. Ona obavlja standardiziranje značajki uklanjanjem srednje vrijednosti te skaliranjem na varijancu. Oznaka za varijancu je σ^2 , a za kovarijancu σ . Nakon skaliranja obavlja se konstrukcija kovarijacijske matrice. Njena veličina je $d \times d$ gdje oznaka d predstavlja dimenziju, odnosno broj značajki. Jednadžba za kovarijancu je

$$\sigma_{jk} = \frac{1}{n} \sum_{i=1}^n \left(x_j^{(i)} - \mu_j \right) \left(x_k^{(i)} - \mu_k \right), \quad (3.9)$$

gdje su j i k značajke, a μ_j i μ_k njihove srednje vrijednosti. Prikaz kovarijacijske matrice s m značajki je

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \cdots & \sigma_{34} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \sigma_{m3} & \cdots & \sigma_m^2 \end{bmatrix}. \quad (3.10)$$

Stupanj pozitivnosti ili negativnosti kovarijance ovisi o razlici vrijednosti svakog para značajki od srednje vrijednosti. Kovarijanca će biti pozitivnija ako postoji veći broj parova vrijednosti koji se razlikuju od srednjih vrijednosti u istom smjeru. Svojstvo istog smjera označava da obje vrijednosti imaju zajedničko odstupanje, koje može biti pozitivno ili negativno. Suprotni smjer prikazuje da jedna vrijednost ima negativno, a druga pozitivno odstupanje od srednje vrijednosti. Sukladno tome, negativnost kovarijance raste za svaki par vrijednosti koji se razlikuje od srednjih vrijednosti u suprotnim smjerovima.

Nad dobivenom kovarijacijskom matricom provodi se dekompozicija kojom se dobiju svojstveni parovi. Oni se sastoje od svojstvenih vektora (engl. *eigenvectors*) i svojstvenih vrijednosti (engl. *eigenvalues*). Svojstveni vektori predstavljaju glavne komponente, a svojstvene vrijednosti njihove veličine. Odnos svojstvenog vektora i vrijednosti može se prikazati jednadžbom

$$\Sigma \mathbf{v} = \lambda \mathbf{v}. \quad (3.11)$$

Svojstveni vektor prikazan je oznakom \mathbf{v} , a svojstvena vrijednost skalarnom oznakom λ . Implementacija dekompozicije kovarijacijske matrice obavlja se pomoću Python funkcije `np.linalg.eig(covariance_matrix)`. Glavne komponente potrebno je poredati prema svojstvenim vrijednostima, počevši od najveće. Nakon sortiranja glavnih komponenti (svojstvenih vektora), provodi se smanjenje dimenzionalnosti. Ovisno o željenom broju dimenzija novog potprostora značajki, k , odabire se prvih k glavnih komponenti. Pri tome se odabiru glavne komponente s najvećom varijancom koja se računa korištenjem jednadžbe

$$\sigma^2 = \frac{\lambda_j}{\sum_{j=1}^d \lambda_j}. \quad (3.12)$$

U šestom koraku provodi se konstruiranje projekcijske matrice \mathbf{W} koja se sastoji od k svojstvenih vektora ili glavnih komponenti. Veličina matrice \mathbf{W} je $d \times k$, gdje oznaka d predstavlja broj dimenzija početnog skupa podataka. Pomoću dobivene projekcijske matrice i početnog d -dimenzionalnog skupa podataka \mathbf{X} obavlja se posljednji korak analize glavnih komponenti. Jednadžba za pretvorbu početnih informacija u interpretabilniji k -dimenzionalni potprostor značajki je

$$\mathbf{X}' = \mathbf{XW}. \quad (3.13)$$

3.2. Slučajne šume

Ovisno o podacima koji su dostupni, odabire se algoritam strojnog učenja. U slučaju ovoga rada, korištene su metode nadziranog učenja. Tijekom učenja, podaci se dijele na podatke za treniranje, validaciju i testiranje. Podaci za treniranje koriste se za „učenje“ algoritma i podešavanje parametara. Validacijski skup podataka osigurava nepristrano ocjenjivanje i prilagodbu hiperparametara samog modela kako bi model na testnim podacima davao najbolje moguće rezultate [15]. Testni skup podataka služi za procjenu efikasnosti finalnog modela s nikad viđenim podacima.

Jedno predočavanje svih uzoraka naziva se epohom. Učenje može biti pojedinačno (za svaki primjer podešavaju se hiperparametri) ili grupno (cijela epoha u jednoj iteraciji) [16]. Podaci koji se koriste kao ulaz formirani su u značajke (engl. *features*). U ovom slučaju značajka predstavlja pitanje koje je postavljeno u anketi. Smanjivanje dimenzionalnosti značajki, odnosno odabir značajki koje najviše doprinose točnosti modela, ključan je korak za analizu podataka. U takvom se procesu uklanjaju značajke koje su redundantne ili ne doprinose informacijama.

Od različitih modela strojnog učenja koji su se pokazali prikladnima za provođenje analize odgovora anketa, kao najuspješniji okarakteriziran je algoritam slučajnih šuma. Slučajne šume (engl. *random forest, (RF)*) jedna su od metoda strojnog učenja koja se može se koristiti za klasifikaciju, ali i za regresiju. Počeci ove metode datiraju u 1995., kada ju je predložio Tim Kam Ho [17]. Slučajne šume, kakve se danas koriste, nastale su 2001., a razvio ih je Leo Breiman [18]. Ovaj algoritam za raspoznavanje ne koristi samo jedan klasifikator, već kreira mnoštvo klasifikatora u obliku stabla odluke (engl. *decision tree*). Svako stablo trenira se na podskupu ulaznih primjera za učenje te sudjeluje u kreiranju konačnih izlaza. Kako bi se klasificirao primjer, on prolazi svakim kreiranim stablom. Svako stablo daje klasifikaciju i „glasuje“ za klasu koju je dodijelio primjeru. Šuma tada odabire klasu koja je dobila najviše glasova (vrši se klasifikacija).

Kao i u algoritmu stabla odlučivanja, koristi se CART algoritam (engl. *Classification and Regression Trees*) za kreiranje jednog stabla. Odabir koje će se ulazne značajke koristiti vrši se pomoću pohlepnog algoritma s ciljem da funkcija gubitka u svakom trenutku bude minimalna [19]. Stablo se prestaje „graditi“ ako je predefiniiran kriterij zaustavljanja, kao što je maksimalna dubina stabla ili broj primjeraka za učenje koji će biti dodijeljeni svakom čvoru lista stabla. Pohlepni algoritam iznimno je jednostavan algoritam kombinatorne optimizacije. Za svaku točku odluke odabire se izbor

koji je u tome trenutku najbolji. U ovom slučaju odabire se ona varijabla za koju je funkcija gubitka minimalna. Za klasifikacijski problem u pohlepnom algoritmu koristi se gini indeks. Gini indeks je suma kvadratnih vjerojatnosti za svaki primjer značajke:

$$Gini = 1 - \sum_{i=1}^n (p_i)^2, \quad (3.14)$$

gdje je n broj klasa u značajki. Algoritam uzima u obzir sva moguća grananja kako bi pronašao najbolje grananje za točnost modela.

Slučajne šume koriste neparametarsku metodu *bootstrap*. To je jedna od metoda ponovnog uzorkovanja podataka iz originalnog skupa podataka. Ovom metodom nastaju primjeri koji imaju dosta šuma. Glavni razlog za uvođenje šuma je smanjenje varijance i povećanje točnosti modela (teže dolazi do pretreniranosti).

Jedna trećina ukupnih podataka koristi se za ocjenjivanje rada klasifikatora. Ti podaci nazivaju se oob (engl. *out-of-bag data*) te služe za procjenu pogreške. Pogreška se računa kao broj stabala koja su pogrešno klasificirala oob.

Početna varijabla za odrediti jest n . Ona kazuje koliko će algoritam stabala kreirati. Iz skupa za učenje metodom *bootstrap* bira se N uzoraka koji čine skup primjera za uzgoj stabla. Taj se skup bira na slučajan način. Od M ulaznih značajki, izabire se njih m , na slučajan način. Veličina varijable m najčešće se bira na način da je $m \ll M$. U čvoru će biti izabrane one koje maksimiziraju informacijsku dobit. Varijabla m je stalna kroz kreiranje stabala. Svako stablo raste do svoje maksimalne veličine [18].

Cilj algoritma je i smanjiti korelaciju između stabala. Veća korelacija rezultira manjom točnošću modela. Točnost svakog stabla povećava točnost cijele šume. Što je varijabla m veća, to su korelacija i točnost stabla veći [18]. Iz tog razloga potrebno je pronaći optimalni m koji će povećati točnost pojedinog stabla, ali ih neće učiniti previše međusobno koreliranim.

3.2.1. Važnost značajki

Bitna funkcionalnost algoritma slučajnih šuma je i mogućnost mjerenja važnosti značajki (engl. *feature importance*). *Sklearn* biblioteka ima implementirane metode koje mjere važnost značajki gledajući koliko čvorovi stabala, koji koriste tu značajku, smanjuju nečistoću na svim stablima u šumi. Ona izračunava ovaj rezultat automatski za svaku značajku nakon treniranja i skalira rezultate, tako da je zbroj svih važnosti

jednak 1 [20].

Za određivanje važnosti značajke pr. N_1 , potrebno je prvo svakim stablom odrediti klasifikaciju oob primjera. Nakon toga se primjeri N_1 značajke slučajno permutiraju unutar oob skupa podataka i ponovno klasificiraju. Od zbroja točno klasificiranih početnih oob podataka za svako stablo oduzme se zbroj točno klasificiranih oob uzoraka nakon permutiranih vrijednosti atributa N_1 . Rezultat se usrednjava brojem stabala u šumi i naziva se važnost atributa N_1 [21].

4. Političke preferencije studentske populacije

U sklopu ovog rada provedena je analiza odgovora ispitanika s ciljem kreiranja modela glasača. Postavljeno je pitanje postoje li odgovori koji se vezuju za određenog kandidata, te može li se pomoću strojnog učenja predvidjeti za koga će netko glasovati na izborima temeljem njegovih odgovora u anketi. Za predviđanje su korištene dvije ankete opisane u poglavlju 2.3.

Na početku istraživanja, analizirani su odgovori te vizualno prikazani glasači pojedinih kandidata iz prvoga i drugoga kruga. U poglavlju 3.1 opisana je korištena metoda za analizu glavnih komponenti koja je bila primijenjena kako bi se dimenzionalnost podataka svela s 23 dimenzije na dvije dimenzije.

U drugom dijelu istraživanja pomoću algoritma slučajnih šuma analizirano je kako pojedino pitanje ankete (značajka) doprinosi predviđanju za koga će netko glasovati. Posljednji dio istraživanja donosi predviđanja ishoda drugog kruga na temelju prve ankete pomoću algoritma slučajnih šuma.

4.1. Analiza prve ankete

Tablica 4.1 prikazuje rezultate prve ankete u sklopu koje je prikupljen ukupno 1551 uzorak. Ispitanici su većinski pripadnici studentske populacije s prosjekom godina između 20 i 22. Prema rezultatima ankete, najveći broj glasova, ukupno 270, pripada Miroslavu Škori. Slijede ga Zoran Milanović s 246 i Dario Juričan s 225 glasova, dok ostali kandidati imaju značajno manje. Kandidati Nedjeljko Babić i Anto Đapić nisu uključeni u daljnju analizu zbog vrlo malog broja glasača. Takvi odgovori ne bi bili statistički značajni, tj. sveukupnom rezultatu doprinose s manje od 5%.

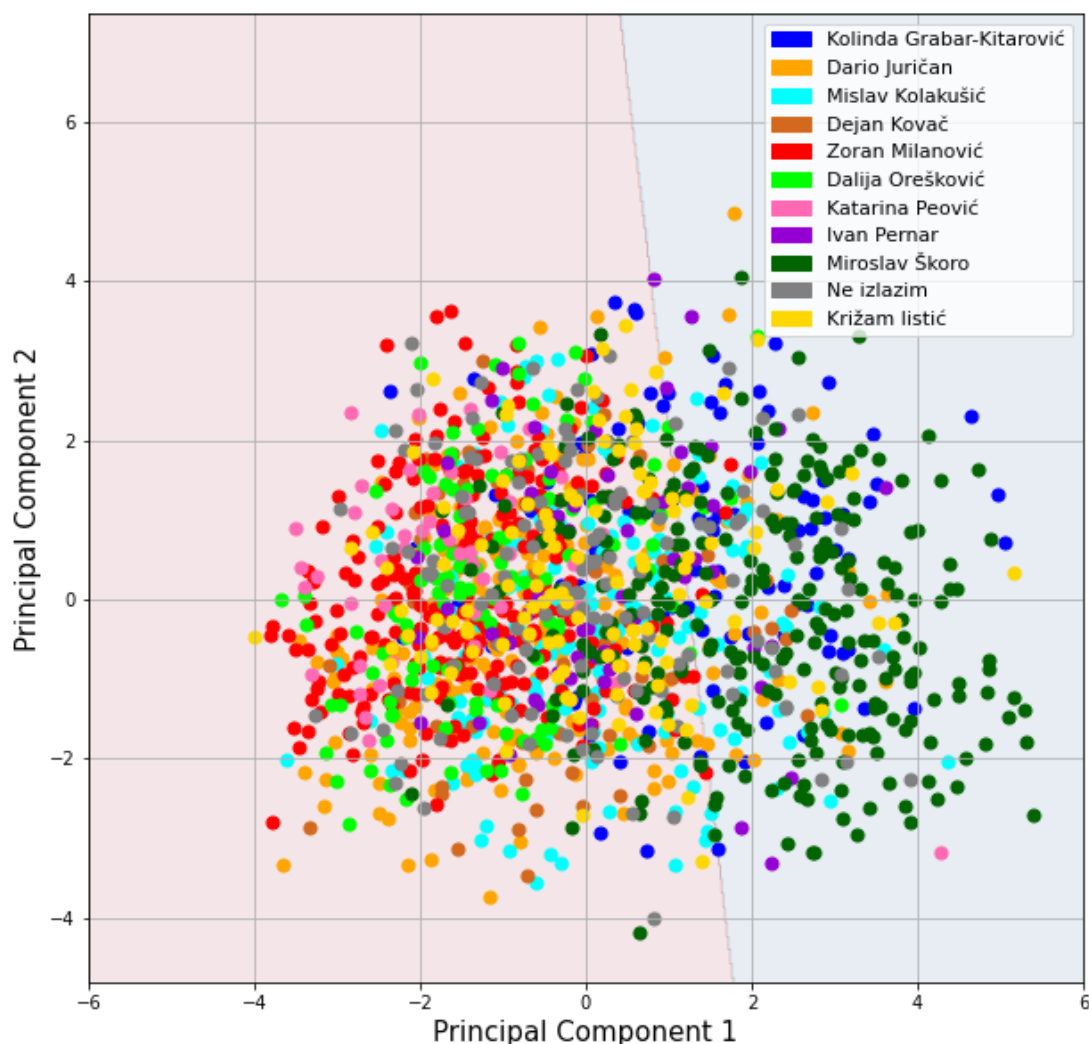
Tablica 4.1: Prikaz stavova glasača za pojedinog kandidata u prvom izbornom krugu

