

Predviđanje uspjeha studenata na temelju njihova pisanoga izražavanja

Rešicki, Tomislav

Master's thesis / Diplomski rad

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:200:710902>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-14**

Repository / Repozitorij:

[Faculty of Electrical Engineering, Computer Science and Information Technology Osijek](#)



**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I
INFORMACIJSKIH TEHNOLOGIJA OSIJEK**

Sveučilišni diplomski studij

**PREDVIĐANJE USPJEHA STUDENATA NA TEMELJU
NJIHOVA PISANOGA IZRAŽAVANJA**

Diplomski rad

Tomislav Rešicki

Osijek, 2018.

SADRŽAJ

1. UVOD	1
1.1. Zadatak diplomskog rada.....	1
2. PRIJAŠNJA ISTRAŽIVANJA	3
3. METODOLOGIJA	5
3.1. Ispitanici i korpus istraživanja.....	5
3.2. Uporaba alata (LIWC)	7
4. PROGRAMSKO RJEŠENJE	9
4.1. Razvojno okruženje <i>Visual studio</i>	9
4.2. <i>Windows forms</i>	9
4.3. Izrada koda	13
4.4. Testiranje koda na bazi podataka.....	18
5. REZULTATI I RASPRAVA	21
5.1. Ograničenja istraživanja	28
6. ZAKLJUČAK	34
6.1. Samoanaliza.....	35
6.2. Prijedlozi za buduća istraživanja i primjenjivost rezultata	36
LITERATURA.....	37
SAŽETAK.....	39
ABSTRACT	40
ŽIVOTOPIS	41
PRILOZI.....	42
Kod	42

1. UVOD

Visoka razina očekivanja za upis na studije uglednih sveučilišta i/ili zaposlenje u tvrtkama stvorila je potrebu za procjenjivanjem i profiliranjem uspješnih kandidata. Naše istraživanje ima za cilj proučiti potencijalne korelacije između studentskih radova i njihove ocjene s ispita. U istraživanje je uključeno 268 studenta tehničkih smjerova koji su predali svoje eseje za računalnu (LIWC) analizu. Rezultati su pokazali da su bolje ocijenjeni studenti koristili opskurniji i složeniji vokabular, opširnije rečenice te više imenica. S druge strane, lošije ocijenjeni studenti koristili su manje složene rečenice i puno jednostavniji rječnik. Neovisno o polju ili struci studija, svaki student u jednom stadiju svog studiranja, suoči se s izradom radova različitih vrsta i dužina; od kraćih eseja do diplomskih radova i doktorskih disertacija. Tijekom svih razina školovanja, studenti uče razne tehnike i stilove pisanja radova kako bi ih što bolje i kvalitetnije strukturirali. Uz edukaciju na samom studiju, na internetu se nalazi obilje materijala i savjeta za pisanje raznih vrsta eseja tako da studenti imaju profesionalno vođenje kroz ovaj, za mnoge, obeshrabrujući zadatak. No, što ako naši eseji otkrivaju više nego što je napisano u njima? Rad je podijeljen u nekoliko dijelova. Prvi dio predstavlja prijašnja istraživanja u području i opisuje metodologiju istraživanja, dok se u drugome dijelu objašnjava izrada programskoga koda i testira njegov rad na bazi studentskih eseja. Nakon analize i rasprave rezultata, koji su jedino mjerilo učinkovitosti provedenog istraživanja, doneseni su neki zaključci o varijablama koje su se pokazale značajne za potencijalno predviđanje uspjeha studenata.

1.1. Zadatak diplomskog rada

Velika tržišna konkurentnost i sve veća potreba za upisivanjem najboljih maturanata na studije stvorili su potrebu za procjenjivanjem i profiliranjem kandidata. Zadatak je ovog diplomskog rada proučiti rezultate dobivene računalnom analizom te na temelju ocjena studenata iz kolegija engleskoga jezika izabrati varijable (jezične kategorije) i postaviti parametre za svaku ocjenu time stvarajući podlogu za buduća istraživanja o predviđanju uspjeha studenata. Istraživanje će se raditi na korpusu eseja prikupljenih na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija Osijek. Eseji su napisani na engleskome jeziku, a u njima su studenti ukratko predstavili sebe, svoje ambicije, hobije i sl. Istraživanje će imati nekoliko etapa. U prvoj

će etapi prikupljeni eseji biti računalno analizirani uporabom *Linguistic Inquiry and Word Count* (LIWC) programskoga alata koji analizira riječ po riječ, uspoređuje ju sa svojim unutarnjim rječnikom, koji je prošao stručnu evaluaciju i kategorizira analizirane riječi u predodređene kategorije (kategorije se kreću od zamjenica, priloga i dugačkih riječi do riječi vezanih uz osjećaje, novac, uspjeh i sl.). U drugoj će se fazi koristiti rezultati prethodnih sličnih istraživanja i programerski postaviti jezične kategorije za koje je dokazano da su vezane uz određenu razinu uspjeha. U posljednjoj će se etapi koristiti rezultati računalne analize studentskih eseja kako bi se, na temelju programiranih jezičnih kategorija, dokazala korelacija odabranih jezičnih kategorija i polučenoga uspjeha na kolegiju Engleskoga jezika.

2. PRIJAŠNJA ISTRAŽIVANJA

Načini na koje se koriste riječi odražavaju načine nečijeg razmišljanja i osobnih interesa. Postoje sve veće tendencije za uvođenjem pisanja uvodnih eseja na preddiplomskim sveučilišnim i stručnim studijskim programima bez obzira na odabrane studijske smjerove kandidata [1-5]. Slično je i prilikom zaposlenja; pored životopisa, koji su najčešće informativnoga karaktera o obrazovanju i prošlosti budućeg kandidata, poslodavci nemaju podatke o osobnosti kandidata. Stoga se uvode kratke pisane izjave svakog kandidata o njegovoj prošlosti kao i ambicijama, prednostima i životnim ciljevima. Razlog tomu jest da poslodavci bliže upoznaju potencijalne kandidate za posao ili, u ovom slučaju, studente koji bi profilom odgovarali za određeni studijski program. Temeljna je ideja pisanja eseja da kandidat pokaže svoje sposobnosti, interese, prijašnja iskustva i buduće planove. Također, otkrivaju se njihova emocionalna stanja, zabrinutosti i slično, što je važno pri odabiru najboljih kandidata. Budući da su intervjui izazovni i dugotrajni, pisanje uvodnih eseja izvrsna je prilika da kandidati predstavljaju svoju akademsku obuku, radno iskustvo, vještine i ambicije, a također i prilika da poslodavci lakše procijene kandidate koji su potencijalno najbolji izbor. Tradicionalno, brojne studije [6-7] koje se bave pitanjem uspješnog pisanja eseja, koristile su pristup „odozgo prema dolje“ - linearni modelni pristup - usredotočujući se na unaprijed određene karakteristike za ocjenjivanje eseja. Pristup „odozgo prema dolje“ pretpostavlja postojanje jedinstvene definicije uspješnog pisanja na temelju unaprijed definiranih obrazaca kao što su struktura, jasna linija razvoja sadržaja, leksička sofisticiranost i kohezija. Međutim, vještine pisanja mogu se obučavati, stoga često tradicionalna procjena eseja koju provodi čovjek može biti previše dugotrajna, ali i subjektivna. Posljedično, novije studije [8-10] usvajaju automatsko ocjenjivanje eseja. Svrha alata za procjenu jest olakšati opterećenje evaluatora eseja, prepoznavanje pogrešaka i davanje povratne dijagnostičke informacije. Ipak, usprkos pomoći alata pri ocjenjivanju eseja, još je uvijek velika usredotočenost na način slaganja rečenica. Trenutačno tržište rada i sveučilišna konkurencija ipak zahtijevaju mnogo više od toga.