	Kolinda Grabar-Kitarović	Dario Juričan	Mislav Kolakušić	Dejan Kovač	Zoran Milanović	Dalija Orešković	Katarina Peović	Ivan Pernar	Miroslav Škoro	Ne izlazim	Križam listić
Broj glasova	98	225	166	59	246	115	50	59	270	136	122
Prosječni broj godina	21,54	21,52	21,10	21,39	21,37	21,60	21,62	20,64	20,88	21,29	21,29
Studentska populacija	90%	91%	92%	92%	94%	93%	100%	81%	94%	0,89	0,92
AIN1	3,19	3,13	3,83	3,29	2,59	2,89	2,72	3,80	4,05	3,08	3,36
AIN2	2,26	2,85	2,58	2,98	2,24	2,51	2,62	2,47	3,17	2,41	2,22
AIN3	3,50	2,59	2,40	2,54	2,94	2,76	3,04	2,80	2,87	2,90	3,11
AIN4	3,58	3,72	4,06	3,69	3,57	3,80	4,16	4,29	3,99	3,91	3,92
AIN5	3,96	4,44	4,39	4,53	4,33	4,34	4,14	4,17	4,19	4,26	4,40
AIN6	3,36	4,12	4,15	3,98	3,99	4,02	3,48	3,81	3,88	3,77	3,84
AIN7	3,38	3,98	4,19	4,08	3,88	3,89	3,98	4,24	4,07	3,82	3,96
AIN8	3,08	3,72	3,39	3,47	3,65	3,77	2,98	3,46	3,40	3,74	4,11
AIN9	2,89	3,04	2,46	2,36	2,71	2,59	1,96	2,69	2,93	2,57	2,63
AIN10	3,73	3,33	3,66	3,12	3,24	3,23	2,72	3,64	3,81	3,47	3,61
AIN11	3,73	2,61	2,95	2,31	2,27	2,22	1,88	2,86	3,60	2,88	2,89
AIN12	3,23	2,00	2,22	2,22	1,85	1,96	1,36	2,10	3,22	2,38	2,49
AIN13	2,61	2,56	2,35	2,31	2,61	2,49	2,34	1,83	1,91	2,40	2,29
AIN14	3,18	3,04	2,80	2,86	2,48	2,67	2,82	3,68	3,77	3,10	3,03
AIN15	3,08	1,52	1,72	1,54	1,17	1,25	1,10	1,85	3,03	1,79	1,86
AIN16	2,92	2,27	2,39	2,36	2,69	2,45	2,82	2,54	2,50	2,18	2,20
AIN17	3,05	2,48	2,81	2,61	1,96	2,06	1,46	2,61	3,56	2,59	2,52
AIN18	2,81	1,77	2,07	1,78	1,22	1,37	1,26	2,07	3,48	2,06	1,98
AIN19	2,16	1,94	2,07	2,24	1,73	1,78	1,96	3,54	2,23	2,14	2,19
AIN20	3,43	2,17	2,41	2,58	2,08	2,17	1,86	1,61	3,4	2,68	2,68
Prosjek odgovora	3,16	2,86	2,95	2,84	2,66	2,71	2,53	3,00	3,35	2,91	2,96

Prosječni odgovori glasača za pojedine tvrdnje iz ankete ovisno o kandidatima prikazani su u tablici 4.1. Originalni numerički podaci stavova koji polariziraju glasače organizirani su tako da liberalne predstavlja ocjena 3, lijeva polovica intervala lijevo orijentirane (interval $[1,3)$), a desna polovica desno orijentirane (interval $\langle 3, 5]$). Konkretno, podaci o slobodama rodni ideologija, pobačaju i marihuani dobiveni anketom izmijenjeni su tako da visoke ocjene uvijek predstavljaju očekivane desne stavove, dok se niže ocjene mogu poistovjetiti s lijevim stavovima. Zbog semantike pitanja to dotad nije bio slučaj. U poglavlju 2.3 ostavljena su originalno postavljena pitanja. No, u kasnijoj analizi ona su obrnuta tako da niže ocjene odgovaraju lijevim, odnosno više ocjene desnim stavovima.

Zanimljivo je uočiti da je ukupna prosječna ocjena glasača pojedinih kandidata u rasponu od 2.53 do 3.35 što je relativno mala razlika u odnosu na ukupan raspon. Generalni stav svih ispitanika je liberalan, odnosno vrijednost većine odgovora kreće se oko centra. Glasači desno orijentiranih kandidata, Miroslava Škore i Kolinde Grabar Kitarović, ocjenama većim od 3 blago naginju desnim stajalištima. Glasači Ivana Pernara ocjenom 3 predstavljaju strogi centar. Kod svih ostalih kandidata, glasači su bliži lijevim stavovima. Najveće razlike među stavovima glasača vidljive su u pitanju o ograničavanju prava žena na pobačaj i treba li predsjednik biti vjernik. Svi ispitanici, osim glasača Miroslava Škore i Kolinde Grabar Kitarović, u oba pitanja izrazili su lijevi stav. Isto je uočljivo na pitanju o Domovinskom ratu. Također, stav glasača desno orijentiranih kandidata o legalizaciji marihuane značajno je konzervativniji.

Svi glasači



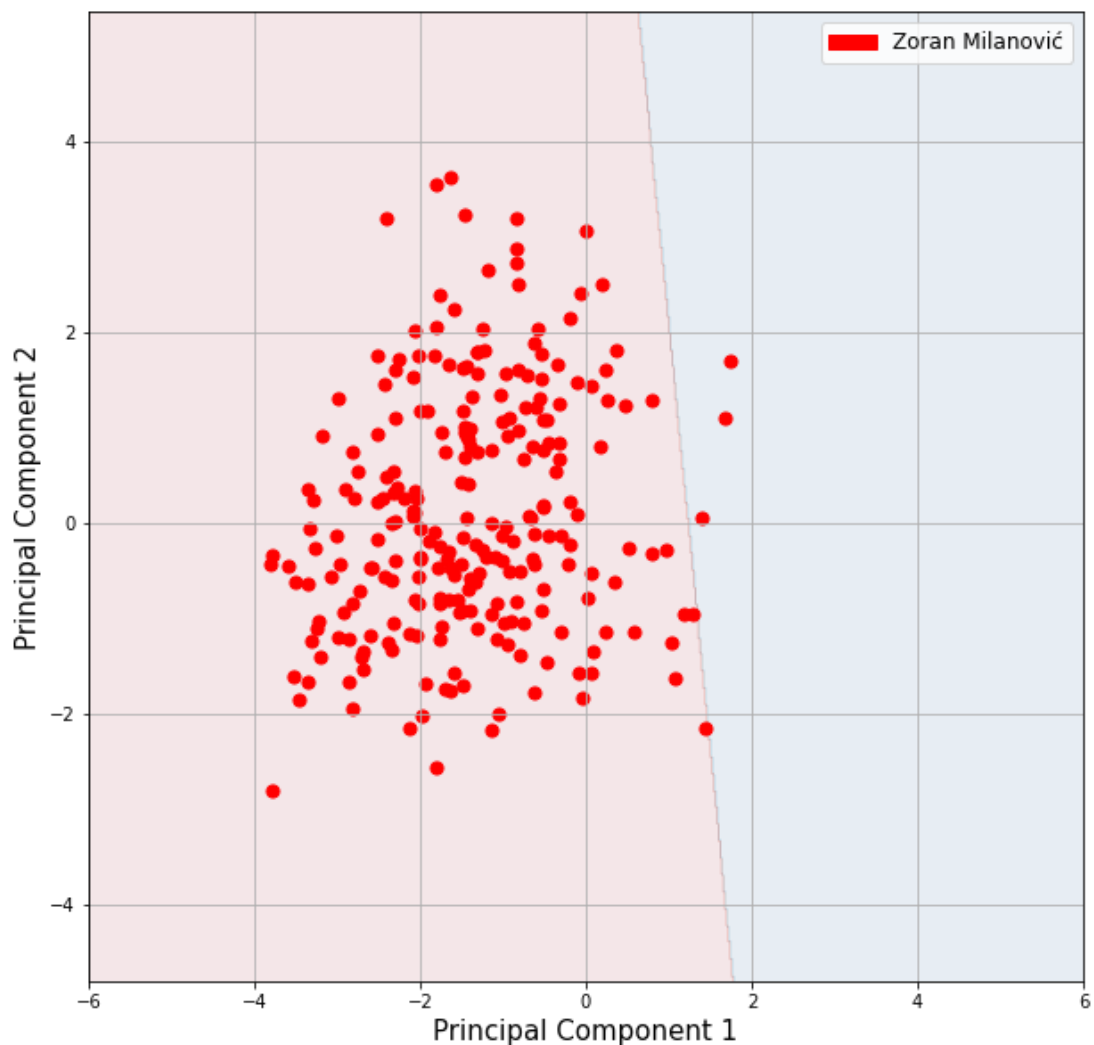
Slika 4.1: Prikaz svih glasača u prvome krugu

Na slici 4.1 grafički su predloženi svi glasači ponuđenih kandidata i oni koji odabiru opciju „Ne izlazim“ i „Križam listić“. Kako bi se dobio intuitivan vizualni prikaz glasača, graf je kreiran uz smanjenje dimenzionalnosti značajki pomoću analize glavnih komponenta. Ispitanici prikazani točkama s lijeve strane središnje osi predstavljaju lijevo orijentirane, dok su ispitanici s desne strane središnje osi desno orijentirani. Većina ispitanika koncentrirana je oko centra što ukazuje na veliku prisutnost liberalnih stavova. Najveći dio takvih ispitanika u svojim stavovima ipak blago naginje stavovima lijevo orijentiranih kandidata. Na desnoj strani grafa vidljiv je veći broj ispitanika s ekstremnijim stavovima. Raspršenost odgovora na desnoj strani grafa ukazuje na veću raznolikost između stavova desno orijentiranih u odnosu na lijevo orijentirane ispitanike.

Separator na grafu dobiven je pomoću linearne regresije. Tim se postupkom pronalazi linearna funkcija, odnosno pravac koji opisuje vezu između varijabli ispitanika. Parametri dobivenog pravca procijenjeni su na osnovi prikupljenih podataka i određeni tako da najbolje opisuju vezu između varijabli. To se radi metodom najmanjih kvadrata tako da se minimizira zbroj kvadrata vertikalnih odstupanja svih točaka od pravca. Separator odjeljuje glasače Kolinde Grabar-Kitarović (plava boja) i Zorana Milanovića (crvena boja).

Općenito, graf koji sadrži sve kandidate zbog svoje složenosti može biti nejasan za analizu. Iz tog razloga na sljedećim su stranicama izdvojeni pojedini kandidati te njihovi glasači. Uz glasače prikazan je i separator koji razdvaja glasače kandidata Zorana Milanovića i Kolinde Grabar-Kitarović.

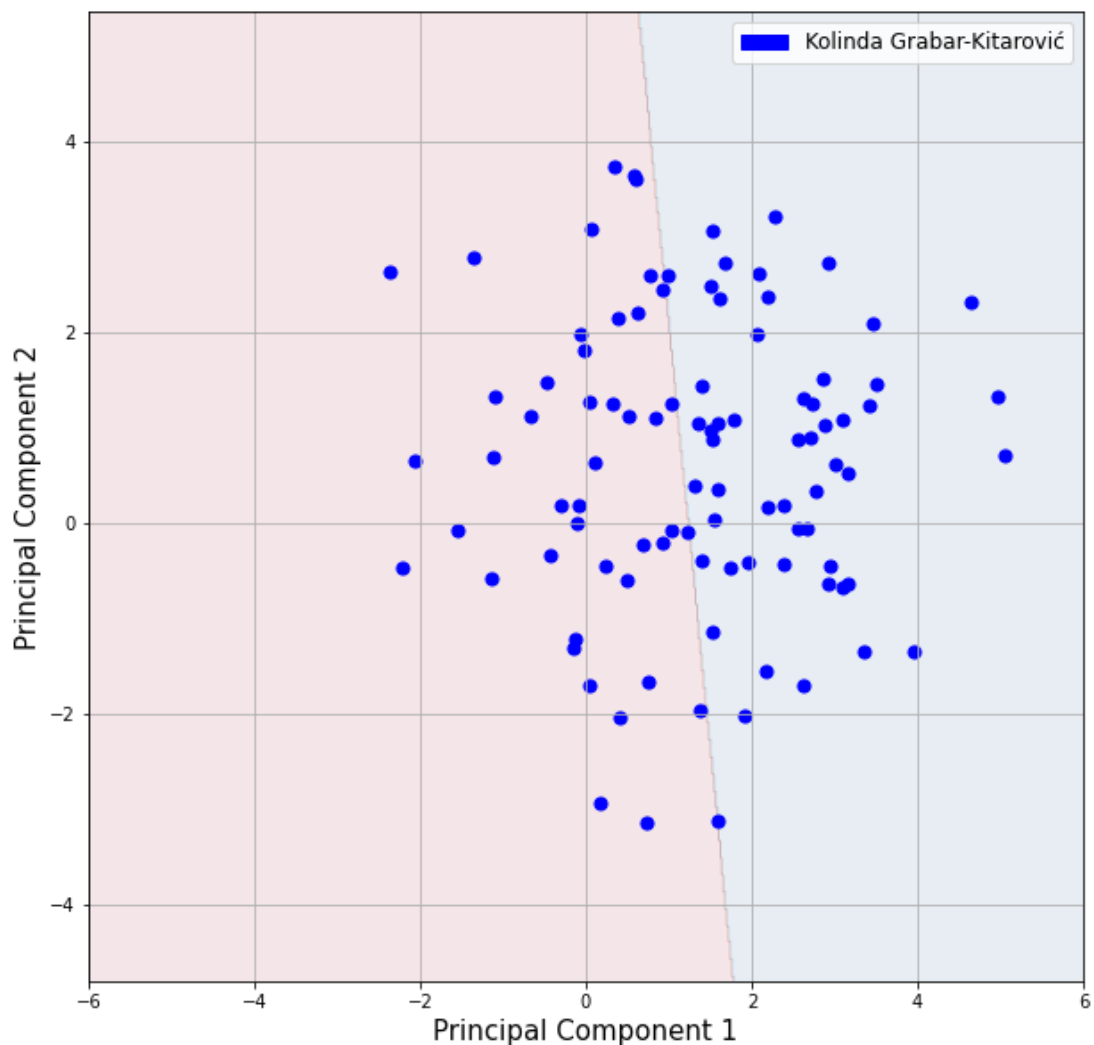
Glasači Zorana Milanovića



Slika 4.2: Prikaz glasača Zorana Milanovića

Na slici 4.2 vidljiva je raspodjela ispitanika koji bi glasali za Zorana Milanovića nakon provedene prve ankete. Iako postoje ispitanici koji su stavovima bliži centru i manji broj njih blago naginje prema desnim stavovima, gotovo svi glasači Zorana Milanovića lijevo su orijentirani. Na grafu nije vidljiva značajna raspršenost ispitanika što ukazuje na činjenicu da njegovi glasači daju slične odgovore. Stavovi većine glasača su između liberalnih i lijevo orijentiranih što se poklapa sa samim stavom i programom kandidata.

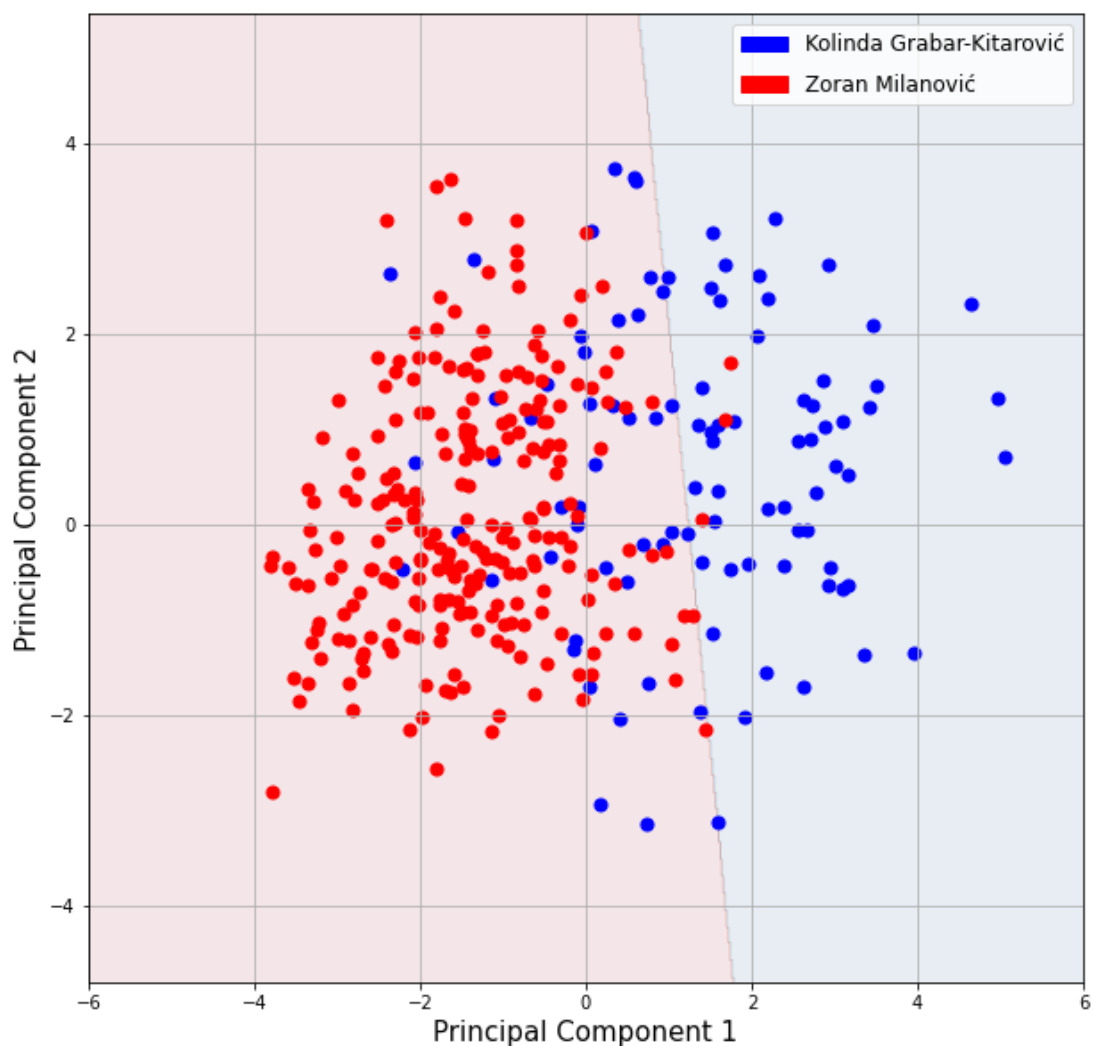
Glasači Kolinde Grabar-Kitarović



Slika 4.3: Prikaz glasača Kolinde Grabar-Kitarović

Ispitanici koji bi glasali za Kolindu Grabar Kitarović prikazani su na slici 4.3. Za razliku od kandidata Zorana Milanovića, njezini glasači većinski su desno orijentirani. Očekivano je da takvi glasači podržavaju konzervativnije ideje i program. U odnosu na prethodni, na ovome grafu uočljiva je veća raspršenost i veći broj glasača s ekstremnim stavovima. Pojedini ispitanici, glasači Kolinde Grabar Kitarović, međusobno imaju vrlo različite stavove. Dio glasača pokazuje lijevo orijentirane stavove.

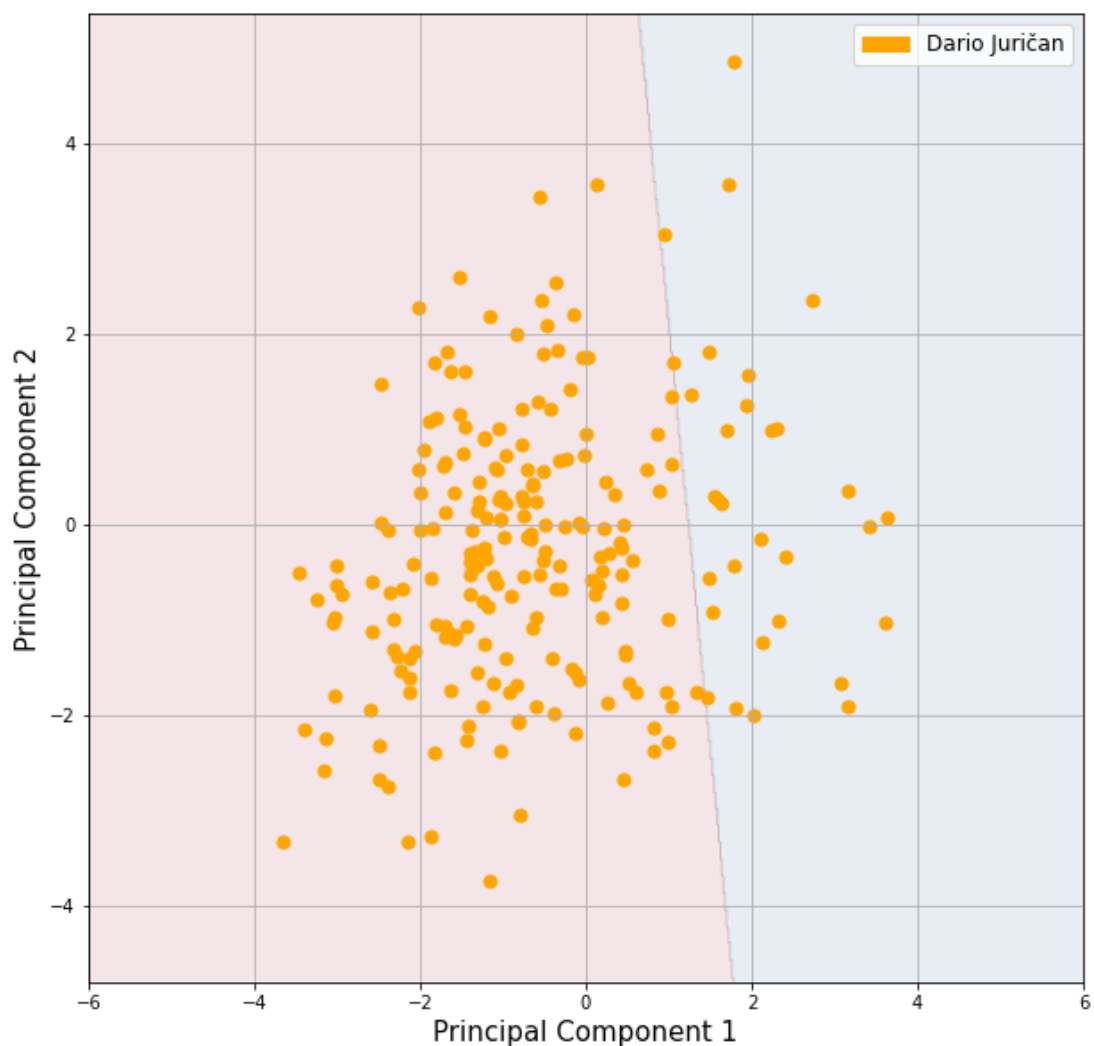
Usporedba glasača Kolinde Grabar-Kitarović i Zorana Milanovića



Slika 4.4: Prikaz glasača Kolinde Grabar-Kitarović i Zorana Milanovića

Na slici 4.4 vidljiva je usporedba raspodjele glasača Kolinde Grabar-Kitarović i Zorana Milanovića nakon provedenog prvog anketnog upitnika. U odnosu na centar, očekivano većina glasača Zorana Milanovića nalazi se s lijeve strane, dok su glasači Kolinde Grabar Kitarović većinski s desne strane. S obje strane postoje ispitanici čiji se stavovi nužno ne poklapaju s njihovim očekivanim izborom pri glasanju. Separator je bliži glasačima Kolinde Grabar Kitarović što dovodi do zaključka da su prvi anketni upitnik ispunjavale osobe koje su glasovale za Zorana Milanovića, a osim lijevih imaju i blago desne stavove. Općenito, ispitanici koji bi glasali za Zorana Milanovića međusobno imaju sličnije stavove u usporedbi s glasačima Kolinde Grabar Kitarović koji pokazuju veću raspršenost.

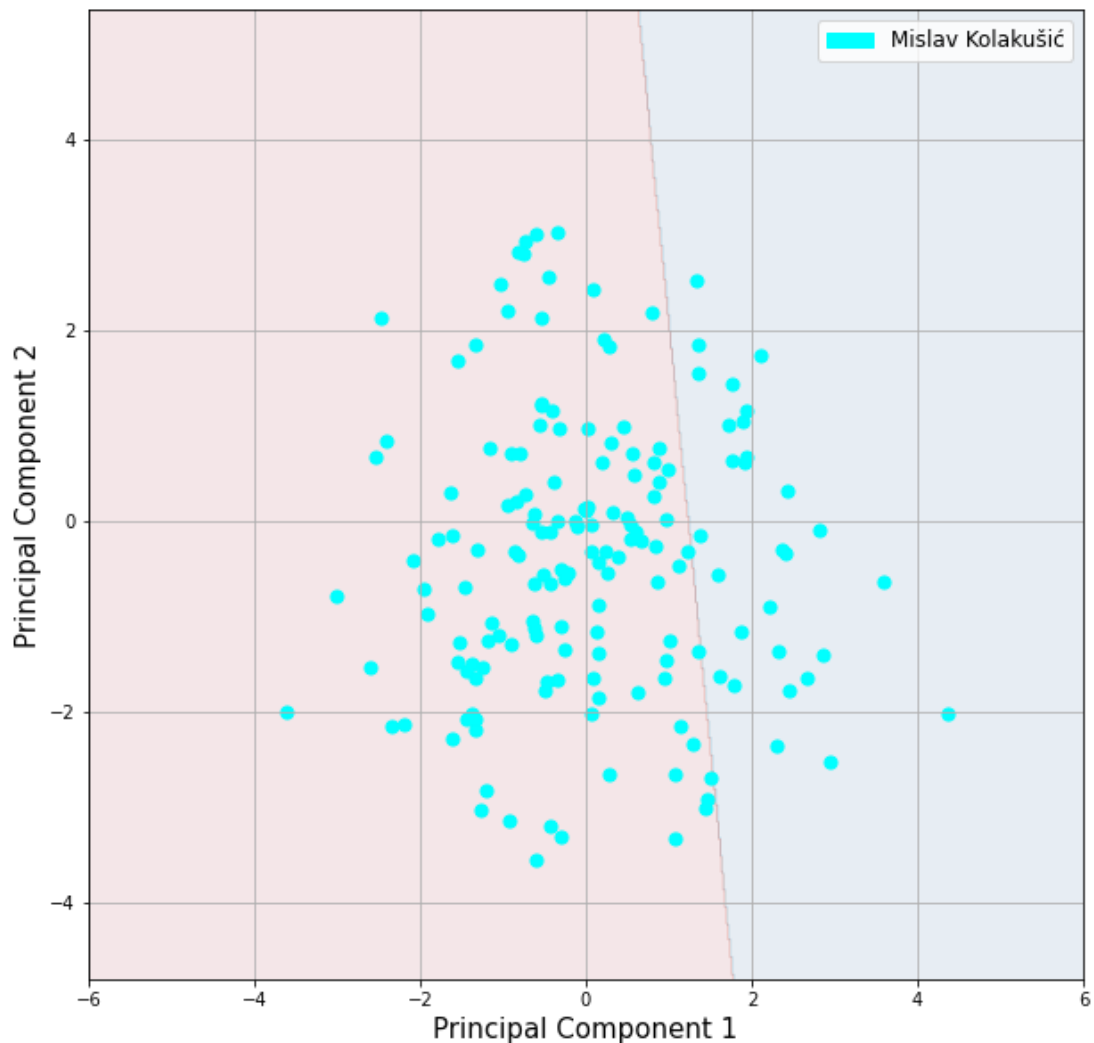
Glasači Darija Juričana



Slika 4.5: Prikaz glasača Darija Juričana

Na slici 4.5 vidljiva je politička orijentiranost ispitanika koji bi glasali za Darija Juričana. Primjećuje se da je više ispitanika koji planiraju glasati za ovog kandidata s lijeve strane središnje osi nego s desne. To daje naznaku da glasači kandidata imaju pretežito lijevo orijentirane stavove, jednako kao i kandidat kojega su odabrali. Samo manji dio glasača ima blago desne stavove.

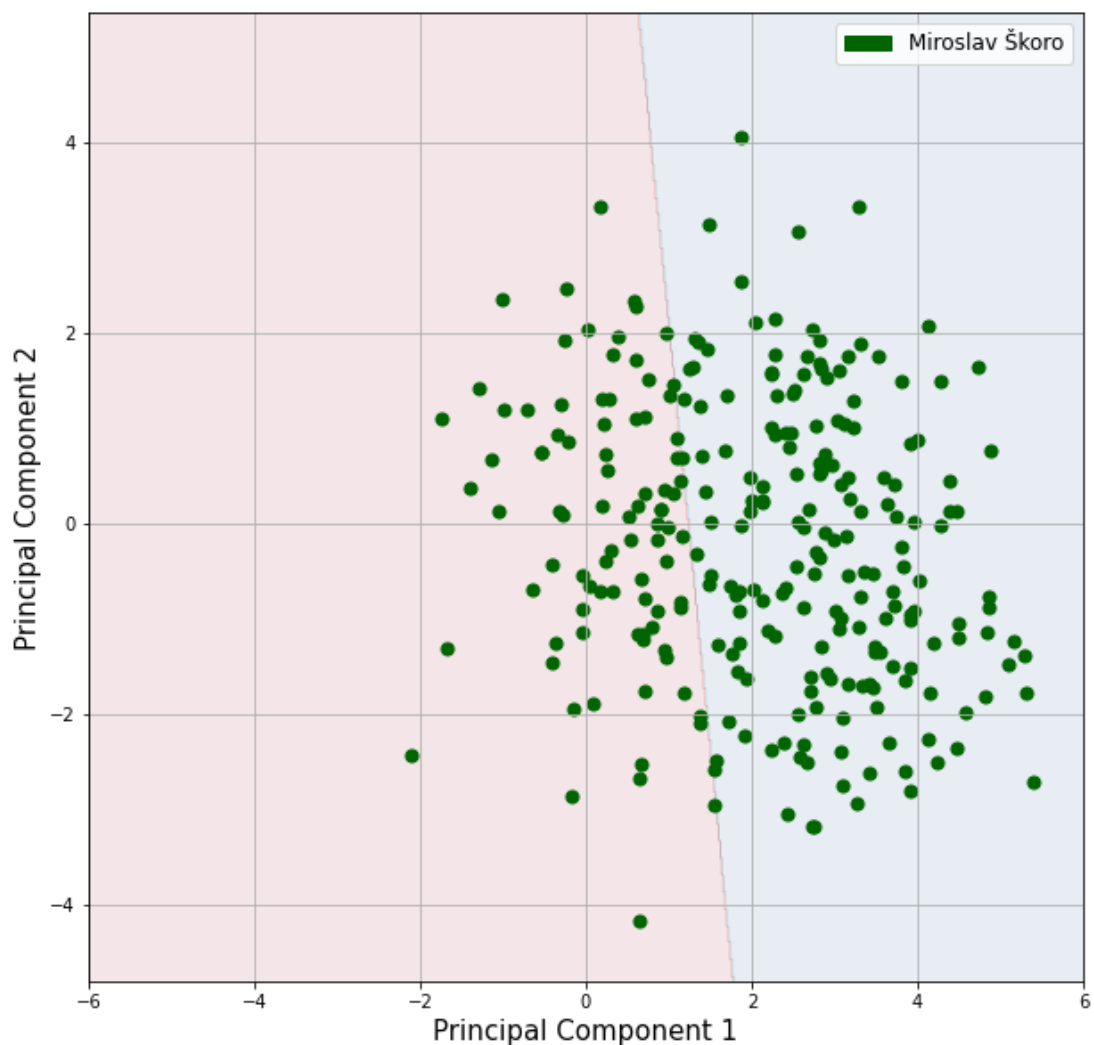
Glasači Mislava Kolakušića



Slika 4.6: Prikaz glasača Mislava Kolakušića

Za kandidata Mislava Kolakušića uočava se kako je omjer ispitanika s lijeve i desne strane središnje osi vrlo sličan kao i kod prethodnog kandidata, Darija Juričana, što je vidljivo na grafu prikazanom na slici 4.6. Iz toga je moguće naslutiti da glasači ovoga kandidata pretežito naginju lijevim idejama ili centru.