Nove studije [11-15] fokusiraju se na profiliranje kandidata na temelju njihova pisanoga izražavanja. Koncentrirajući se na sintaktičku razinu, Haswell [16] je utvrdio da napredniji autori oblikuju složenije rečenice. Skupina je autora prevođena Crossleyjem [17] zaključila da napredniji studenti koriste sintaktički složene rečenice izmjerene brojem modifikatora u rečenici kao i manje osnovnih glagolskih oblika. Nadalje, Robinson [14] i Pennebaker [15] otkrili su da su učenici koji postižu bolji uspjeh koristili složeniji rječnik i konkretne imenice koje ih

definiraju kao kategorijske mislioce koji se usredotočuju na objekte i stvari. S druge strane, koncentriranjem na radnje i promjene, učenici s lošijim uspjehom koristili su više zamjenica i glagole pa su u očima ocjenjivača dobivali dinamični predznak. Međutim, autori [14-15] nisu tvrdili da su kategorijski kandidati pametniji od dinamičnijih; samo su nagrađeni višim ocjenama jer su obrazovni sustavi postavljeni tako da nagrađuju studente za njihovo razumijevanje svijeta na kategorički način. Fredricksonova istraživačka studija [18] utvrdila je korelaciju pozitivnih emocija riječi i ocjena studentskog uspjeha. Pozitivne emocije nisu samo odraz nečijeg zadovoljstva, nego one prilično potiču pojedinca da nastavi širok raspon ciljeva i aktivnosti koji doprinose većoj ocjeni. Drugi razlog zašto je korištenje pozitivnih emocija zaradilo veću ocjenu jest da takve emocije potiču osjećaje empatije i podrške prema ostalim zaposlenicima ili studentima kao i poslodavcima. Još jedna kategorija koja je dokazano povezana s predviđanjem uspjeha jest kategorija osobne zamjenice *mi (we)*, čije korištenje ukazuje na povezivanje autora s nekim grupama ili afilijacijama čiji je član.

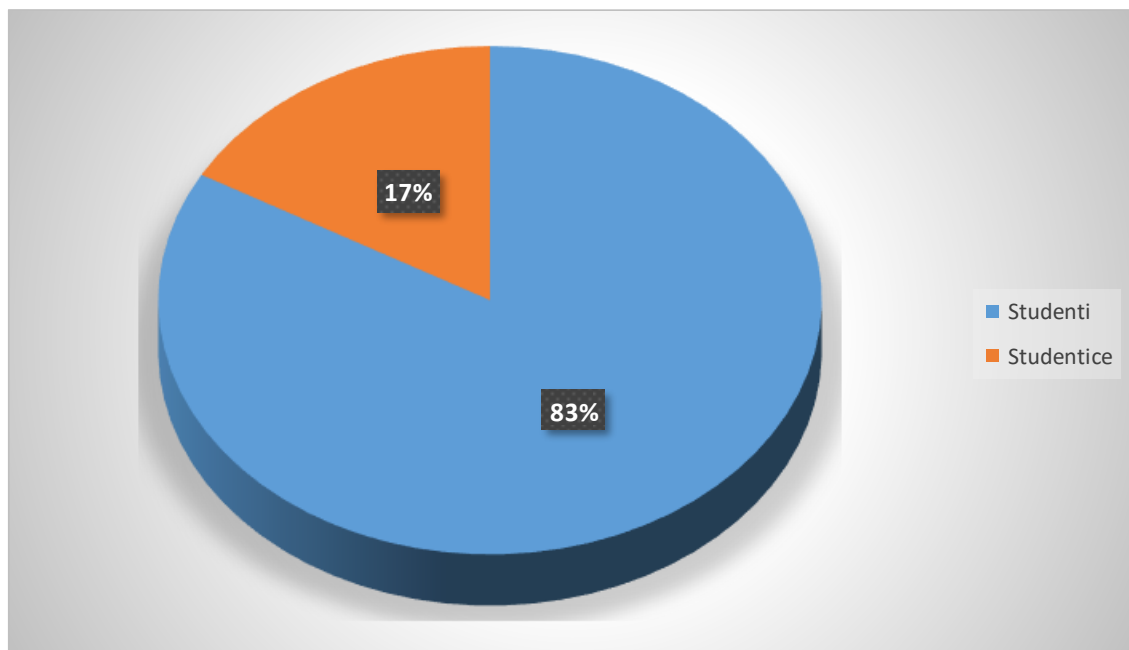
Sve veći broj istraživanja [15-17] ukazuje na novi način pristupa esejima, koji stavljaju naglasak na dobru oblikovanost eseja studenta ili kandidata za posao i njegov doprinos zajednici i izvan akademskog uspjeha, odnosno radnog mjesta.

3. METODOLOGIJA

Koristeći se rezultatima računalne (LIWC) analize koja je rađena na korpusu eseja studenata Fakulteta elektrotehnike, računarstva i informacijskih tehnologija Osijek, cilj je ovoga diplomskoga rada bio analizirati izabrane jezične kategorije, usporediti raspone svake kategorije s ocjenama studenata te postaviti granične vrijednosti za svaku jezičnu kategoriju i ocjenu u izrađenoj aplikaciji. U radu smo koristili induktivnu metodu gdje smo sustavno proučavali rezultate računalne analize eseja i ocjene studenata te krećući se od pojedinačnih slučajeva, postavili granične vrijednosti za raspon četiriju pozitivnih ocjena kod svake jezične kategorije. Razlog odabira induktivne metode jest u dostupnosti korpusa prikupljenih eseja i, za naše istraživanje, relativno primjerenom broju ispitanika koji su svojim esejima sudjelovali u istraživanju.

3.1. Ispitanici i korpus istraživanja

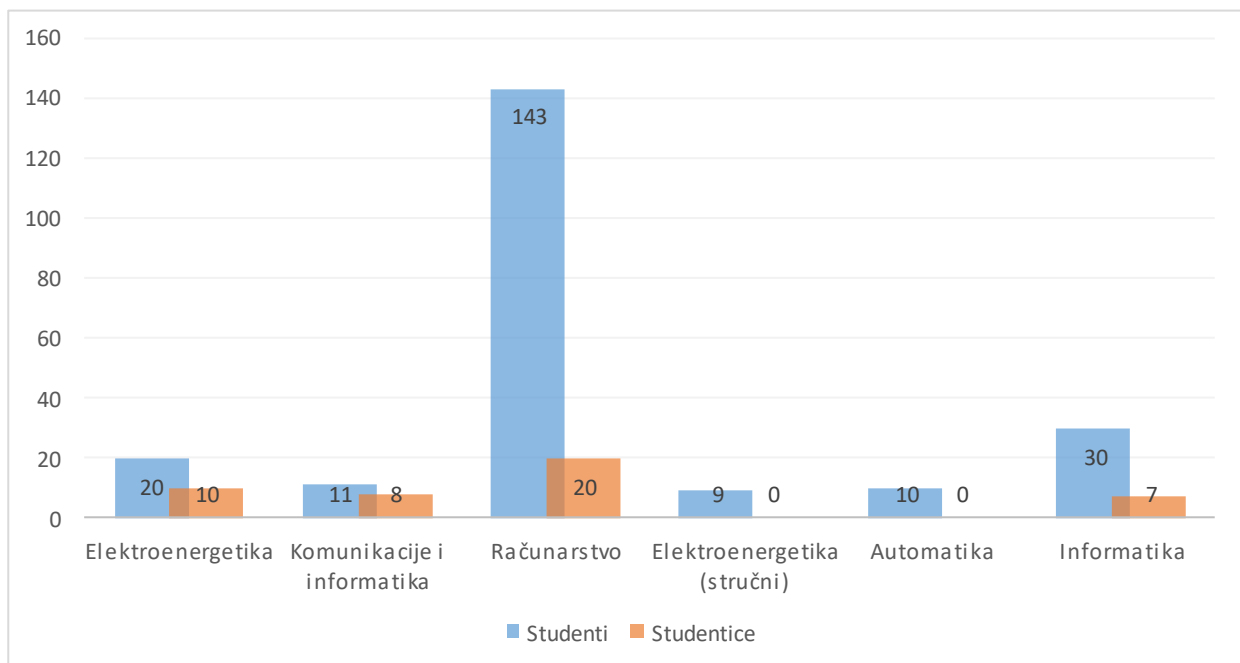
Od ukupno 512 mogućih kandidata s različitih razina studiranja tijekom 2016./2017. i 2017./2018. akademske godine, 268 studenata i studentica (52 %) dobrovoljno je napisalo eseje. Kao što je prikazano na Sl. 3.1., od 268 studenata i studentica, u istraživanju je sudjelovalo 223 studenta (83 %) i 45 studentica (17 %).