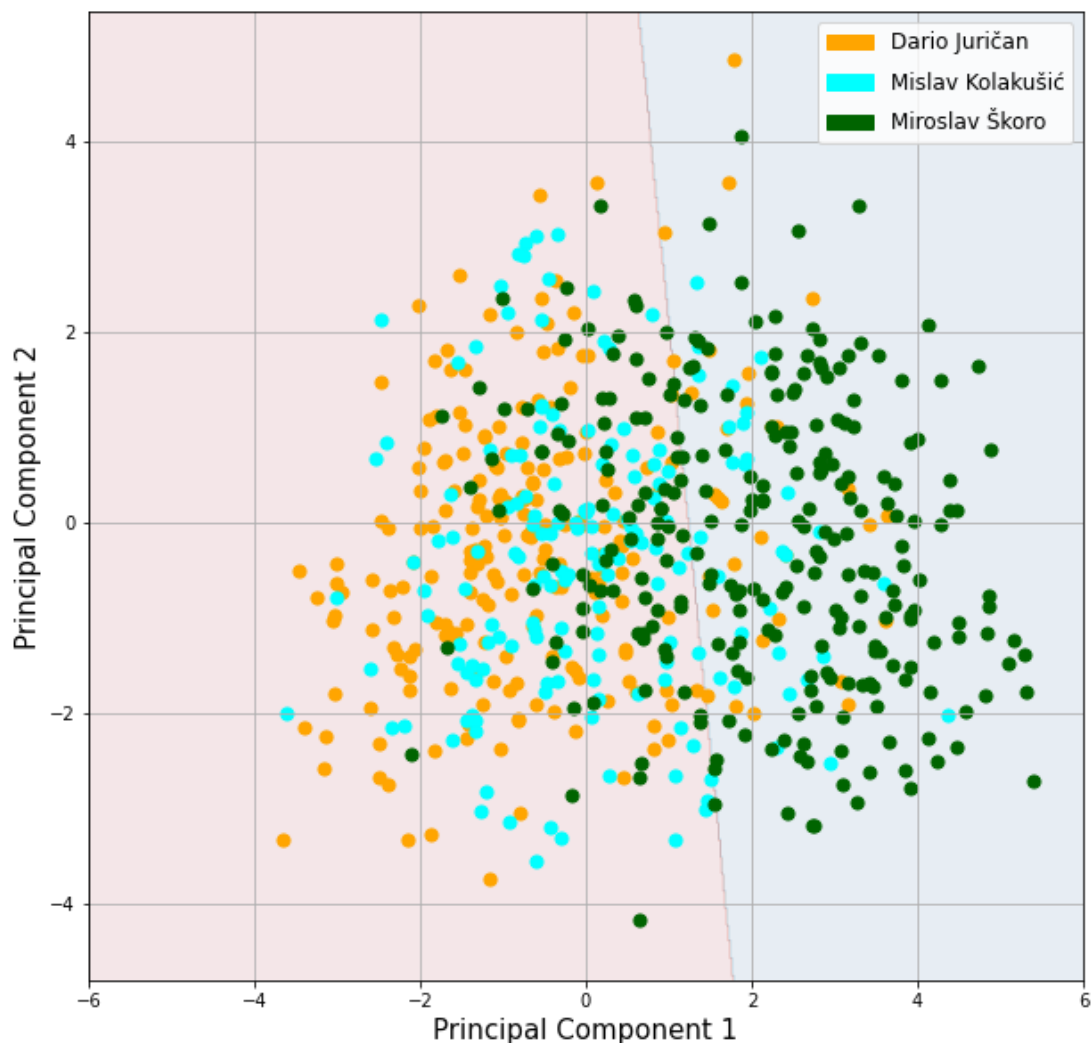
Glasači Miroslava Škore



Slika 4.7: Prikaz glasača Miroslava Škore

Za razliku od prethodna dva kandidata, ispitanici koji bi glasovali za Miroslava Škoro, prikazani na slici 4.7, nalaze se pretežno na desnoj strani središnje osi. To dovodi do zaključka da njegovi glasači imaju desno orijentirane stavove i vjerojatno podržavaju konzervativnije ideje i program. Usporedbom s glasačima Kolinde Grabar-Kitarović, vidljivo je da su glasači Miroslava Škore više desno orijentirani nego glasači bivše predsjednice.

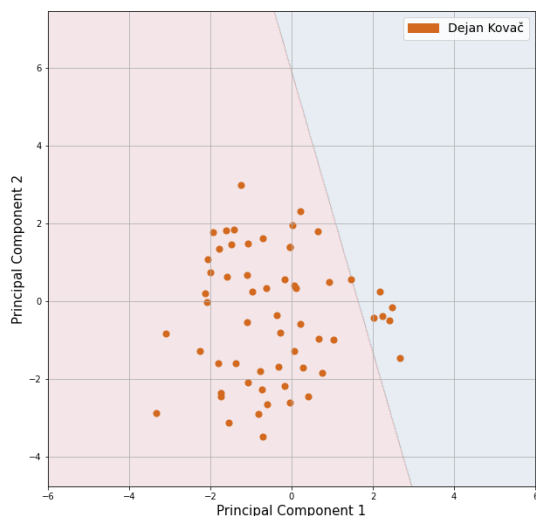
Usporedba glasača Darija Juričana, Mislava Kolakušića i Miroslava Škore



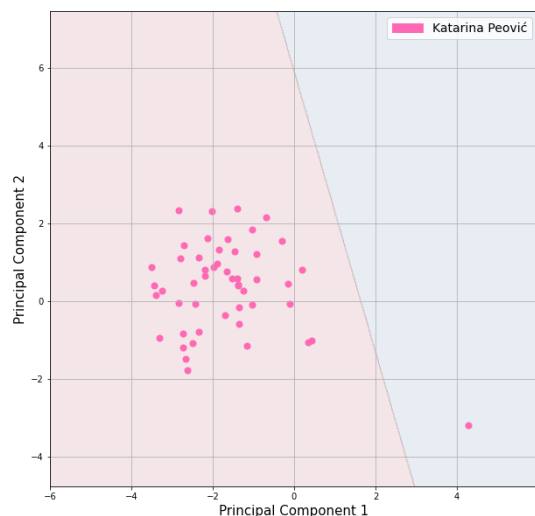
Slika 4.8: Prikaz glasača Darija Juričana, Mislava Kolakušića i Miroslava Škore

Na slici 4.8 vidljiva je usporedba glasača prethodnih triju kandidata: Darija Juričana, Mislava Kolakušića i Miroslava Škore. Ova trojica kandidata imala su najviše glasova u prvome krugu, ne uzimajući u obzir finaliste (Kolindu Grabar-Kitarović i Zorana Milanovića). U respektivnim odnosima moguće je usporediti njihove političke stavove prema odnosima točaka na grafu. Jasno je dočarano kako je Dario Juričan kandidat koji privlači osobe s najviše lijevim političkim stajalištima, Miroslav Škoro kandidat koji privlači osobe s najviše desnim političkim stajalištima, dok Mislav Kolakušić, iako bliže centru od ostalih dvaju kandidata, ipak više naginje ljevici nego desnici.

Glasači Dejana Kovača i Katarine Peović



Slika 4.9: Prikaz glasača Dejana Kovača

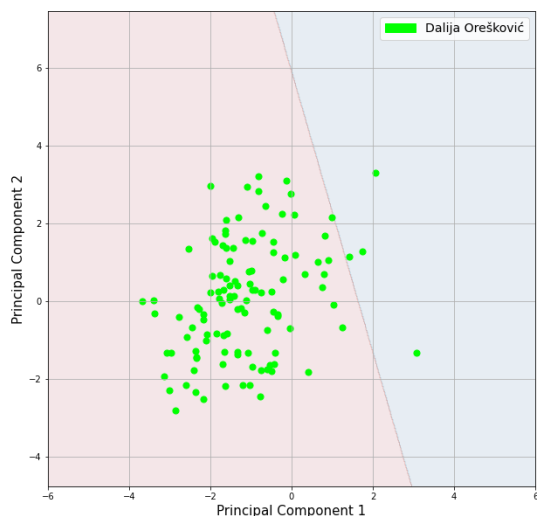


Slika 4.10: Prikaz glasača Katarine Peović

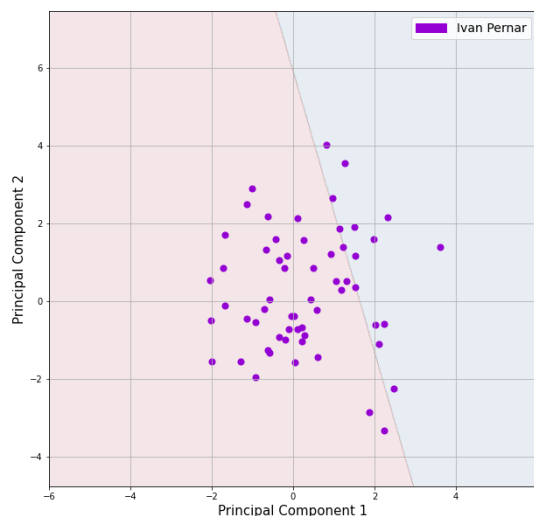
Na grafu prikazanom na slici 4.9 vidljiva je raspodjela političke orijentiranosti ispitanika koji će glasati za Dejana Kovača. Moguće je jasno vidjeti kako se velika većina njih nalazi s lijeve strane središnje osi, iz čega zaključujemo kako se radi o kandidatu koji predstavlja lijevo orijentirane ideje.

Graf prikazan na slici 4.10 prikazuje gotovo stopostotnu pripadnost lijevoj strani središnje osi. Iz njega se lako da zaključiti kako se kandidatkinja Katarina Peović zalaže za ideje koje privlače skoro isključivo lijevo orijentirane ispitanike. Od svih dosad prikazanih grafova ovaj je najviše jednostran i kao takav dobro prikazuje uspješnost korištenog algoritma i njegovu odlučivost.

Glasači Dalije Orešković i Ivana Pernara



Slika 4.11: Prikaz glasača Dalije Orešković

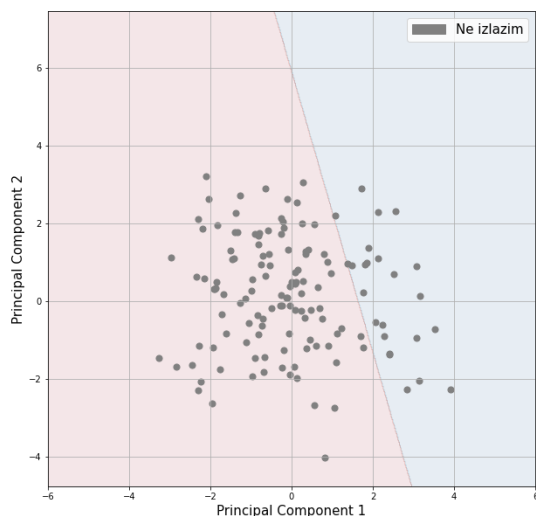


Slika 4.12: Prikaz glasača Ivana Pernara

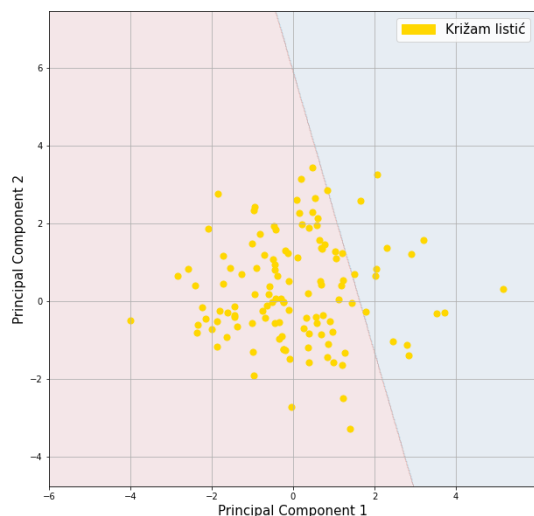
Slično kao i na grafu za prethodnog kandidata, na grafu ispitanika koji planiraju glasati za Daliju Orešković, prikazanom na slici 4.11, vidljiva je većinska opredijeljenost za lijevu stranu središnje osi.

Podjela ispitanika koji bi glasali za Ivana Pernara vidljiva je na slici 4.12. Njihov broj prevladava na lijevoj strani grafa, ali ni desna strana ne zaostaje mnogo. Velik broj ispitanika nalazi se u blizini središnje osi, a poneki su i na njoj samoj. Ovaj kandidat mogao bi se svrstati među kandidate koji zastupaju politički centar, odnosno ideje koje nisu naročito ni lijevo ni desno orijentirane.

Glasači koji su birali odgovor „Ne izlazim“ i „Križam listić“



Slika 4.13: Prikaz glasača koji su odabrali odgovor „Ne izlazim“

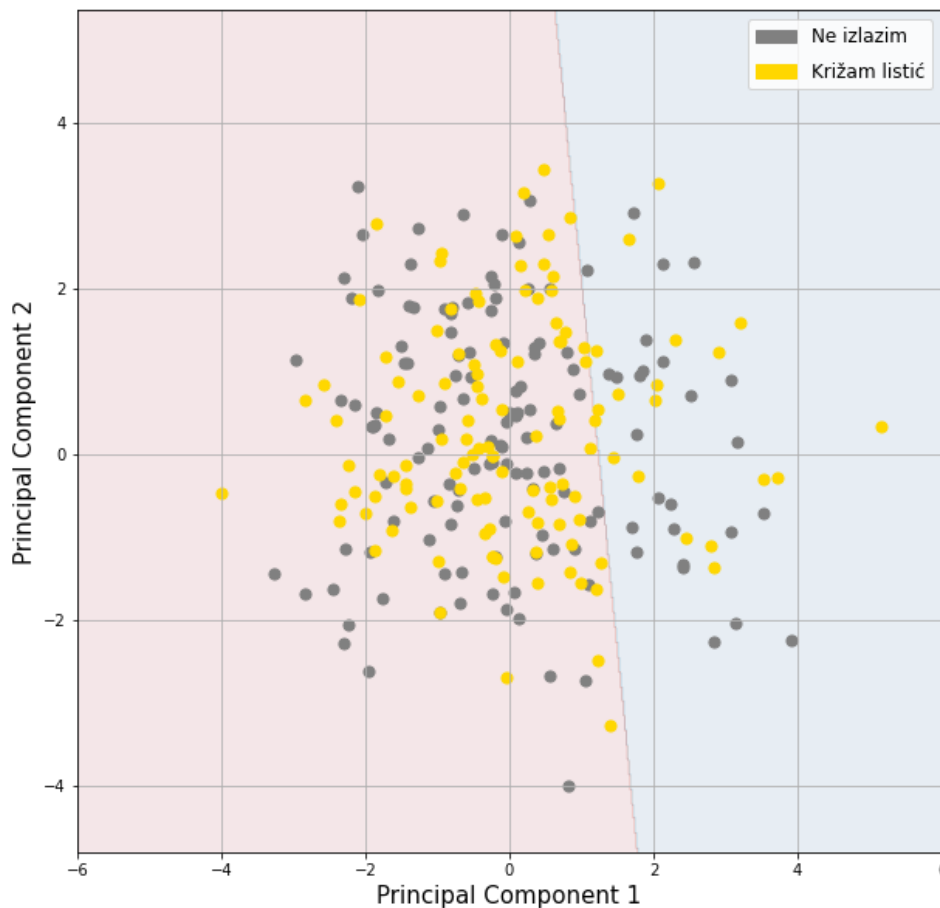


Slika 4.14: Prikaz glasača koji su odabrali odgovor „Križam listić“

Graf ispitanika koji ne planiraju izaći na izbore prikazan je na slici 4.13. Većina takvih ispitanika nalazi se s lijeve strane grafa. Brojni su kandidati, čije smo grafove do sada prikazali, imali glasače sa sličnom raspodjelom na grafu. Zanimljivo je da ovi ispitanici nisu uspjeli pronaći kandidata kojemu bi udijelili svoj glas, unatoč tome što imaju raspodjelu na grafu sličnu pojedinim kandidatima.

Ispitanici koji planiraju prekrížiti svoj listić prikazani su grafom na slici 4.14. Većina ih je s lijeve strane središnje osi, iako se može primijetiti i kako je određen broj njih blizu središta. Isto kao i kod ispitanika koji ne planiraju izaći na izbore, vide se sličnosti s raspodjelom ispitanika koji će glasovati za pojedine kandidate.

Usporedba glasača koji su birali odgovor „Ne izlazim“ ili „Križam listić“



Slika 4.15: Prikaz glasača koji su odabrali odgovor „Križam listić“ ili „Ne izlazim“

Na slici 4.15 usporedno je vidljiva raspodjela ispitanika koji ne planiraju izaći na izbore i onih koji planiraju prekrižiti listić. Raspodjela je vrlo slična u oba slučaja, pretežito s lijeve strane središnje osi uz veliku prisutnost pobornika političkog centra. Također, može se primijetiti i velika disperzija među stavovima ovih ispitanika. Otuda bi bilo moguće zaključiti kako su razlozi za neizlazak na izbore ili križanje listića brojni što čini teškim određivanje generalnog razloga za takve postupke.

4.2. Analiza druge ankete

U ovom dijelu analiziran je drugi anketni upitnik. Ispitani su i provjeravani stavovi glasača u drugom izbornom krugu ovisno o njihovu opredjeljenju u prvome krugu. Cilj je bio utvrditi mogućnost iščitavanja pravilnosti, sličnosti u stavovima glasača istog dijela političkog spektra; otkriti uzorke pripadnosti političkoj ljevici, odnosno desnici i izvesti zaključke o kandidatima koje biraju.