Sl. 3.1. Grafički prikaz odnosa studenata i studentica

Ovaj neravnomjeran udio studenata i studentica koji su sudjelovali u istraživanju odgovara postotku studenata i studentica koji studiraju na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija Osijek.

U istraživanju su sudjelovali studenti i studentice sa svih smjerova koji se nude na sveučilišnome i stručnome preddiplomskom studiju, što je prikazano na Sl. 3.2. Naš se konačan uzorak sastojao od 20 studenata (8,98 %) i 10 studentica (22,22 %) sa smjera elektroenergetike, 11 studenata (4,93 %) i 8 studentica (17,78 %) sa smjera komunikacija i informatike, 143 studenata (64,12 %) i 20 studentica (44,44 %) sa smjera računarstva, 9 studenata (4,04 %) sa stručnog studija elektroenergetike, 10 studenata (4,48 %) sa smjera automatike te 30 studenata (13,45 %) i 7 studentica (15,56 %) sa smjera informatike.



Sl. 3.2. Grafički prikaz odnosa studenata i studentica po smjerovima

Raspodjela studenata u odnosu na smjerove studiranja prilično je neujednačena s obzirom da je većina naših studenata bila sa smjera računarstva. Iako su svi studenti bili prilično motivirani zarađivanjem dodatnih bodova na kolegiju, istraživanje je pokazalo da su studenti i studentice računarstva bili izrazito motivirani da sudjeluju u ovom testiranju, što opravdava činjenica da je u analiziranim akademskim godinama najviše studenata upisano upravo na smjer računarstvo. Osim toga, zanimala ih je i povratna informacija o njihovim rezultatima, što ide u prilog njihovoj intrinzičnoj motivaciji za sudjelovanjem u istraživanju.

Navedenih 268 studenata i studentica sudjelovalo je u istraživanju na način da su dobrovoljno napisali esej o sebi. Podaci su prikupljeni na uzorku polaznika kolegija Engleskog jezika 1, 2 i 3 u slijednim semestrima kako bi se dobio velik broj reprezentativnih uzoraka. Eseji su napisani na engleskome jeziku jer odabrani alat za računalnu obradu podataka LIWC nema mogućnost analiziranja tekstova napisanih na hrvatskome jeziku. Eseji su morali brojati između 200 i 250 riječi, što je i standardna norma pisanoga izražavanja na engleskome jeziku na državnoj maturi koju moraju položiti svi učenici da bi stekli pravo upisa na studije. Isto tako, tekstove koji broje manje od 100 riječi, alat za računalnu obradu podataka LIWC ne može adekvatno analizirati, tako da je postavljanjem uvjeta za esej od 200 do 250 riječi ispunjen i tehnički uvjet za računalnu obradu teksta navedenim alatom. Osim tehničke upute o broju riječi, studentima je dana uputa da predstave sebe, svoje hobije, interese, strahove, budućnost i sl., no svaki je student mogao odabrati samo jedan aspekt i detaljnije ga opisati ili manje detaljno opisati sve aspekte. Studenti su se prilikom pisanja eseja mogli koristiti online i tiskanim rječnicima kao i internetskim stranicama za pomoć prilikom odabira vokabulara i drugih problematičnih područja. Studenti nisu znali koji će se aspekti njihova pisanja analizirati tako da nisu mogli utjecati na rezultate istraživanja. Zbog vremenskoga ograničenja, studenti su eseje pisali za domaću zadaću, čime su mogli utjecati na rezultate istraživanja na način da je netko drugi umjesto njih napisao esej, što je ograničenje ovoga istraživanja.

3.2. Uporaba alata (LIWC)

Nakon prikupljanja, svaki je esej skeniran za pravopisne pogreške, ispravljen i tehnički pripremljen za računalnu analizu alatom LIWC-om, što bi u slobodnom prijevodu značilo *Lingvistički upiti i brojač riječi*. LIWC su prvi puta implementirali Pennebaker i Francis 1993., a alat radi na principu da analizira tekst riječ po riječ, uspoređuje svaku riječ sa svojim internim rječnikom koji sadrži oko 4 500 riječi te izračunava stupanj do kojeg pojedini sudionik koristi različite kategorije riječi. Analizirane riječi kategoriziraju su u 80 LIWC izlaznih varijabli koje su se prethodno bile prikupile i ocjenjivale u nekoliko ocjenjivačkih faza. U ocjenjivačkim su fazama različiti brojni ocjenjivači kategorizirali riječi koje su morale biti kategorizirane u iste kategorije da bi se kategorizacija prihvatila. One riječi koje ocjenjivači nisu svrstali u jednake kategorije prolazile su ponovne ocjenjivačke faze i rasprave. Po završetku je ocjenjivačkih faza načinjen unutarnji LIWC rječnik koji sadrži 4 500 riječi (korijena i izvedenica riječi). 2001. godine performanse su alata ažurirane i vidno poboljšane, a 2007. je godine unutarnji rječnik

znatno proširenim te su dodane opcije koje se nude korisnicima. Upravo ta verzija korištena je u ovome radu [19].

LIWC radi na sljedećem principu – pronalaskom se riječi ista uspoređuje s unutarnjim rječnikom i svrstava u određene kategorije. Jedna se riječ može svrstati u više kategorije. Primjerice, riječ *razveselio se* svrstat će se u kategoriju glagola, prošlog glagolskog vremena i riječi koje se koriste za izražavanja pozitivnih emocija. 80 je kategorija u koje LIWC svrstava riječi, a kreću se od numeričkih kategorija broja riječi u rečenici, duljine riječi i interpunkcijskih znakova preko gramatičkih kategorija svih vrsta riječi (zamjenice, glagoli, prilozi, brojevi, itd.) do riječi vezanih uz emocije, novac, uspjeh, interese i sl. Ideja za stvaranja ovoga alata bila je u psihološkim istraživanja koje su provodili Pennebaker, Francis i Booth [19] koji su smatrali da sretniji ljudi više koriste riječi sreće i zadovoljstva, dok ljutiti ljudi većinom koriste riječi vezane uz gnjev i razočaranost. Također, dokazali su da korištenje funkcijskih riječi, kojima se inače ne pridaje tolika pažnja u rečenici, poput uporabe zamjenica, članova, prijedloga, negacija i sl., može puno saznati o profilu autora. S obzirom na funkcije i, nama, dostupnost ovoga alata, odlučili smo se za njegovu uporabu za računalnu analizu teksta, no za potrebe smo ovoga diplomskoga rada odabrali osam jezičnih kategorija.