Tablica 4.2: Prikaz stavova glasača Kolinde Grabar-Kitarović ovisno o izboru u prvom glasačkom krugu

	Kolinda Grabar-Kitarović	Dario Juričan	Miroslav Škoro	Ne izlazim
Broj glasova	40	8	44	28
Prosječni broj godina	21,63	21,38	20,50	21,61
Studentska populacija	88%	100%	95%	100%
A1N1	2,88	3,50	3,61	3,18
A1N2	1,80	2,63	2,80	2,00
A1N3	3,40	2,50	3,07	3,36
A1N4	3,55	4,13	3,80	3,86
A1N5	3,93	4,13	4,05	4,11
A1N6	3,43	3,50	3,50	3,93
A1N7	3,20	3,75	4,05	4,18
A1N8	2,95	4,00	3,73	4,32
A1N9	3,25	2,75	2,93	3,00
A1N10	3,68	3,13	4,00	3,71
A1N11	3,43	2,63	3,95	3,43
A1N12	3,43	3,13	3,41	3,29
A1N13	2,30	2,25	2,20	2,36
A1N14	3,38	3,50	3,84	3,25
A1N15	2,98	2,50	3,45	2,93
A1N16	3,05	2,63	2,77	2,71
A1N17	4,13	4,13	3,48	4,11
A1N18	3,93	3,63	3,45	3,86
A1N19	1,95	2,50	2,00	1,96
A1N20	3,40	4,38	3,45	3,86
Prosjek odgovora	3,2	3,27	3,38	3,37

U tablici 4.2 analizirani su stavovi glasača Kolinde Grabar-Kitarović u drugom krugu ovisno o izboru kandidata u prvom krugu. Ima ih 138 od ukupno 727 glasača čiji su riješeni anketni upitnici prikupljeni. Od anketiranih glasača Nedjeljka Babića nitko se nije odlučio glasati za Kolindu Grabar-Kitarović. S obzirom na to da se vrlo malen broj glasača Ante Đapića, Mislava Kolakušića, Dejana Kovača, Zorana Milanovića, Dalije Orešković, Katarine Peović, Ivana Pernara te onih koji su u prvom krugu prekrížili listić naknadno odlučio glasati za Kolindu Grabar-Kitarović, njihovi odgovori nisu uzeti u obzir pri analizi. Statistički bi se moglo smatrati da je njihovo opredjeljenje za Kolindu Grabar-Kitarović u potpunosti slučajno. Očekivano, rezultati pokazuju da je Kolinda Grabar-Kitarović u drugom krugu mogla očekivati najviše glasova onih glasača koji su i u prvom krugu glasali za nju, kao i onih koji su glasali za Miroslava Škoru, također pripadnika desnice. Glasove je osvojila i od nekih glasača koji u prvom krugu nisu izlazili na izbore. Među statistički značajnije glasače pripadaju i oni Darija Juričana, ali u izrazitije manjem postotku.

Zanimljivo je vidjeti da ukupna prosječna ocjena glasača pojedinih kandidata, kao i prosječna ocjena na većinu pitanja, prelazi granicu političkog centra, ocjenu 3, naginjući višem prosjeku koji predstavlja desna stajališta. To je posebno uočljivo u visokom prosjeku u pitanjima o nedopuštanju prevelike slobode po pitanju rodni ideologija, o ograničavanju prava žena na pobačaj te o nepotrebnosti legalizacije marihuane, ali i činjenici da predsjednik treba biti pristupačna osoba iz naroda. Unatoč naginjanju višem prosjeku, prosječna ocjena ipak se nalazi u liberalnim okvirima političkog centra.

Tablica 4.3: Prikaz stavova glasača Zorana Milanovića ovisno o izboru u prvom glasačkom krugu

	Dario Juričan	Mislav Kolakušić	Zoran Milanović	Dalija Orešković	Miroslav Škoro	Ne izlazim
Broj glasova	43	25	149	32	22	61
Prosječni broj godina	21,72	21,36	21,83	21,38	21,27	21,48
Studentska populacija	95%	92%	90%	100%	91%	93%
A1N1	2,63	3,60	2,66	2,31	3,14	3,02
A1N2	2,72	2,52	2,03	2,13	2,64	2,13
A1N3	2,79	2,24	2,83	2,53	2,50	3,30
A1N4	3,40	3,44	3,42	3,09	3,68	3,67
A1N5	4,44	4,48	4,37	4,41	4,55	4,31
A1N6	4,00	4,36	4,09	3,91	4,00	3,82
A1N7	3,98	4,20	4,03	3,75	3,68	4,00
A1N8	4,05	3,44	3,26	3,59	3,86	3,97
A1N9	3,19	2,24	2,78	2,09	2,86	2,79
A1N10	3,67	4,00	3,82	3,41	3,86	3,82
A1N11	2,70	2,72	2,28	2,06	3,05	2,52
A1N12	2,05	2,08	1,95	2,00	2,73	2,15
A1N13	2,74	2,40	2,67	2,63	2,55	2,75
A1N14	2,84	2,64	2,58	2,38	3,45	2,82
A1N15	1,19	1,56	1,16	1,03	2,32	1,30
A1N16	2,84	2,44	2,72	2,75	2,23	2,48
A1N17	2,56	2,60	1,86	1,41	2,73	1,97
A1N18	1,30	1,80	1,27	1,25	2,95	1,66
A1N19	2,12	1,88	1,89	1,63	2,41	2,23
A1N20	2,00	2,64	2,23	2,06	3,00	2,30
Prosjek odgovora	2,86	2,86	2,7	2,52	3,11	2,85

U tablici 4.3 analizirani su stavovi glasača Zorana Milanovića u drugome krugu ovisno o izboru kandidata u prvom krugu. Ukupno ih je 380 od 727 anketiranih glasača, što je znatno veći broj u odnosu na glasače Kolinde Grabar-Kitarović. Iako su glasovi nekih glasača zanemareni zbog nedovoljnog doprinosa statističkoj značajnosti, veoma je istaknuto kako je Zoran Milanović osvojio simpatije glasača većeg broja kandidata nego Kolinda Grabar-Kitarović. To je također očekivan ishod zbog činjenice da je bilo više kandidata pripadnika političkoj ljevici. Zoranu Milanoviću u drugom su krugu ostali vjerni njegovi glasači, a pridružen mu je velik broj glasača koji u prvom krugu nisu izlazili na izbore, kao i mnogi glasači Darija Juričana, Dalije Orešković te Mislava Kolakušića. U statistički značajne glasače profilirani su i glasači Miroslava Škore iako u manjem postotku nego kod Kolinde Grabar-Kitarović. U usporedbi s tablicom 4.2, u tablici 4.3 prosječne ocjene mnogo su niže.

Zanimljivo je da promatrajući pojedina pitanja i njihove prosječne ocjene često nije moguće utvrditi opredjeljenje glasača. Iz toga je moguće izvesti zaključak da je potrebno analizirati stajališta iz različitih kutova. Potvrđena je hipoteza kako je

bitno oblikovati pitanja ankete po različitim tematskim cjelinama. Pitanja o povećanju ovlasti predsjednika, o potrebi političkog iskustva predsjednika, kao i prevlast osobnog dojma nad samim programom ne čine razliku među glasačima Kolinde Grabar-Kitarović i Zorana Milanovića. Prosječna ocjena tih pitanja nalazi se vrlo blizu samog centra. Visoke ocjene pri smatranjima kako bi predsjednik trebao povećati standard, kako je državni aparat prevelik, kako bi trebalo smanjiti broj radnih mjesta u javnom i državnom sektoru minimalno 20% i promijeniti izborni zakon također ne diferenciraju glasače. Glasajući obaju kandidata smatraju da biraju „manje zlo“. Potporu stranke manje očekuju glasači Zorana Milanovića, ali očekuju veću autoritarnost. Neznatno veći postotak glasača Zorana Milanovića smatra da bi trebalo uvesti euro kao sredstvo plaćanja. Glasajući Kolinde Grabar-Kitarović u značajnijoj mjeri smatraju Domovinski rat ishodišnom točkom moderne Hrvatske i naglašavaju potrebu ulaganja u Oružane snage Republike Hrvatske. Pitanja o preuzimanju odgovornosti za vlastite poteze i o slobodi cijepjenja ne odjeljuju glasače.

Najveća razlika stavova vidljiva je u pitanjima o potrebi da predsjednik bude vjernik, o slobodama rodnih ideologija, pobačaju i marihuani, gdje niže ocjene kod glasača Zorana Milanovića označavaju stavove o većoj slobodi.

Ukupna prosječna ocjena glasača Zorana Milanovića prema kandidatima prvog kruga (osim Miroslava Škore), kao i prosječna ocjena na većinu pitanja koja odjeljuju stajališta glasača, očekivano ne prelaze granicu političkog centra, ocjenu 3, već ostaju na lijevoj polovici cjelokupnog intervala. Kao i u tablici 4.2, moguće je primijetiti da ukupna prosječna ocjena ne odskače znatno od centra. Zanimljivo je primijetiti da je sustav glasače Dalije Orešković okarakterizirao kao najviše lijeve. Glasajući Miroslava Škore i u tablici 4.2 i u tablici 4.3 imaju prosječnu ocjenu veću od 3, što ih smješta u desni interval. Glasajući koji su se u drugom krugu opredijelili za Kolindu Grabar-Kitarović imaju veću prosječnu ocjenu odgovora na pitanja nego oni koji su se opredijelili za Zorana Milanovića, što je u skladu s dosadašnjim razmatranjima.

Tablica 4.4: Prikaz stavova glasača koji su odabrali odgovor „Ne izlazim“ ovisno o izboru u prvom glasačkom krugu

	Miroslav Škoro	Ne izlazim
Broj glasova	12	44
Prosječni broj godina	20,33	21,52
Studentska populacija	100%	95%
A1N1	3,08	3,14
A1N2	3,92	2,43
A1N3	2,25	3,41
A1N4	3,42	3,77
A1N5	4,42	4,20
A1N6	4,00	4,00
A1N7	4,25	3,82
A1N8	3,50	3,43
A1N9	2,50	2,77
A1N10	3,58	3,61
A1N11	3,83	2,52
A1N12	2,58	2,14
A1N13	2,50	2,52
A1N14	3,92	3,07
A1N15	2,92	1,73
A1N16	2,75	2,14
A1N17	3,83	2,75
A1N18	3,92	2,16
A1N19	2,08	2,20
A1N20	3,92	2,48
Prosjek odgovora	3,36	2,91

Osim opredjeljenja za Kolindu Grabar-Kitarović ili Zorana Milanovića, na izbore u drugom krugu bilo je moguće ne izlaziti ili prekriziti listić. Prikaz stavova glasača koji ne izlaze na izbore ovisno o izboru kandidata u prvom krugu vidljiv je u tablici 4.4. U kategoriji „Ne izlazim“ našlo se 80 od ispitanih 727 glasača. Najveći doprinos kategoriji dali su oni koji na glasanje nisu izlazili ni u prvom krugu te glasači Miroslava Škore koji u ponuđenim kandidatima nisu vidjeli zadovoljavajući odabir. Kao i u dvije prethodno razmatrane tablice, samostalno proučavajući mnoga pitanja nije moguće zaključiti opredjeljenje glasača niti razloge neizlaženja na izbore. Pitanja koja diferenciraju glasače Kolinde Grabar-Kitarović i Zorana Milanovića na isti način ovdje diferenciraju glasače Miroslava Škore, desne, i one koji u prvom krugu nisu izlazili, lijeve. Usrednjavanjem njihovih doprinosa prosjek taman pada na centar.

Tablica 4.5: Prikaz stavova glasača koji su odabrali odgovor „Križam listić“ ovisno o izboru u prvom glasačkom krugu

	Dario Juričan	Mislav Kolakušić	Miroslav Škoro	Ne izlazim	Križam listić
Broj glasova	19	16	47	11	13
Prosječni broj godina	21,58	21,88	21,04	21,36	20,92
Studentska populacija	95%	81%	96%	91%	92%
A1N1	2,58	4,00	3,85	3,36	2,69
A1N2	2,89	2,88	3,21	1,82	2,31
A1N3	2,74	2,25	2,60	2,73	2,92
A1N4	3,53	3,88	3,62	3,45	3,54
A1N5	4,58	4,38	4,40	4,36	3,85
A1N6	4,11	4,13	3,98	3,73	3,77
A1N7	4,21	4,38	4,23	4,27	4,00
A1N8	3,63	2,88	2,94	4,00	3,46
A1N9	2,74	2,25	2,57	2,73	2,46
A1N10	3,47	4,00	3,87	3,64	4,00
A1N11	2,53	2,81	4,00	2,45	2,15
A1N12	2,37	2,19	2,96	2,09	2,62
A1N13	2,68	2,13	1,83	2,27	2,08
A1N14	3,05	3,63	3,77	2,55	2,54
A1N15	1,42	2,19	2,94	1,64	1,92
A1N16	2,00	1,69	2,47	2,00	2,38
A1N17	2,47	3,19	3,79	2,64	2,38
A1N18	2,26	2,19	3,83	1,73	2,15
A1N19	1,58	2,50	2,19	1,73	2,23
A1N20	2,11	1,81	3,49	1,82	2,38
Prosjek odgovora	2,85	2,97	3,33	2,75	2,79

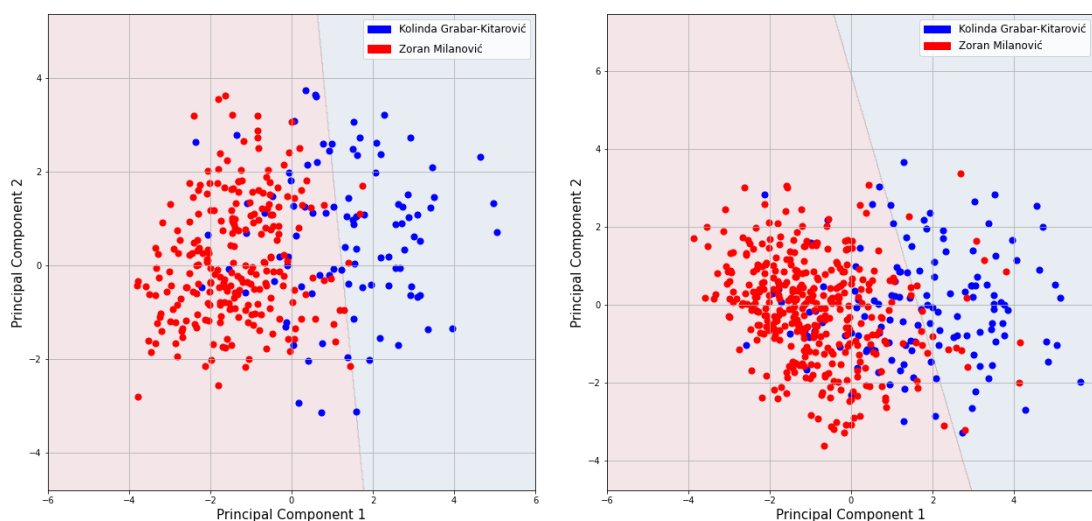
Tablica 4.5 prikazuje odgovore na pitanja glasača koji u drugom krugu križaju listić. U kategoriji „Križam listić“ našlo se 129 od 727 glasača, broj sličan količini glasača Kolinde Grabar-Kitarović. Najveći udio među njima pripao je glasačima Miroslava Škore. Manjim, ujednačenim postotkom doprinijeli su glasači Darija Juričana, Mislava Kolakušića te oni koji su i u prvom krugu križali listić ili nisu ni izlazili na izbore. Brojke dovode do sličnih zaključaka kao pri analizi prethodnih triju tablica. Glasajući Darija Juričana smješteni su u lijevi interval blisko političkom centru, glasači Mislava Kolakušića pripadnici su političkog centra, dok su glasači Miroslava Škore prema stavovima pozicionirani desno. Glasajući koji su u prvom krugu križali listić ili nisu izlazili na izbore pripadaju većinom političkom centru naginjući lijevim stavovima.