3.3. Jezične kategorije

Osam jezičnih kategorija koje su odabrane za računalnu analizu su *broj riječi po rečenici*, *riječi duže od 6 grafema*, *interpunkcijski znak točka*, *zamjenica ja* i njene izvedenice poput *moj*, *mene* i sl., *osobne zamjenice*, *članovi*, *riječi vezane uz slobodno vrijeme* i *riječi vezane uz postignuća*. Razlog je odabira navedenih kategorija u prijašnjim provedenim istraživanjima koja su dokazala povezanost istih s polučenim uspjehom, a o svakoj ćemo jezičnoj kategoriji i prijašnjim znanstvenim istraživanjima detaljnije reći u poglavlju 5.

4. PROGRAMSKO RJEŠENJE

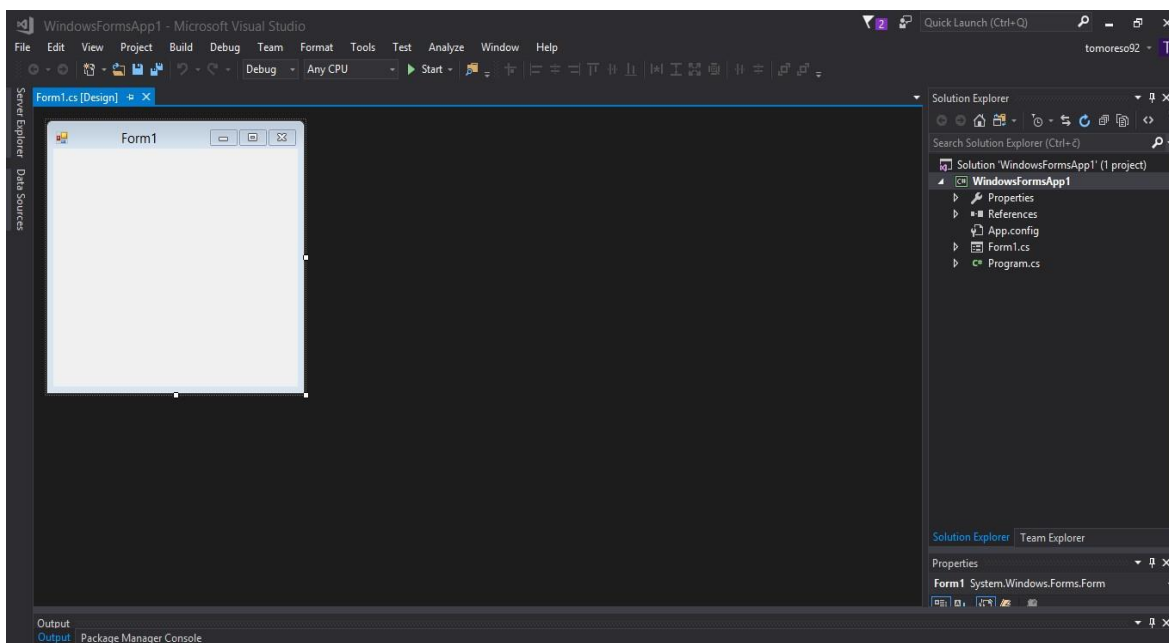
Razvojem tehnologija i mogućnosti u programerskom svijetu, stvorio se vrlo širok spektar alata i rješenja pomoću kojih se učinkovito može dobiti kvalitetan ishod. Tako je danas mnoštvo internetskih i mobilnih aplikacija kojima je moguće riješiti i najsloženije probleme. U mnoštvu tehnologija, važno je pronaći adekvatnu kako bi se dobilo na jednostavnosti, učinkovitosti, čitljivosti, razumljivosti kao i održivosti projekta. Za izradu aplikacije kojom ćemo analizirati bazu podataka, korišteno je vrlo popularno i moćno razvojno okruženje *Visual studio*.

4.1. Razvojno okruženje *Visual studio*

Visual studio razvojno je okruženje namijenjeno razvoju raznih vrsta aplikacija na *Windows* platformi. Razvoj je ovoga okruženja konstantan, stoga je sve osim običnoga uređivača za razvoj aplikacija. Širok spektar mogućnosti *Visual studija* nudi korisniku upoznavanje s poznatim tehnologijama u programerskim razgovorima. Tako osim objektno-orijentiranog programiranja i programiranja u programskom jeziku C++, *Visual studio* podržava pisanje kodova u Java Scriptu, ASP.net, HTML i CSS editore i još mnoge slične varijante programiranja. Za potrebe je ovoga rada korištena C# tehnologija koja sama po sebi ima veliku prednost pri izgradnji *Windows* i internetskih aplikacija. Bitna je stavka koja razlikuje *Windows* od konzolnih aplikacija samo sučelje, a alat za izradu *Windows* aplikacija jest *Windows forms*.

4.2. *Windows forms*

Windows forms aplikacija veoma je jednostavna za pokretanje. Otvorimo *Visual studio* te odabiremo projekt za *Windows forms* aplikaciju. Nakon što smo pokrenuli, dobijemo prozorčić koji se naziva forma. Unutar forme oblikujemo izgled naše aplikacije. Sučelje nakon kreiranja *forms* aplikacije prikazano je na slici 4.1.

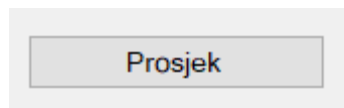


Sl. 4.1. Sučelje izrađene forms aplikacije

Nakon što je uspješno stvorena forma za aplikaciju, potrebno je dodati alate koje ćemo programirati kako bismo dobili uspješnu cjelinu kao konačno rješenje. Mnogo je naredbi i opcija pomoću kojih se može na različite načine doći do jedinstvenog, a opet približno jednakog rješenja. Na slici 4.1. vidljivo je sučelje kompletne aplikacije i njene mogućnosti, a prvi je korak u izradi dodavanje potrebnih alata. Pritiskom na pregled (eng. *View*), otvara se lista prozora od kojih je za izradu naše aplikacije najzanimljivija paleta s alatima (eng. *Toolbox*), gdje nam se nalaze alati za formu. Jedno je od najosnovnijih svojstava svakog alata ime koje je svojstveni identifikator istoga. Sličnost je ovog svojstva u dodjeljivanju imena varijabli u programskome kodu (bez razmaka i sličnih pravila).

U aplikaciji su korišteni sljedeći alati:

- a) Gumb (engl. *button*) - jedan od najjednostavnijih alata koji korisniku nudi mogućnost pokretanja (zaustavljanja) određenog procesa ili akcije. Programerskim žargonom rečeno, pokreće određeni segment u kodu. Specifičnost je gumba svojstvo *text* koje daje naziv gumbu. Česte su varijante naziva na gumbovima *start*, *stop*, *pokreni*, *zaustavi*, *izračunaj* i slično. Na slici 4.2., vidljiva je funkcija gumba koji je nazvan *prosjek*. Uloga je gumba u ovome slučaju da pritiskom na njega program prođe kroz sve ocjene koje smo dobili i generira konačnu ocjenu na temelju aritmetičke sredine.



Sl. 4.2. Izgled gumba u forms aplikaciji

- b) Oznake (engl. *label*) - alat koji većinom nema određenu programersku važnost per se, ali je vrlo koristan kako bismo se lakše snašli u aplikaciji; konkretno, opisuje značenje određenog dijela aplikacije. Oznaka je korištena kao tekstualni prikaz naziva vrste riječi, kao i raspona u kojemu korisnik može unijeti broj. Osim toga, oznake se koriste kao podnaslovi i naslovi. Na slici 4.3. vidimo široku primjenu oznaka: *Vrsta riječi*, *Raspon*, *Unesi vrijednost*, *Ocjena*, *Words per sentence* i *11,00-37,50*. Oznaka *vrsta riječi* služi za popis svih kategorija koje ćemo koristiti za analizu, a pomaže nam pri lakšem snalaženju u situacijama kada obrađujemo podatke. Oznaka *raspon* pomaže korisniku koji analizira na način da prikazuje u kojim je granicama moguće unijeti vrijednost, odnosno da sve vrijednosti koje su unesene izvan raspona koji je upisan ne ulaze u analizu.