Tablica 4.6: Ukupna prosječna ocjena stavova glasača u drugom krugu

	Kolinda Grabar-Kitarović	Zoran Milanović	Ne izlazim	Križam listić
Broj glasova	138	380	80	129
Prosječni broj godina	21,19	21,64	21,35	21,36
Studentska populacija	0,95	0,92	0,98	0,93
A1N1	3,28	2,80	3,13	3,48
A1N2	2,28	2,20	2,70	2,70
A1N3	3,22	2,82	3,06	2,74
A1N4	3,75	3,46	3,76	3,63
A1N5	4,07	4,39	4,23	4,34
A1N6	3,61	4,00	3,96	3,91
A1N7	3,81	4,01	4,04	4,15
A1N8	3,67	3,60	3,45	3,21
A1N9	3,01	2,69	2,68	2,51
A1N10	3,73	3,77	3,66	3,77
A1N11	3,50	2,43	2,66	3,05
A1N12	3,31	2,04	2,26	2,55
A1N13	2,25	2,66	2,68	2,07
A1N14	3,56	2,72	3,20	3,26
A1N15	2,98	1,28	1,98	2,18
A1N16	2,83	2,66	2,33	2,26
A1N17	3,08	2,03	2,89	2,97
A1N18	3,14	1,48	2,36	2,8
A1N19	2,09	2,05	2,14	2,16
A1N20	3,36	2,21	2,66	2,61
Prosjek odgovora	3,23	2,76	2,99	3,02

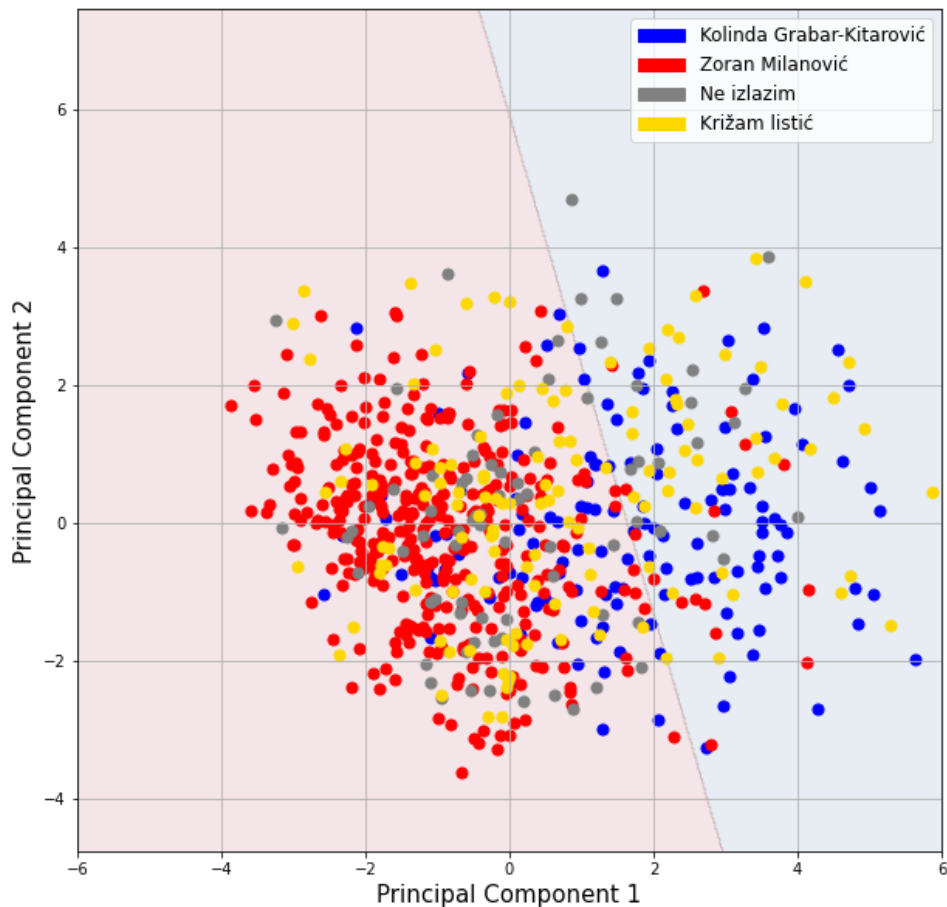
U tablici 4.6 prikazane su prosječne ocjene svih pitanja, kao i ukupne prosječne ocjene po opcijama za koje su se glasači mogli odlučiti u drugom krugu. Rezultati su očekivani. Glasaci Kolinde Grabar-Kitarović pripadaju desnom intervalu, a glasači Zorana Milanovića lijevom. Gotovo su podjednako udaljeni od centra u različitim smjerovima. S obzirom na to da kategoriju „Ne izlazim“ čine desno orijentirani glasači Miroslava Škore i lijevo orijentirani glasači koji nisu izlazili ni u prvom krugu, zbrajanjem se njihovi doprinosi usrednjavaju te se dobije rezultat političkog centra. U kategoriji „Križam listić“ stvar je veoma slična. Listić su odlučili prekriziti glasači mnogih lijevo orijentiranih kandidata, ali zbog činjenice da ih je uvjerljivo najviše među glasačima Miroslava Škore, rezultat biva smještan na samom centru. Zanimljivo je usporediti ukupne prosječne ocjene opcija drugog kruga s odgovarajućim opcijama prvog kruga. Prosjek od 3.16 Kolinde Grabar-Kitarović u prvom krugu prešao je u

3.23 u drugom krugu; udaljila se od centra dobitkom glasača desnije pozicioniranog Miroslava Škore. Prosjek od 2.66 Zorana Milanovića u prvom krugu prešao je u 2.76 u drugom krugu; on se približio političkom centru dobitkom glasača desnije pozicioniranog Miroslava Škore, centriranog Mislava Kolakušića te velikog broja glasača Darija Juričana koji je u prvom krugu okarakteriziran kao pripadnik ljevice, ali ipak bliži centru od Zorana Milanovića. Razlog povećanja prosjeka kategorije „Ne izlazim“ s 2.91 na 2.99 leži u doprinosu glasača Miroslava Škore koji se nisu željeli opredijeliti za ponuđene kandidate drugoga kruga. Isti je razlog i za povećanje prosjeka kategorije „Križam listić“ s 2.96 na 3.02.



Slika 4.16: Usporedba raspodjele glasača Kolinde Grabar-Kitarović i Zorana Milanovića nakon prvog i drugog izbornog kruga

Na slici 4.16 vidljiva je usporedba raspodjele glasača Kolinde Grabar-Kitarović i Zorana Milanovića nakon provedenog prvog i drugog anketnog upitnika čije provođenje odgovara vremenima prvog i drugog izbornog kruga. Očekivano, separator grafa u obje ankete odvaja glasače Zorana Milanovića kao lijevo orijentirane, a glasače Kolinde Grabar-Kitarović kao većinski desno orijentirane. Moguće je primijetiti da separator nema isti nagib u prvom i drugom krugu zbog različitog postotka glasača koji su riješili ankete, a pripadnici su pojedinih profila političkog spektra. U drugom krugu granica je nagnutija lijevoj strani, što uz prethodno razmatranje pokazuje da su drugi anketni upitnik ispunjavali pripadnici koji više nagnju centru i blago desnoj ideologiji, što je dovelo do pomicanja separatora u desnu stranu. U oba se slučaja općeniti stav svih ispitanika bolje poklapa sa stavovima glasača Kolinde Grabar Kitarović zbog većeg raspršenja u odnosu na pravilnije smještene glasače Zorana Milanovića.



Slika 4.17: Prikaz glasača koji su odabrali odgovor „Križam listić“ ili „Ne izlazim“

Na slici 4.17 vidljiva je raspodjela glasača koji su se u drugom glasačkom krugu opredijelili za Kolindu Grabar-Kitarović, Zorana Milanovića, odlučili ne izlaziti na izbore ili prekrížiti listić. Najveći broj ispitanika smješten je s lijeve strane u odnosu na centar. Više od polovice ispitanika (380/727) u drugom je krugu kao svoj odabir odredio Zorana Milanovića pa takav rezultat ne čudi. Zanimljiva je velika količina glasača koji su se svrstali u kategorije „Ne izlazim“ ili „Križam listić“. Moguće je izvesti zaključak da glasači neće odabrati kandidata koji predstavlja određen dio političkog spektra samo zato što se njihovi politički stavovi poklapaju s tim dijelom spektra.

U nastavku su provjerene nedosljednosti stavova u odgovorima pojedinih glasača. Uspoređivani su odgovori na pitanja treba li predsjednik biti vjernik i o pravu na izbor u kontekstu pobačaja. U prvom anketnom upitniku, glasači koji smatraju da bi predsjednik trebao biti vjernik i ne slažu se s pravom izbora pobačaja malobrojni su među pripadnicima političke ljevice. Konkretno, među glasačima Darija Juričana (njih 225 od ukupno ispitanih 1551) ima ih 10, među glasačima Mislava Kolakušića (166/1551) ima ih 7, među glasačima statistički manje značajnih Dejana Kovača, Katarine Peović i Ivana Pernara po 1. U kategoriji „Ne izlazim“ (136/1551) ima ih 9, a u kategoriji „Križam listić“ (122/1551) ima ih 5. U prethodno spomenute dvije kategorije u obzir su uzeti oni odgovori čiji se ukupni prosjek ocjena može smjestiti u lijevi interval. Svi ti rezultati ne doprinose statističkoj značajnosti jer doprinose ukupnom postotku s manje od 5% pa je moguće zaključiti i da se radi i o namjerno ili nenamjerno pogrešno označenim odgovorima. S druge strane, u prvom anketnom upitniku, glasači koji smatraju da nije potrebno da predsjednik bude vjernik i slažu se s pravom izbora pobačaja značajnije doprinose ukupnom rezultatu analize političke desnice. U slučaju Kolinde Grabar-Kitarović (98/1551) čak 24 glasača pokazuje takve stavove, dok kod Miroslava Škore (270/1551) njih 51. Za kategorije „Ne izlazim“ i „Križam listić“ uzeti su u obzir oni odgovori čiji se ukupni prosjek ocjena može smjestiti u desni interval. U kategoriji „Ne izlazim“ takve stavove pokazuje 86 glasača, a u kategoriji „Križam listić“ 81.

U drugom anketnom upitniku razmatranja su veoma slična onima iz prvog anketnog upitnika. Glasači koji smatraju da bi predsjednik trebao biti vjernik i ne slažu se s pravom izbora pobačaja malobrojni su među pripadnicima političke ljevice. Konkretno, među glasačima Zorana Milanovića (380/727) samo ih je 4. Među lijevo orijentiranim glasačima koji nisu izlazili na izbore (80/727) također ih je samo 4, a među onima koji su križali listić (129/727) 9. Glasači koji smatraju da nije potrebno da predsjednik bude vjernik i slažu se s pravom izbora pobačaja ponovno znatnije utječu na ukupni rezultat političke desnice. Među glasačima Kolinde Grabar-Kitarović (138/727) ima ih čak 36. Među desno orijentiranim glasačima koji nisu izlazili na izbore ima ih 45, a među onima koji su križali listić 56.

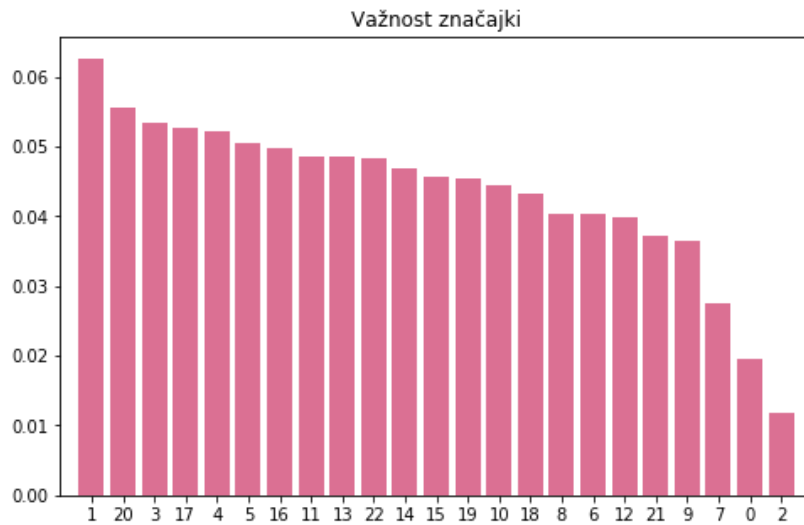
Iz takvih podataka moguće je zaključiti da među pripadnicima političke ljevice nema ekstremnih nekonzistentnosti stajališta, odnosno nema simpatizera stavova političke desnice. Međutim, vidljivo je da glasači političke desnice nisu u potpunosti ujedinjeni. Među desno orijentiranim glasačima postoje oni koji podržavaju stavove političke ljevice. Takvi stavovi vidljivi su i na grafovima, gdje lijevo od centra možemo primijetiti više od nekoliko plavih točkica (glasača) u oba kruga izbora.

4.3. Važnost značajki

Kao što je u prijašnjem poglavlju analizirano, iz same statistike glasova ne može se točno odrediti zbog čega je netko odabrao nekog kandidata. Iz samih odgovora, ljudsko oko ne može zaključiti za koga će netko glasovati. Stoga se u analizu uključuje strojno učenje. Kao što je prije spomenuto, slučajne šume imaju funkciju koja određuje koliko je koja varijabla važna. U svrhu određivanja važnosti varijabli te dubljeg razumijevanja kako stablo funkcionira, korištena je funkcija *feature_importances*. Teorija iza ove funkcije opisana je u poglavlju 3.2.1. Podaci korišteni za analizu značajki prve ankete jesu odgovori cijele prve ankete bez kandidata Nedjeljka Babića i Ante Đapića. Za analizu značajki druge ankete korištena je cijela druga anketa. Na slikama 4.18 i 4.19 te u tablicama 4.7 i 4.8 prikazane su značajke prve i druge ankete te njihova važnost.

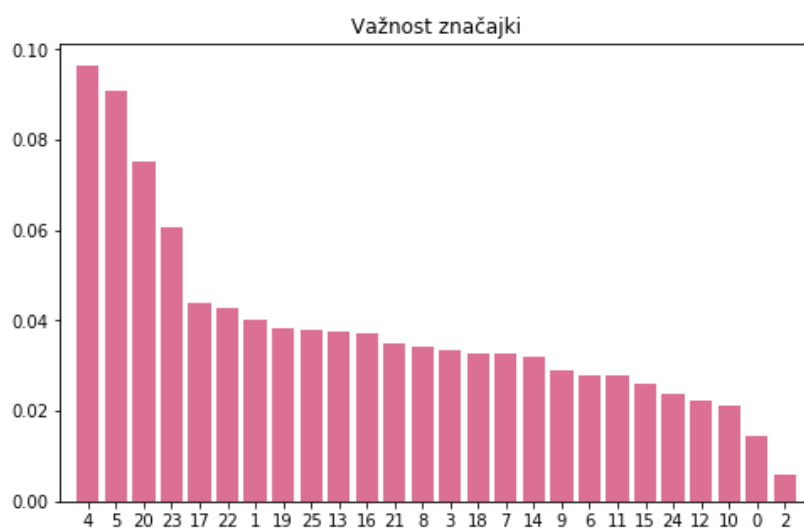
Tablica 4.7: Značajke i njihova važnost za prvu anketu [22]

Broj značajke	Oznaka značajke	Važnost značajke
1	Godine	0.062573
20	A1N18	0.055476
3	A1N1	0.053457
17	A1N15	0.052689
4	A1N2	0.052117
5	A1N3	0.050373
16	A1N14	0.049688
11	A1N19	0.048468
13	A1N11	0.048430
22	A1N20	0.048395
14	A1N12	0.046864
15	A1N13	0.045664
19	A1N17	0.045353
10	A1N8	0.044321
18	A1N16	0.043183
8	A1N6	0.040435
6	A1N4	0.040234
12	A1N10	0.039936
21	A1N19	0.037077
9	A1N7	0.036408
7	A1N5	0.027527
0	Spol	0.019610
2	Student	0.011720



Slika 4.18: Prikaz važnosti značajki za prvu anketu

Procjenom prikaza važnosti značajki prve i druge ankete (slike 4.18 i 4.19) primjetno je kako prvih nekoliko značajki druge ankete ima iznimnu važnost te znakovito informacijom doprinosi odluci klasifikacije. Kod prve ankete gotovo nije moguće izdvojiti nekoliko bitnih značajki. Za prvu anketu, na prvom mjestu, kao najvažnija značajka nalazi se značajka godine, a slijede ju značajke koje provjeravaju stavove o pobačaju, povećanju ovlasti, treba li kandidat biti vjernik te treba li imati prethodno političko iskustvo i potporu stranke. S druge strane, najmanje bitne značajke upravo su pitanja je li ispitanik student te kojeg je spola.