Vrsta riječi	Raspon	Unesi vrijednost	Ocjena
Words per sentence	11,00 - 37,50	<input type="text"/>	<input type="text"/>

Sl. 4.3. Izgled oznaka u forms aplikaciji

- c) Tekstualno polje (engl. *textbox*) - ovaj alat omogućuje korisniku unos podataka koji će se koristiti za obradu. Dobra je karakteristika tekstualnoga polja da korisnik može ručno unositi svoje vrijednosti, ali ih i, po potrebi, mijenjati u trenutku kada je aplikacija pokrenuta. Konkretno, vratimo se na sliku 4.3. gdje vidimo dva tekstualna polja. U prvo polje unosimo vrijednost koja se treba obraditi u zadanom rasponu. Drugo polje kao odgovor daje vrijednost (ocjenu) koja je u rasponu unesene vrijednosti.

Ubacivanjem svih potrebnih alata i graničnih vrijednosti za rad aplikacije, dobili smo testnu formu. Na slici 4.4. prikazano je okvirno sučelje sa zadanim graničnim vrijednostima unutar kojih korisnik postavlja pojedinačnu vrijednost za svaku vrstu riječi. Za početak, izabrani su rasponi za 8 vrsta riječi zbog usporedbe funkcionalnosti koda na manjem uzorku.

Vrsta riječi	Raspon	Unesi vrijednost	Ocjena	Vrsta riječi	Raspon	Unesi vrijednost	Ocjena
Words per sentence	11,00 - 37,50	<input type="text"/>	<input type="text"/>	Auxiliary verb	4,37 - 15,49	<input type="text"/>	<input type="text"/>
Pronoun I	5,51 - 14,85	<input type="text"/>	<input type="text"/>	Past	0,00 - 6,92	<input type="text"/>	<input type="text"/>
Pronoun We	0,00 - 5,26	<input type="text"/>	<input type="text"/>	Present	3,16 - 14,92	<input type="text"/>	<input type="text"/>
Article	0,00 - 8,40	<input type="text"/>	<input type="text"/>	Future	0,00 - 4,24	<input type="text"/>	<input type="text"/>
Verb	7,29 - 18,97	<input type="text"/>	<input type="text"/>	Adverb	1,15 - 8,70	<input type="text"/>	<input type="text"/>
Achievement	0,00 - 6,88	<input type="text"/>	<input type="text"/>	Preposition	7,49 - 19,25	<input type="text"/>	<input type="text"/>
Leisure	0,00 - 7,63	<input type="text"/>	<input type="text"/>	Conjunction	3,43 - 11,85	<input type="text"/>	<input type="text"/>
Period	3,34 - 9,49	<input type="text"/>	<input type="text"/>	Quantifiers	0,47 - 8,40	<input type="text"/>	<input type="text"/>
				Numbers	0,00 - 4,15	<input type="text"/>	<input type="text"/>
				Social	0,91 - 11,95	<input type="text"/>	<input type="text"/>
				Six-letter words	10,67 - 33,20	<input type="text"/>	<input type="text"/>
				Personal pronouns	6,36 - 17,65	<input type="text"/>	<input type="text"/>

Sl. 4.4. *Kompletna prvobitna forma aplikacije*

Programerski dio funkcionalnosti i povezanosti između pojedinih tekstualnih polja za unos detaljno je opisan u sljedećem poglavlju. Na slici 4.5. prikazan je jedan nasumičan unos podataka kako bi se prikazala ispravnost koda. Korisnik pokreće aplikaciju pritiskom na *start* te se aktivira cijela forma. Nakon pokretanja aplikacije, unosi se vrijednost u svako tekstualno polje pored oznake za raspon. Kada korisnik unese vrijednost koja je u rasponu, automatski se popunjava sljedeće tekstualno polje za ocjenu.

Vrsta riječi	Raspon	Unesi vrijednost	Ocjena
Words per sentence	11,00 - 37,50	15	2
Pronoun I	5,51 - 14,85	9	4
Pronoun We	0,00 - 5,26	0,79	3
Article	0,00 - 8,40	2,46	2
Verb	7,29 - 18,97	7,31	5
Achievement	0,00 - 6,88	2,65	3
Leisure	0,00 - 7,83	7	3
Period	3,34 - 9,49	5,55	4

Otvori datoteku Prosjeak 3.25

Ocjena: 3

Sl. 4.5. *Unos nasumičnih podataka za provjeru funkcionalnosti koda*

Vidljivo je kako program pritiskom na gumb *prosjeak* računa aritmetičku sredinu ukupnog broja ocjena i zaokružuje ukupan prosjek na dvije decimale. Ova je stavka vrlo važna u analizi

uspješnosti kod onih studenata kojima je ocjena na granici. Kao i u ocjenjivanju, tako i u aplikaciji, ocjena se generira bez decimala pa je program postavljen da prepozna o kojoj se ocjeni radi iz prosjeka pomoću decimala i upiše ju.

4.3. Izrada koda

Nakon izrađene forme aplikacije, potrebno je definirati svaku od pojedinih kategorija koje će se koristiti za analizu. Veći broj kategorija koje se uzimaju u obzir daju veću vjerojatnost ispravnosti koda i same procjene. Analizirali smo osam kategorija i to *WPS (Words per sentence)*, *SLW (Six letter words)*, *Period*, *Personal Pronouns*, *Pronoun I*, *Article*, *Leisure* i *Achievement*. LIWC analizom prikupljeni su podaci za 80 različitih kategorija. Svaki kandidat ima 2 upisane verzije ocjene i to iz dviju kontrolnih zadaća i ukupnu ocjenu na kraju kolegija. Osim toga, popunjena je vrijednost izvučena iz LIWC računalne analize za svaku od kategorija koju smo analizirali. Na taj način, u jednom dokumentu imamo pregled za svih 268 kandidata koji se analiziraju. Vrijednosti su prikazane u Excel formatu po stupcima i redcima. Izgled Excela s popunjenim kandidatima i njihovim rezultatima vidljiv je na slici 4.6.

Student	Branch	Gender	REgrade	Fingrade	WC	WPS	Sixltr	Dic	Funct	Pron	Ppron	I	We	You	Heshe	They
Abraham Kostic - Racunarstvo.doc	3	1	3	3	282,00	23,50	16,67	88,65	54,96	18,44	14,54	13,83	0,00	0,00	0,35	0,35
Adam Martinek - Elektroenergetika.	1	1	4	4	201,00	20,10	23,88	85,07	56,22	14,43	8,46	6,47	0,00	1,49	0,50	0,00
Adrian Gere - Informatika.doc	6	1	3	3	435,00	43,50	11,95	89,66	63,45	20,23	13,79	11,95	0,23	1,15	0,23	0,23
Alisa Koprivic - Racunarstvo.doc	3	2	3	3	312,00	24,00	22,76	83,33	53,21	14,42	11,54	11,54	0,00	0,00	0,00	0,00
Anamarija Blavicki - Komunikacije.doc	2	2	5	5	231,00	23,10	23,81	86,58	55,84	16,88	12,99	11,69	0,43	0,43	0,00	0,43
Andreas Boc - Informatika.doc	6	1	2	2	218,00	13,62	14,22	90,83	58,72	20,18	14,68	12,84	0,00	0,00	1,83	0,00
Ante Coric - Informatika.doc	6	1	3	2	163,00	16,30	19,63	87,73	53,99	15,34	12,27	10,43	0,00	1,23	0,00	0,61
Ante Ravnjak - Automatika.doc	5	1	3	4	219,00	15,64	18,72	88,13	62,10	15,98	9,13	4,57	1,37	2,74	0,00	0,46
Antonio Falak - Racunarstvo.doc	3	1	4	4	185,00	16,82	27,03	81,62	52,43	12,43	9,19	7,03	0,54	0,54	0,00	1,08
Antonio Jambresic - Racunarstvo.doc	3	1	4	4	218,00	19,82	18,35	89,91	54,59	14,22	11,93	9,17	0,46	1,38	0,46	0,46
Antonio Popic - Informatika.doc	6	1	4	4	235,00	13,06	16,60	93,19	60,43	18,72	11,91	10,64	0,85	0,00	0,00	0,43
Antonio Vrbic - Racunarstvo.doc	3	1	4	4	247,00	20,58	17,41	92,71	61,54	16,19	10,53	7,69	0,81	0,40	1,62	0,00
Antun Cicic - Racunarstvo.doc	3	1	4	4	247,00	19,00	17,41	85,02	58,30	16,60	10,53	9,72	0,00	0,00	0,81	0,00
Antun Kakuk - Racunarstvo.doc	3	1	5	4	206,00	18,73	20,87	88,35	55,83	15,53	9,71	9,71	0,00	0,00	0,00	0,00

Sl. 4.6. Prikaz Excel tablice s podacima kandidata

Kao što je prikazano na slici 4.6., svaki student ima svoje vrijednosti koje su jedinstvene i kroz njih se definira konačna ocjena analizom i graničnim vrijednostima koje se postavljaju za svaku kategoriju. Vrijednosti nisu ovisne jedne o drugima; svaka je zasebna i ne temelje se na principu „veća vrijednost veća ocjena“ i slične varijante. Granične vrijednosti kategorija rađene su na temelju računalne analize programskim alatom LIWC-om, a izbor je kategorija uvrštenih u analizu rađen uzimajući u obzir prethodna slična istraživanja. Kategorije imaju svoje težinske vrijednosti koje se prikupljaju za konačnu analizu kao i ocjenu. Pojednostavljeno, izvršena je

analiza za svih 268 kandidata i tražene su granične vrijednosti koje bi dale najtočniju i najprecizniju analizu. Problem kod ovakvih analiza jest što na velikom uzorku kandidata postoji veliko odstupanje od vrijednosti. Samim time mogući su različiti događaji koji utječu na konačnu ocjenu. Primjerice, dva kandidata imaju jednaku vrijednost u zadanoj kategoriji, ali u konačnici jedan je kandidat dobio ocjenu *izvrstan*, a drugi kandidat ocjenu *dovoljan*. Kod takvih situacija, jasno je da ostale kategorije imaju svoju ulogu.

Princip prikaza forme aplikacije opisan je u prethodnome poglavlju, stoga je potrebno programirati kod koji će pouzdano analizirati podatke dobivene računalnom LIWC analizom. Prvi je korak u izradi koda za potrebe analize bio pronaći model po kojemu će se dobiti povezanost između pojedinih dijelova u formi. Tako je, nakon postavljanja svih potrebnih alata i određivanja kategorija koje ćemo analizirati, bilo potrebno povezati ključne alate. Postavljanjem alata oznaka (engl. *label*), dobili smo tekstualne vizualne opise kategorija, raspona u kojem se nalazi ocjena te polja za unos vrijednosti i ocjene. Oznaku nije potrebno dodatno programirati, stoga je njena uloga prilično jednostavna. Sljedeći je korak postaviti nazive svih tekstualnih polja (engl. *textbox*) kako bi snalaženje u kasnijemu pisanju bilo jednostavnije. Tako je, primjerice, za kategoriju *slobodno vrijeme* (engl. *leisure*) naziv *textBoxLeisure*, i po sličnom modelu za sve ostale kategorije. Tekstualna polja u aplikaciji služe da bi se u njima unio broj iz raspona koji je dodijeljen ili, u ovom konkretnom slučaju, automatski se popunila vrijednost za spomenutu kategoriju iz Excel dokumenta. Dakle, funkcija tekstualnih polja za oznaku *unesi vrijednost* jest prikaz brojčanih vrijednosti koju je dobio svaki student. Iduće je tekstualno polje koje se koristi u aplikaciji polje za ocjenu. Stoga se naziv *textBoxLeisureOcjena* odnosi na ocjenu koju student dobije iz polja kategorije (u ovom slučaju) *leisure*. Povezanost je između tekstualnih polja veoma bitna jer je potrebno definirati da vrijednost iz polja unesene vrijednosti kategorije odgovara tekstualnom polju za ocjenu. Pojednostavljeno, vrijednost koja se unese u polje *unesi vrijednost* automatski generira ocjenu u polju *ocjena* za određenu kategoriju. Najvažnije je u ovom automatizmu pronaći granične vrijednosti za svaku pojedinačnu ocjenu kako bi analiza bila što preciznija. Na slici 4.7., prikazan je kod koji služi za prikaz povezivanje tekstualnih polja i graničnih vrijednosti za kategoriju *leisure*.

```

private void textBoxLeisure_TextChanged(object sender, EventArgs e)
{
    double broj;
    if (double.TryParse(textBoxLeisure.Text, out broj))
    {
        double ocjena = (broj >= 2.009 && broj <= 3.001) ? 2 : 1;
        if (broj >= 3.009 && broj <= 7.83)
            ocjena = 3;
        if (broj >= 1.01 && broj <= 2.001)
            ocjena = 4;
        if (broj >= 0.00 && broj <= 1.009)
            ocjena = 5;
        if (broj > 7.83)
            ocjena = 0;
        textBoxLeisureOcjena.Text = ocjena.ToString();
    }
}

```

Sl. 4.7. Povezivanje tekstualnih polja i graničnih vrijednosti za jednu kategoriju

Na slici 4.7. vidljivo je kako su ocjene određene rasponom vrijednosti i to na način da je u kategoriji *leisure* za ocjenu *dovoljan* raspon od 2,01 do 3,00, *dobar* od 3,01 do 7,83, *vrlo dobar* od 1,1 do 2,00 te *izvrstan* od 0 do 1,00. Bitna napomena jest da su sve vrijednosti studenata u tim spomenutim rasponima. Konkretno, u ovoj kategoriji nemoguć je raspon 7,84, stoga je u kodu za raspon veći od 7,83 ocjena 0. Kod čita decimalni broj (*double* tip podataka) iz Excel tablice i ovisno u kojem je rasponu vrijednost, definira ocjenu te ju upisuje u tekstualno polje za ocjenu te kategorije. Svaka je kategorija definirana kolekcijama, odnosno svaki *textbox* predstavlja pojedinu kategoriju. Varijabla *tbCollection* sprema kategorije, kao što je vidljiva na slici 4.8. te će njena svrha biti objašnjena u nastavku rada.

```

private IEnumerable<TextBox> GetCollection()
{
    var tbCollection = new Collection<TextBox>();
    tbCollection.Add(textBoxWPSocjena);
    tbCollection.Add(textBoxPronounIOcjena);
    tbCollection.Add(textBoxArticleOcjena);
    tbCollection.Add(textBoxAchievementOcjena);
    tbCollection.Add(textBoxLeisureOcjena);
    tbCollection.Add(textBoxPeriodOcjena);
    tbCollection.Add(textBoxSLWOcjena);
    tbCollection.Add(textBoxPPOcjena);
    return tbCollection;
}