Slika 4.19: Prikaz važnosti značajki za drugu anketu

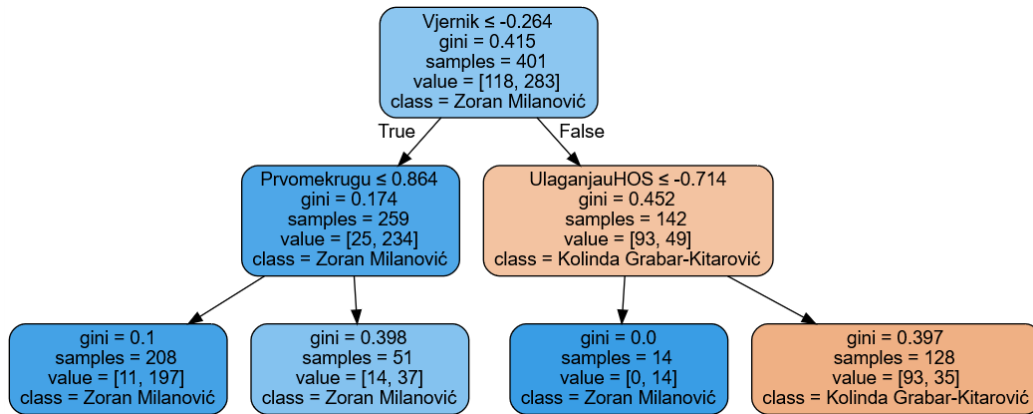
Tablica 4.8: Značajke i njihova važnost za drugu anketu [22]

Broj značajke	Kratki opis značajke	Važnost značajke
4	Za koga ste glasovali u prvom krugu?	0.096475
5	Za koga ćete glasovati na nadolazećim parlamentarnim izborima?	0.090684
20	A2N15	0.075286
23	A2N18	0.060753
17	A2N12	0.043903
22	A2N17	0.042895
1	Godine	0.040071
19	A2N14	0.038135
25	A2N20	0.037855
13	A2N8	0.037650
16	A2N11	0.037143
21	A2N16	0.035051
8	A2N3	0.034291
3	Za koga ste glasovali u drugom krugu predsjedničkih izbora 2015 godine?	0.033476
18	A2N13	0.032852
7	A2N2	0.032725
14	A2N9	0.032104
9	A2N4	0.029114
6	A2N1	0.027986
11	A2N6	0.027975
15	A2N10	0.026194
24	A2N19	0.023934
12	A2N7	0.022328
10	A2N5	0.021036
0	Spol	0.014373
2	Student	0.005710

Za drugu anketu prve su četiri značajke puno izraženije nego ostale. To su pitanja za koga se glasovalo u prvome krugu, za koga će se glasovati na nadolazećim parlamentarnim izborima, pitanje treba li kandidat biti vjernik te stav ispitanika o pobačaju. Zanimljivo je vidjeti da posljednja dva pitanja također ne doprinose važnosti s velikom ocjenom, kao i kod prve ankete. U nastavku je prikazan i primjer kako jedno stablo, koje čini šumu, koristi u svojim čvorovima značajke.

Primjer modela stabla s maksimalnom dubinom 2 nalazi se na slici 4.20. Stablo radi binarnu klasifikaciju za kandidate Kolindu Grabar-Kitarović i Zorana Milanovića

s podacima druge ankete. Stablo u čvorovima uzima u obzir samo značajke koje najbolje dijele kandidate. Prva značajka je treba li predsjednik/ca biti vjernik. Druga bitna značajka je pitanje za koga se glasovalo u prvome krugu izbora. Konačno, treća značajka pita treba li povećati ulaganja u Oružane snage Republike Hrvatske. Točnost ovakvog jednostavnog modela je 82%. Bitno je napomenuti da se ovim stablom ne radi predviđanje koje povezuje prvi i drugi krug. Stablo uči na jednom skupu podataka iz druge ankete i testira naučeno na ostatku skupa podataka druge ankete.



Slika 4.20: Primjer stabla s dubinom 2

Usporedbom slike 4.20 i poretka značajki u tablici 4.8 vidljivo je da su sve značajke korištene u stablu među pet najvažnijih u tablici. Poredak značajki, kao i njihov doprinos, mijenja se promjenom dubine stabla i promjenom podataka.

5. Predviđanje rezultata drugog kruga izbora

U ovom dijelu rada prikazani su testovi i analiza veze prve i druge ankete. U svakom testu korišten je algoritam slučajnih šuma. Sve značajke prve ankete su zadržane. Iz druge izbačene su značajke koje se razlikuju od prve, kao što su pitanja o parlamentarnim izborima i predsjedničkim izborima 2015. godine. Obje ankete imaju 23 značajke (23 pitanja) te odgovor na pitanje za koga će se glasovati na izborima. Za prvi dio analize korištene su tri verzije anketa. Anketa1_2odgovora označava podatke prve ankete gdje su konačni odgovori Kolinda Grabar-Kitarović i Zoran Milanović. Anketa2_2odgovora podaci su druge ankete, sa značajkama identičnim prvoj anketi i odgovorima Kolinda Grabar-Kitarović i Zoran Milanović. Druga verzija anketa su Anketa1_3odgovora i Anketa2_3odgovora gdje su značajke u obje ankete identične, ali je uz dva kandidata dodan i odgovor „Ostalo“, koji označava spoj odgovora „Ne izlazim“ i „Križam listić“. Treća verzija su Anketa1_4odgovora i Anketa2_4odgovora u kojima su odgovori „Ne izlazim“ i „Križam listić“ zasebni odgovori uz odgovore Kolinda Grabar-Kitarović i Zoran Milanović. U nastavku se nalazi tablica 5.1 s brojem osvojenih glasova za pojedini odgovor.

Tablica 5.1: Odnos broja glasova za odgovore prve i druge ankete

	Anketa 1	Anketa 2
Kolinda Grabar-Kitarović	98	138
Zoran Milanović	246	380
Ne izlazim na izbore	136	80
Križam listić	122	129

Prvi korak analize bilo je dokazivanje da su prva i druga anketa međusobno zavisne. Iz prve se pokušala „predvidjeti“ druga te obrnuto. Prilikom inicijalizacije hiperparametara slučajne šume, potrebno je postaviti podatke za treniranje i testiranje. Isprobane

su sve kombinacije gdje prva anketa predstavlja podatke iz kojih šuma uči, a druga se anketa pokušava predvidjeti. Nakon toga slijedi obrat gdje se iz druge ankete predviđa prva anketa. Svaki test napravljen je tako da se testiraju različite kombinacije broja stabala koja odlučuju u šumi ($n_estimators$) te maksimalna dubina stabla (max_depth). Broj mogućih stabala: 5, 10, 50, 100, 250, 300. Moguća maksimalna dubina: 2, 4, 8, 16, 32, 64, None. None označava da stablo nema restrikciju kolika može biti maksimalna dubina. U tablicama su prikazane kombinacije koje daju najveću točnost.

Tablica 5.2: Točnost algoritma slučajne šume kod analize veze prve i druge ankete

Trening podaci	Test podaci	$n_estimators$	max_depth	Točnost
Anketa1_2odgovora	Anketa2_2odgovora	50	4	85%
Anketa2_2odgovora	Anketa1_2odgovora	250	None	88%
Anketa1_3odgovora	Anketa2_3odgovora	300	64	58%
Anketa2_3odgovora	Anketa1_3odgovora	300	8	56%
Anketa1_4odgovora	Anketa2_4odgovora	100	64	56%
Anketa2_4odgovora	Anketa1_4odgovora	100	16	51%

Najbolji rezultati su kada se radi o binarnoj klasifikaciji, o odabiru između Kolinde Grabar-Kitarović i Zorana Milanovića. Rezultati su u tome slučaju 85% i 88%. Lošiji rezultati dobivaju se kada se uvedu nove varijable koje nisu tako jasno odjeljive kao ova dva odgovora. Najlošiji rezultat imaju testovi u kojima ankete sadrže četiri odgovora.

Drugi skup testova analizira koliko precizno slučajne šume mogu predvidjeti za koga će glasovati glasači pojedinog kandidata. U ovom slučaju odabrani su kandidati Dario Juričan, Mislav Kolakušić i Miroslav Škoro. Odabrani su jer su to tri kandidata koja su uz vodeća dva (Kolindu Grabar-Kitarović i Zorana Milanovića) osvojili najviše glasova u prvoj anketi. Algoritam uči na podacima prve ankete. Neka primjer bude Anketa1_2odgovora i Anketa2_DarioJuričan_2odgovora. Testni podaci sačinjeni su od druge ankete, s tim da su u obzir uzeti samo odgovori u kojima je rečeno da je u prvom krugu ispitanik glasovao za Darija Juričana. Anketa2_DarioJuričan se, dakle, sastoji od uzoraka druge ankete. Ima sve iste značajke kao prva anketa, no selektirani su samo uzorci u kojima je na pitanje o izboru prvog kruga odgovor bio Dario Juričan. Po istom principu selektirani su i uzorci za preostale dvije ankete, Anketa2_MislavKolakušić i Anketa2_MiroslavŠkoro. Ponovno su postojale tri opcije:

uzimanje u obzir samo dvaju kandidata, izbor uz pridodan odgovor „Ostalo“ te biranje između četiriju odgovora.

Tablica 5.3: Točnost predviđanja odgovora drugog kruga gdje je u prvom krugu ispitanik glasovao za Darija Juričana

Trening podaci	Test podaci	n_estimators	max_depth	Točnost
Anketa1_2odgovora	Anketa2_DarioJuričan_2odgovora	250	4	91%
Anketa1_3odgovora	Anketa2_DarioJuričan_3odgovora	300	32	60%
Anketa1_4odgovora	Anketa2_DarioJuričan_4odgovora	100	None	53%

Tablica 5.4: Točnost predviđanja odgovora drugog kruga gdje je u prvom krugu ispitanik glasovao za Mislava Kolakušića

Trening podaci	Test podaci	n_estimators	max_depth	Točnost
Anketa1_2odgovora	Anketa2_MislavKolakušić_2odgovora	100	4	74%
Anketa1_3odgovora	Anketa2_MislavKolakušić_3odgovora	250	32	55%
Anketa1_4odgovora	Anketa2_MislavKolakušić_4odgovora	250	8	52%

Tablica 5.5: Točnost predviđanja odgovora drugog kruga gdje je u prvom krugu ispitanik glasovao za Miroslava Škoru

Trening podaci	Test podaci	n_estimators	max_depth	Točnost
Anketa1_2odgovora	Anketa2_MiroslavŠkoro_2odgovora	100	4	75%
Anketa1_3odgovora	Anketa2_MiroslavŠkoro_3odgovora	100	32	45%
Anketa1_4odgovora	Anketa2_MiroslavŠkoro_4odgovora	250	8	39%

Točnost očekivano opada uvođenjem novih odgovora. Također, broj glasova za pojedini odgovor vrlo je disbalansiran. Najbolji rezultat je za model u kojem su testni podaci glasači Darija Juričana, gdje je točnost klasifikacije 91%. To znači da ne znajući za koga će netko glasovati u drugoj anketi, algoritam s točnošću od 91% prepozna iz odgovora na anketna pitanja za koga će glasovati pojedini glasač koji je u prvom krugu zaokružio Darija Juričana. Isti testovi provedeni su i za glasače kandidata Miroslava Škore i Mislava Kolakušića. Nešto lošiji rezultati dobivaju se za njihove glasače pretežito zbog manjka glasača s desno orijentiranim stavovima.

Treća analiza zadržava iste podatke za treniranje kao i u prethodnom testu, no mijenjaju se podaci za testiranje. Model ovoga puta ne gleda točnost na način određivanja ispravnosti smještanja glasača u političkom spektru, već sam predviđa kako će izgledati distribucija glasova. Model se testira nad podacima prve ankete i uspoređuje s distribucijom odgovora u drugoj anketi. Za primjer moguće je uzeti kandidata Darija Juričana i njegove glasače. Model uči nad odgovorima prve ankete gdje su ispitanici zaokružili Kolindu Grabar-Kitarović ili Zorana Milanovića. Podatke za testiranje će činiti svi uzorci koji su u prvoj anketi dali odgovor da će u prvom krugu izbora glasovati za Darija Juričana. Distribucija odgovora u drugoj anketi (gdje je na pitanje o prvom krugu zaokružen odgovor Dario Juričan) čini podatke za provjeru predviđanja.

Tablica 5.6: Usporedba predviđenih i stvarnih distribucija odgovora druge ankete, gdje je na pitanje o prvom krugu ispitanik zaokružio Darija Juričana

Trening podaci	Test podaci	n_estimators	max_depth	Raspodjela odgovora	Stvarna raspodjela odgovora
Anketa1_2odgovora	Anketa1_DarioJuričan	100	8	Kolinda Grabar-Kitarović: 18% Zoran Milanović: 82%	Kolinda Grabar-Kitarović: 14% Zoran Milanović: 86%
Anketa1_3odgovora	Anketa1_DarioJuričan	300	4	Kolinda Grabar-Kitarović: 1% Zoran Milanović: 54% Ostalo: 45%	Kolinda Grabar-Kitarović: 9% Zoran Milanović: 55% Ostalo: 36%
Anketa1_4odgovora	Anketa1_DarioJuričan	300	8	Kolinda Grabar-Kitarović: 8% Zoran Milanović: 72% Ne izlazim na izbore: 12% Križam listić: 8%	Kolinda Grabar-Kitarović: 9% Zoran Milanović: 55% Ne izlazim na izbore: 9% Križam listić: 27%

Tablica 5.7: Usporedba predviđenih i stvarnih distribucija odgovora druge ankete, gdje je na pitanje o prvom krugu ispitanik zaokružio Mislava Kolakušića

Trening podaci	Test podaci	n_estimators	max_depth	Raspodjela odgovora	Stvarna raspodjela odgovora
Anketa1_2odgovora	Anketa1_MislavKolakušić	50	4	Kolinda Grabar-Kitarović: 23% Zoran Milanović: 77%	Kolinda Grabar-Kitarović: 10% Zoran Milanović: 90%
Anketa1_3odgovora	Anketa1_MislavKolakušić	100	4	Kolinda Grabar-Kitarović: 1% Zoran Milanović: 32% Ostalo: 67%	Kolinda Grabar-Kitarović: 6% Zoran Milanović: 50% Ostalo: 44%
Anketa1_4odgovora	Anketa1_MislavKolakušić	300	16	Kolinda Grabar-Kitarović: 7% Zoran Milanović: 54% Ne izlazim na izbore: 26% Križam listić: 13%	Kolinda Grabar-Kitarović: 6% Zoran Milanović: 50% Ne izlazim na izbore: 14% Križam listić: 30%

Usporedbom distribucija u tablicama 5.6, 5.7 i 5.8, moguće je zaključiti kako najslabije distribucije između predviđenog i stvarnog ima tablica 5.6. Nešto lošiji rezultati su u tablici 5.7. Kada je cilj modela predvidjeti binarnu podjelu predviđajući distribuciju glasova, algoritam radi iznimno dobar posao. Razlika između stvarnog postotka i predviđenog je minimalna.

Tablica 5.8: Usporedba predviđenih i stvarnih distribucija odgovora druge ankete, gdje je na pitanje o prvom krugu ispitanik zaokružio Miroslava Škoru

Trening podaci	Test podaci	n_estimators	max_depth	Raspodjela odgovora	Stvarna raspodjela odgovora
Anketa1_2odgovora	Anketa1_MiroslavŠkoro	50	21	Kolinda Grabar-Kitarović: 73% Zoran Milanović: 27%	Kolinda Grabar-Kitarović: 65% Zoran Milanović: 35%
Anketa1_3odgovora	Anketa1_MiroslavŠkoro	10	4	Kolinda Grabar-Kitarović: 25% Zoran Milanović: 5% Ostalo: 70%	Kolinda Grabar-Kitarović: 36% Zoran Milanović: 19% Ostalo: 45%
Anketa1_4odgovora	Anketa1_MiroslavŠkoro	100	32	Kolinda Grabar-Kitarović: 46% Zoran Milanović: 11% Ne izlazim na izbore: 25% Križam listić: 18%	Kolinda Grabar-Kitarović: 35% Zoran Milanović: 20% Ne izlazim na izbore: 9% Križam listić: 36%

Četvrti i posljednji test predviđanje je kako će model raspodijeliti odgovore druge ankete te ih usporediti sa stvarnom distribucijom odgovora druge ankete. Ponovno se model uči na prvoj anketi. Kao i prije, tri su različite varijante: Anketa1_2odgovora, Anketa1_3odgovora, Anketa1_4odgovora, a skup za testiranje ovoga je puta druga anketa.