```

Sl. 4.8. Dodavanje kategorija u kolekciju

Nakon što su dodijeljene ocjene pojedinačno po kategorijama, potrebno je finalizirati analizu ukupnom ocjenom. Aritmetička sredina definira se na način da zbraja sve vrijednosti iz tekstualnih polja *ocjena* te ih dijeli na broj kategorija koje se analiziraju. Kod prepozna je u kojim je poljima upisana ocjena i samo njih uzima za izračun. Polje u koje se unosi decimalna vrijednost nazvana je *textBoxProsjeck*. Na slici 4.9., prikazana je funkcija za izračunavanje aritmetičke sredine.

```
private float Average()
{
    var tbCollection = GetCollection();

    float sum = 0;
    int br = 0;

    foreach(var item in tbCollection)
    {
        if(!String.IsNullOrEmpty(item.Text))
        {
            br++;
            sum += float.Parse(item.Text);
        }
    }
    return sum / br;
}
```

Sl. 4.9. Izračun aritmetičke sredine ocjena iz svih kategorija

Metoda *average* (prosjeck) služi za izračun prosjeka ukupnih ocjena koje su unesene. Kako bi se grupirali *textbox-ovi* s ocjenama, koristimo metodu *GetCollection*. Nakon grupiranja podataka, oni se spremaju u varijablu *tbCollection* te postavljamo početne parametre. Sljedeća je stavka petlja kroz koju će program proći i provjeravati količinu unesenih podataka. To se izvršava kroz *foreach* petlju. *foreach* petlja koristi se za rad s nizovima, prolazi kroz svaki element u istom te obavlja zadanu naredbu. U konkretnom slučaju, ova petlja dohvaća svaki *textbox* iz kolekcije te provjerava kroz naredbu *if* ima li dohvaćeni *textbox* vrijednost *null* ili je prazan. Ukoliko nije prazan, zaključuje se kako se u *textboxu* već nalazi neka vrijednost koja opisuje ocjenu za analogni element ocjenjivanja. Nakon što je odrađena provjera, varijabli *sum* (zbroj) dodajemo vrijednost te se varijabla *br* uvećava za 1. Tako ćemo za svaku unesenu ocjenu (iz *textboxa*) povećavati vrijednost varijable *br*. Kada *foreach* petljom prođemo kroz sve elemente, metoda *average* vraća vrijednost *sum/br*, što predstavlja zbroj svih ocjena (numeričke

vrijednosti iz textbox-ova) te ih dijeli s brojem varijabli *br* odnosno broja *textbox-ova* koji su popunjeni ocjenama. Nakon što smo napravili izračun aritmetičke sredine, potrebno je napraviti funkciju koja će prepoznavati ocjenu koja se generira iz prosjeka. Na slici 4.10. prikazan je kod koji će biti zadužen za prikaz ocjene.

```
private void buttonProsjek_Click(object sender, EventArgs e)
{
    float srednja = Average();
    float cjelobrojno = srednja * 100;
    int br = (int)cjelobrojno % 100;
    br /= 10;
    int ocjena = (int)srednja;
    if (br >= 5)
    {
        ocjena += 1;
    }
    if (float.IsNaN(srednja) || float.IsInfinity(srednja))
    {
        textBoxProsjek.Text = "Krivo";
        labelOcjena.Text = "Ne valja";
    }
    else
    {
        textBoxProsjek.Text = Math.Round(srednja, 2).ToString();
        labelOcjena.Text = ocjena.ToString();
    }
}
```

Sl. 4.10. Zaokruživanje ocjena na temelju decimalne vrijednosti

Kao što je vidljivo na slici 4.10., klikom na *buttonProsjek* generira se poziv za izračun konačne ocjene. Varijabla *srednja* predstavlja nam rezultat metode *average* te se ta vrijednost u varijabli *cjelobrojno* množi sa 100 kako bi se dobio broj koji će predstavljati cjelobrojnu vrijednost varijable *srednja* te se sprema u varijablu *cjelobrojno*. U sljedećem koraku, izračunava se modul (ostatak) dijeljenja vrijednosti varijable *cjelobrojno* sa 100 te još dodatno sa 10 kako bi se dobila desetinka decimalnog broja, koja je vrlo važna za izračun prosjeka. Decimalna vrijednost koja se generira daje svoj rezultat koji se zatim uspoređuje s brojem 5. Ako je vrijednost jednaka ili veća od broja 5, ukupna se zaokružena ocjena uvećava za jedan. Primjerice, generira se ocjena 4,24. Prva je decimala broj 2, koji je manji od broja 5, stoga ukupna ocjena ostaje 4. U slučaju da je ocjena 4,66, tada se, također, uspoređuje prva decimala s brojem 5. Ovdje je to broj 6, stoga se ukupna ocjena povećava za 1, odnosno ukupna ocjena iz prosjeka 4,66 jest 5. Ukoliko ne postoji niti jedan *textbox* u metodi *average*, a korisnik želi provjeriti prosjek, potrebno je poslati obavijest o pogrešci. Tako je u ovom slučaju kod dijeljenja s 0 rezultat metode *NaN* (eng. *Not a number*) ili beskonačan. Stoga, metodama *IsNaN* i *IsInfinity*

ispisujemo poruku o čemu se radi. Ukoliko srednja vrijednost nije *NaN* ili pak beskonačna, *textBoxProsjek* će poprimiti vrijednost varijable zaokruženu na dvije decimale pomoću metode *Math.Round*, a *labelOcjena* vrijednost zaokružene ocjene. Rješenje s implementiranom tablicom stvarnih podataka za metodu *average* prikazano je na slici 4.11.

The screenshot shows a data entry interface with the following components:

- Form:** A table with 4 columns: 'Vrsta riječi', 'Raspon', 'Unesi vrijednost', and 'Ocjena'. Rows include categories like 'Words per sentence', 'Pronoun I', 'Article', 'Achievement', 'Leisure', 'Period', 'Six-letter words', and 'Personal pronouns'. Each row has a range and a text input field for a value.
- Buttons:** 'Otvori datoteku' and 'Prosjek'.
- Summary:** 'Ocjena: 4' and a display box showing '4,38'.
- Table:** A table with 10 columns: 'Student', 'Fingrade', 'REgrade', 'WPS', 'I', 'Artic', 'Achie', 'Leisu', 'Period', and 'SL'. It contains data for several students, with 'Adam Martinek' highlighted in blue.

Sl. 4.11. Prikaz funkcionalnosti metode average

Iz slike 4.11. uočava se kako je svaka kategorija poprimila vrijednost iz tablice, sumirale su se sve ocjene te je metodom *average* prikazan ukupni prosjek i ukupna ocjena. Pritiskom na gumb *Prosjek*, popunjava se *textBoxProsjek*. Iz *textBoxProsjek*, provjeravaju se decimale te se konačna ocjena upisuje u *labelOcjena*. Povezivanje Excel tablica sa stvarnim podacima ocjena studenata i kodom opisana je u sljedećem poglavlju.

4.4. Testiranje koda na bazi podataka

Nakon što je forma za popunjavanje kategorija isprobana na nasumičnim podacima te povezana s funkcijama za računanje prosjeka i konačne ocjene, potrebno je povezati kod sa stvarnim podacima. U radu je opisivano kako će se u Excel tablici nalaziti svi podaci vezani za analizu kod svakog pojedinačnog studenta. S obzirom da je pri analizi korišteno 80-ak kategorija, a odlučeno je da će se raditi s osam kategorija, bilo je potrebno prilagoditi kod da što brže i funkcionalnije analizira podatke te ih grupira na pravi način. Za povezivanje Excel tablica

i aplikacije, potrebno je bilo instalirati paket kako bi se riješila poteškoća (The Microsoft ACE.OLEDB.12.0 provider is not registered on the local machine). Nakon instaliranja ovog dodatka, programira se *dataGridView* u kojemu će se prikazivati popis studenata sa svim vrijednostima za analizu. Slika 4.12. prikazuje postavljanje parametara za prikaz podataka iz Excel tablice sortirano i prilagođeno aplikaciji.

```
private void dataGridView1_CellClick(object sender, DataGridViewCellEventArgs e)
{
    if (e.RowIndex < 0)
        return;
    int rowIndex = e.RowIndex;

    textBoxWPS.Text = dataGridView1.Rows[rowIndex].Cells[3].Value.ToString();
    textBoxPronounI.Text = dataGridView1.Rows[rowIndex].Cells[4].Value.ToString();
    textBoxArticle.Text = dataGridView1.Rows[rowIndex].Cells[5].Value.ToString();
    textBoxAchievement.Text = dataGridView1.Rows[rowIndex].Cells[6].Value.ToString();
    textBoxLeisure.Text = dataGridView1.Rows[rowIndex].Cells[7].Value.ToString();
    textBoxPeriod.Text = dataGridView1.Rows[rowIndex].Cells[8].Value.ToString();
    textBoxSLW.Text = dataGridView1.Rows[rowIndex].Cells[9].Value.ToString();
    textBoxPP.Text = dataGridView1.Rows[rowIndex].Cells[10].Value.ToString();
}
```