Tablica 5.9: Usporedba predviđenih i stvarnih distribucija odgovora druge ankete

Trening podaci	Test podaci	n_estimators	max_depth	Raspodjela odgovora	Stvarna raspodjela odgovora
Anketa1_2odgovora	Anketa2_2odgovora	250	4	Kolinda Grabar-Kitarović: 29% Zoran Milanović: 71%	Kolinda Grabar-Kitarović: 27% Zoran Milanović: 73%
Anketa1_3odgovora	Anketa2_3odgovora	250	None	Kolinda Grabar-Kitarović: 11% Zoran Milanović: 43% Ostalo: 46%	Kolinda Grabar-Kitarović: 19% Zoran Milanović: 53% Ostalo: 28%
Anketa1_4odgovora	Anketa2_4odgovora	300	8	Kolinda Grabar-Kitarović: 19% Zoran Milanović: 59% Ne izlazim na izbore: 14% Križam listić: 8%	Kolinda Grabar-Kitarović: 19% Zoran Milanović: 53% Ne izlazim na izbore: 11% Križam listić: 17%

U tablici 5.9 nalazi se posljednji test u kojem se predviđa distribucija glasova druge ankete. Kada algoritam bira među dvama kandidatima, rezultati su vanredno dobri. Model s nevjerojatnom točnošću predviđa kakva će biti razdioba odgovora druge ankete. Nešto slabiji rezultat dobije se ubacivanjem treće mogućnosti. No, dodavanjem četvrtog odgovora, algoritam ipak prepozna razliku između osoba koje biraju odgovore „Ne izlazim“ i „Križam listić“ te točnost ponovno raste.

6. Zaključak

Politika se u svojoj složenosti neosporno može definirati kao vještina obavljanja društvenih i državnih poslova s krajnjim ciljem postizanja općeg dobra. Prepoznata kao izuzetno zanimljiv aspekt svakodnevice, poslužila je nastanku interdisciplinarnog projekta čija je ideja bila na nov način iskoristiti mogućnosti raspona tehničkog područja interesa autora. Cilj projekta bio je konkretno primijeniti znanja stečena na fakultetu pri analizi društva u kojem živimo; povezati teorijska i praktična znanja stečena tijekom obrazovanja s aktualnom političkom situacijom. Implicitno, namjera je bila osvijestiti važnost političke informiranosti i angažiranosti u današnjem svijetu.

Analizom kandidata i njihovih programa dubinski su proučene razlike stavova pripadnika različitih dijelova političkog spektra. Sudjelovanjem u metodi fokus grupe istaknuta je važnost suradnje. Analizom rezultata metode fokus grupe poboljšana je vještina izvlačenja relevantnih podataka koji su se koristili pri sastavljanju ankete. Privlačenje velike količine podataka uzrokovalo je detaljna istraživanja metoda strojnog učenja koje mogu poslužiti daljnjem radu. Najdjelotvornijim se pokazao algoritam slučajnih šuma. Kao značajke ulaznih podataka definirane su pojedine tvrdnje anketa. Kako bi se dobio intuitivan i interpretabilan prikaz, velik broj značajki reduciran je na manji pomoću tehnike analize glavnih komponenta. Nakon obavljene redukcije dimenzionalnosti i efikasne vizualizacije, bilo je potrebno definirati rezultate i uklopiti ih u problem - otkriti što je moguće zaključiti o političkom spektru svakog ispitanog glasača i kako njegovo političko opredjeljenje utječe na konkretan izbor kandidata te koliko se predikcija poklapa sa stvarnošću.

Konkretno, analizom prve ankete ustanovljena je pripadnost najvećeg broja glasova ispitanika (270/1551) Miroslavu Škori. Zoran Milanović ne zaostaje mnogo (246/1551). Kandidati koji su u stvarnosti dobili mali postotak glasova, dobili su i od ispitanog dijela populacije. Ukupna prosječna ocjena glasača pojedinih kandidata u rasponu je između 2.53 i 3.35; neznatno odstupa od centra koji predstavlja ocjena

3, naginjući lijevim stajalištima. Grafički prikazi daju intuitivan vizualni prikaz smještaja kandidata u političkom spektru. Separator, pravac koji na najbolji mogući način odjeljuje varijable ispitanika, dobiven je linearnom regresijom. Takvim postupkom, gotovo svi glasači Zorana Milanovića okarakterizirani su kao lijevo orijentirani. Glasasči Kolinde Grabar-Kitarović prelaze separator s desne strane što donosi zaključak da podržavaju konzervativnije ideje. Glasasči koji su prekrížili listić ili nisu izlazili na izbore imaju sličnu raspodjelu u političkom spektru. Većina ih je koncentrirana oko centra, uz veći broj pobornika lijevih nego desnih ideja.

Analizom druge ankete zaključeno je kako su Kolinda Grabar-Kitarović i Zoran Milanović većinom očekivano dobili glasove onih glasača koji su pripadnici istog dijela političkog spektra. Prosječna ocjena stavova glasača Kolinde Grabar-Kitarović prelazi 3 naginjući višem prosjeku intervala desnih stajališta. Prosječna ocjena stavova glasača Zorana Milanovića ne prelazi 3, čime je smještena u interval lijevo orijentiranih ideja. Promatrajući pojedina pitanja i njihove prosječne ocjene često nije moguće utvrditi opredjeljenje glasača. Time je potvrđena moć pretpostavke o korisnosti analize stavova iz različitih kutova, po različitim tematskim cjelinama ankete. Najveći doprinos kategoriji onih koji ne izlaze na izbore u drugom krugu dali su oni koji nisu izlazili ni u prvom krugu, kao i glasači Miroslava Škore koji nisu našli prikladnog kandidata među ponuđenima. Listić su u drugom krugu prekrížili najviše glasači Miroslava Škore. Ukupne prosječne ocjene stavova i prosječne ocjene pojedinih pitanja glasača Kolinde Grabar-Kitarović i Zorana Milanovića promijenile su se ovisno o glasačima kandidata prvog kruga koji su im se naknadno pridružili. Velika količina glasača koji su odlučili ne izlaziti na izbore ili prekrížiti listić u drugom krugu dokaz su da glasač neće nužno odabrati kandidata koji predstavlja određeni dio političkog spektra samo zato što je i sam pripadnik tog dijela spektra. Također, jedna od analiza pokazala je da među ispitanim pripadnicima političke ljevice nema ekstremnih nekonzistentnosti stajališta, a da pripadnici političke desnice nisu međusobno ujedinjeni.

Drugi dio rada testira u kojoj se mjeri iz ankete može predvidjeti za koga će netko glasovati. Predviđanje je izvedeno pomoću algoritma slučajnih šuma. Analizirani su glasači trojice predvođćih kandidata koji nisu ušli u drugi krug, Darija Jurićana, Mislava Kolakušića i Miroslava Škore. Rezultati dobiveni predviđanjem iznimno su blizu stvarnim rezultatima. Točnost takvih predviđanja najčešće je oko 90%.

LITERATURA

- [1] Politika | Hrvatska enciklopedija, pristupljeno 20.8.2020. <https://www.enciklopedija.hr/natuknica.aspx?id=49240>.
- [2] Predsjednik republike | Hrvatska enciklopedija, pristupljeno 20.8.2020. <https://enciklopedija.hr/natuknica.aspx?id=50080>.
- [3] Izbori za predsjednika Republike Hrvatske, pristupljeno 22.2.2020. <https://www.izbori.hr/site/izbori-referendumi/izbori-za-predsjednika-republike-hrvatske/66>.
- [4] Predsjednički izbori u Hrvatskoj, pristupljeno 15.4.2020. https://hr.wikipedia.org/wiki/Predsjedni%C4%8Dki_izbori_u_Hrvatskoj.
- [5] Potpuni rezultati izbora za predsjednika Republike Hrvatske, pristupljeno 29.12.2019. https://www.izbori.hr/site/UserDocsImages/2019/Izbori_za_predsjednika_RH_2019/Izvjesce_o_rezultatima_I.kruga.pdf.
- [6] Božo Skoko and Vanesa Benković. Znanstvena metoda fokus grupa-mogućnosti i načini primjene. *Politicka Misao: Croatian Political Science Review*, 46(3), 2009.
- [7] Larissa A Grunig. Using qualitative research to become the “thinking heart” of organizations. In *Public Relations Metrics*, pages 136–152. Routledge, 2009.
- [8] Peter Westfall. Statistical significance, pristupljeno 5.8.2020. <https://www.investopedia.com/terms/s/statistical-significance.asp>.
- [9] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

- [10] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.
- [11] Sidharth P Mishra, Uttam Sarkar, Subhash Taraphder, Sanjay Datta, Devi P Swain, Reshma Saikhom, and M Laishram. Multivariate statistical data analysis-principal component analysis (pca). *International Journal of Livestock Research*, 7(5):60–78, 2017.
- [12] Krunoslav Rajčić. *Usporedba analize glavnih komponenti i faktorske analize*. PhD thesis, University of Zagreb. Faculty of Science. Department of Mathematics., 2015.
- [13] Željko Pedišić and Dražan Dizdar. Priručnik za kvantitativne metode. *Zagreb: Kineziološki fakultet Sveučilišta u Zagrebu*, 2010.
- [14] Lindsay I Smith. A tutorial on principal components analysis. Technical report, 2002.
- [15] Jason Brownlee. What is the difference between test and validation datasets?, pristupljeno 16.5.2020. <https://machinelearningmastery.com/difference-test-validation-datasets/>.
- [16] Bojana Dalbelo Bašić, Marko Čupić, and Jan Šnajder. Umjetne neuronske mreže. *Fakultet elektrotehnike i računarstva*, 2008.
- [17] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.
- [18] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [19] Jason Brownlee. Classification and regression trees for machine learning, pristupljeno 25.5.2020. <https://machinelearningmastery.com/classification-and-regression-trees-for-machine-learning/>.
- [20] Niklas Donges. Slučajni šumski algoritam - strojno učenje, 10.6.2020. <https://hr.sciencewal.com/77885-the-random-forest-algorithm-d457d499ffcd-33>.
- [21] Juraj Petrović. *Prediction of protein-protein interactions using Random Forests*. PhD thesis, Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu, 2010.

- [22] Mariela Markovinović. Model glasača na predsjedničkim izborima baziran na metodama strojnog učenja. Master's thesis, Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu, 2020.

POPIS TABLICA

2.1. Kandidati [5]	4
2.2. Osnovne informacije o uzorku	10
2.3. Distribucija odgovora na pitanja prve ankete	10
2.4. Distribucija odgovora na pitanja druge ankete	10
2.5. Odgovori na pitanje „Za koga ćete glasovati (ste glasovali) u prvome krugu predsjedničkih izbora?“ iz prve i druge ankete	11
2.6. Odgovori na pitanje „Za koga ste glasovali na predsjedničkim izborima 2015. godine?“	11
2.7. Odgovori na pitanje „Za koga ćete glasovati na predsjedničkim izborima?“	11
3.1. Matrice korištene primjenom analize glavnih komponenti	14
4.1. Prikaz stavova glasača za pojedinog kandidata u prvom izbornom krugu	24
4.2. Prikaz stavova glasača Kolinde Grabar-Kitarović ovisno o izboru u prvom glasačkom krugu	38
4.3. Prikaz stavova glasača Zorana Milanovića ovisno o izboru u prvom glasačkom krugu	40
4.4. Prikaz stavova glasača koji su odabrali odgovor „Ne izlazim“ ovisno o izboru u prvom glasačkom krugu	42
4.5. Prikaz stavova glasača koji su odabrali odgovor „Križam listić“ ovisno o izboru u prvom glasačkom krugu	43
4.6. Ukupna prosječna ocjena stavova glasača u drugom krugu	44
4.7. Značajke i njihova važnost za prvu anketu [22]	48
4.8. Značajke i njihova važnost za drugu anketu [22]	50
5.1. Odnos broja glasova za odgovore prve i druge ankete	52
5.2. Točnost algoritma slučajne šume kod analize veze prve i druge ankete	53

5.3.	Točnost predviđanja odgovora drugog kruga gdje je u prvom krugu ispitanik glasovao za Darija Juričana	54
5.4.	Točnost predviđanja odgovora drugog kruga gdje je u prvom krugu ispitanik glasovao za Mislava Kolakušića	54
5.5.	Točnost predviđanja odgovora drugog kruga gdje je u prvom krugu ispitanik glasovao za Miroslava Škoru	54
5.6.	Usporedba predviđenih i stvarnih distribucija odgovora druge ankete, gdje je na pitanje o prvom krugu ispitanik zaokružio Darija Juričana .	55
5.7.	Usporedba predviđenih i stvarnih distribucija odgovora druge ankete, gdje je na pitanje o prvom krugu ispitanik zaokružio Mislava Kolakušića	55
5.8.	Usporedba predviđenih i stvarnih distribucija odgovora druge ankete, gdje je na pitanje o prvom krugu ispitanik zaokružio Miroslava Škoru	56
5.9.	Usporedba predviđenih i stvarnih distribucija odgovora druge ankete .	56

POPIS SLIKA

4.1. Prikaz svih glasača u prvome krugu	25
4.2. Prikaz glasača Zorana Milanovića	27
4.3. Prikaz glasača Kolinde Grabar-Kitarović	28
4.4. Prikaz glasača Kolinde Grabar-Kitarović i Zorana Milanovića	29
4.5. Prikaz glasača Darija Juričana	30
4.6. Prikaz glasača Mislava Kolakušića	31
4.7. Prikaz glasača Miroslava Škore	32
4.8. Prikaz glasača Darija Juričana, Mislava Kolakušića i Miroslava Škore	33
4.9. Prikaz glasača Dejana Kovača	34
4.10. Prikaz glasača Katarine Peović	34
4.11. Prikaz glasača Dalije Orešković	35
4.12. Prikaz glasača Ivana Pernara	35
4.13. Prikaz glasača koji su odabrali odgovor „Ne izlazim“	36
4.14. Prikaz glasača koji su odabrali odgovor „Križam listić“	36
4.15. Prikaz glasača koji su odabrali odgovor „Križam listić“ ili „Ne izlazim“	37
4.16. Usporedba raspodjele glasača Kolinde Grabar-Kitarović i Zorana Mi- lanovića nakon prvog i drugog izbornog kruga	45
4.17. Prikaz glasača koji su odabrali odgovor „Križam listić“ ili „Ne izlazim“	46
4.18. Prikaz važnosti značajki za prvu anketu	49
4.19. Prikaz važnosti značajki za drugu anketu	49
4.20. Primjer stabla s dubinom 2	51

Analiza političkih preferencija studenata metodama strojnog učenja

Filip Jurec, Mariela Markovinović, Marta Novak, Vedrana Raguž, Ivana Žeger

Sažetak

U ovom radu opisano je korištenje strojnog učenja za analizu političkih preferencija studenata. Provedene su i obrađene ankete za dva kruga izbora putem kojih su ispitanici izrazili svoje političke stavove. Kao pomoć pri sastavljanju anketnih pitanja korištena je fokus grupa. Metode korištene za analizu podataka su analiza glavnih komponenti i algoritam slučajnih šuma. Dvije provedene ankete analizirane su pojedinačno uzevši u obzir grafičke i tablične rezultate dobivene prethodno navedenim metodama. Najbolje rezultate predviđanja daje algoritam slučajnih šuma s prosječnom točnosti od oko 90%.

Ključne riječi: anketa, strojno učenje, analiza glavnih komponenti, slučajne šume, predsjednički izbori

Analysis of students' political preferences using machine learning methods

Filip Jurec, Mariela Markovinović, Marta Novak, Vedrana Raguž, Ivana Žeger

Abstract

This paper presents the usage of machine learning in analysing students' political preference. Surveys were conducted for two rounds of elections through which respondents expressed their political views. Focus group has been used to help in compiling surveys. Methods used for data analysis include principal component analysis and random forest algorithm. Two conducted surveys were analyzed separately with respect to both graphic and table results obtained by the above mentioned methods. Best predictions are being made by random forest algorithm with a precision of about 90%.

Keywords: survey, machine learning, principal component analysis, random forest, presidential elections