Sl. 4.12. Prilagodba Excel tablice aplikaciji (upis u ćelije i redove)

Nakon što je Excel dokument prikazan u formi aplikacije, potrebno je povezati kategoriju iz dokumenta s kategorijom u aplikaciji. Na slici 4.12. prikazan je kod koji definira kategorije pomoću indeksa za stupce. Tako se u prvom stupcu u dokumentu s podacima nalazi ime i prezime studenta (indeks = 0), drugi stupac predstavlja stvarnu konačnu ocjenu studenta (indeks = 1), a treći stupac konačnu ocjenu studenta iz dviju kontrolnih zadaća (indeks = 2). Nadalje, ovim slijedom, definirani su indeksi za naše kategorije, prikazani u tablici 4.1.

Tab 4.1. Vrijednosti indeksa za svaku kategoriju

Indeks	Kategorija
3	WPS (broj riječi po rečenici)
4	Pronoun I (zamjenica ja)
5	Article (član)
6	Achievement (dostignuća)
7	Leisure (slobodno vrijeme)
8	Period (točka)

9	SLW (riječi duže od 6 grafema)
10	Personal pronoun (osobne zamjenice)

Kao što je prikazano u tablici 4.1., svaki indeks predstavlja stupac. Ovakvim načinom uspostavljanja veze između Excel stvarnih podataka i aplikacije dobiva se pouzdanost programa na način da je nemoguće pogreškom zamijeniti kategorije. Aplikacija čita podatke iz svakog stupca za svakog pojedinog studenta. Ranije je u radu prikazano sučelje aplikacije, a na slici 4.11. vidimo gumb *Otvori datoteku*. Klikom na taj gumb, otvaramo naš Excel koji želimo analizirati, a njegov je kod prikazan u priložima.

5. REZULTATI I RASPRAVA

1) Words per sentence (WPS) – broj riječi po rečenici

Kategorija *broj riječi po rečenici* numerička je jezična kategorija jer, kao što joj naziv govori, daje nam informaciju o korištenom broju riječi u rečenici. Iako informativna, ovo je najproblematičnija jezična kategorija jer se temelji na prebrojavanju interpunkcijskih znakova (točka, upitnik i uskličnik) koji se koriste za označavanje kraja rečenice. Stoga se, kako bismo imali relevantne podatke za analiziranje, željeni tekst mora „pročistiti“ na način da se pročita i uklone oni interpunkcijski znakovi koji nemaju funkciju označavanja kraja rečenice. Najčešće se radi o uklanjanju interpunkcijskoga znaka točke prilikom korištenja kratica poput *itd.*, *npr.*, a ako se radi o tekstu pisanom na engleskome jeziku, najčešće su to primjeri poput *Mr.*, *U.N.* i decimalni brojevi. Drugim riječima, prilikom upotrebe navedenih kratica, akronima i decimalnih brojeva, programski alat LIWC prepoznaje interpunkcijski znak točke i dotadašnji dio rečenice broji kao cjelinu, a ostatak rečenice, ukoliko, dakako, spomenuti ne označavaju kraj rečenice, broji kao novu cjelinu neovisno o tome što prva riječ iza spomenutih nije pisana velikim početnim slovom. Naime, LIWC ne prepoznaje velika i mala početna slova, što je njegov nedostatak. Interpunkcijski se znakovi upitnika i uskličnika najčešće nalaze u umetnutome i/ili citiranome dijelu rečenice u nekim nazivima poput *Tko se boji Virginije Woolf?*, *Mamma Mia!* i sl. Stoga se za adekvatno korištenje kategorije *broj riječi po rečenici* tekst treba pripremiti uklanjajući interpunkcijske znakove koji nemaju svoju funkciju, što smo mi učinili za svaki analizirani esej.

Kategoriju *red riječi po rečenici* odabrali smo imajući u vidu prethodna istraživanja [20-22]. Naime, u istraživanjima su sudjelovali iskusni ocjenjivači koji su eseje studenata koji su koristili duže rečenice ocijenili višim ocjenama od onih eseja koji su sadržavali kraće i jednostavnije rečenice.

Nakon računalne analize odrađene programskim alatom LIWC-om, proučili smo raspone kretanja sirovih podataka za svaku ocjenu i svakog studenta koji je sudjelovao u našem istraživanju te prema dobivenim podacima, postavili granične vrijednosti podataka računalne analize za svaku pojedinu ocjenu kao što je prikazano u tablici 5.1.

Tab 5.1. Granične vrijednosti ocjena za WPS

Ocjena	Granične vrijednosti
Dovoljan (2)	11,00 – 13,00
Dobar (3)	13,01 – 15,00
Vrlo dobar (4)	15,01 – 22,00
Izvrstan (5)	22,01 – 37,50

U tablici 5.1. prikazane su granične vrijednosti za svaku od ocjena. Vrijednosti se ocjena studenata koji su sudjelovali u našem istraživanju razlikuju na način da studenti koji su imali bolje ocjene (vrlo dobar i izvrstan) koriste duže i složenije rečenice od studenata koji su polučili slabiji uspjeh (dobar i dovoljan). Ovaj je rezultat prilično očekivan jer učenici koji imaju dobro poznavanje engleskog jezika mogu sastavljati složenije i duže rečenice. Također, složenije rečenice mogu ukazivati na detaljnu razradu argumenata i primjera, kao i korištenje prijelaznih fraza koje pridonose koheziji teksta. Stoga je naše istraživanje potvrdilo rezultate prethodnih sličnih istraživanja [20-22].

2) Six letter words (SLW) – riječi duže od 6 grafema

Slično kao i kod prethodne kategorije, jezična je kategorija *riječi duže od šest grafema* numerička, odnosno programski alat LIWC prebrojava sve riječi koje se sastoje od šest i više grafema. Za razliku od prethodne kategorije, kod ove kategorije nije potrebna intervencija „pročišćavanja“ jer ništa ne utječe na broj grafema u rečenici i njihovo računalno prebrojavanje i analiziranje.

Ovu smo kategoriju odabrali za analizu postavljajući hipotezu da studenti ocijenjeni višim ocjenama (vrlo dobar i izvrstan) više koriste složenije riječi od studenata ocijenjenih nižim ocjenama (dobar i dovoljan).

Tab 5.2. Granične vrijednosti ocjena za SLW

Ocjena	Granične vrijednosti
Dovoljan (2)	12,04 – 15,87
Dobar (3)	15,88 – 18,54
Vrlo dobar (4)	18,55 – 25,88
Izvrstan (5)	25,89 – 33,20

U tablici 5.2. nalaze se granične vrijednosti za svaku ocjenu ove jezične kategorije. Kao i u kategoriji *broj riječi po rečenici*, broj grafema upućuju na složenost jezičnog izraza. Studenti ocijenjeni višim ocjenama više su koristili složene riječi iz razloga što im je vokabular bogatiji složenijim i znanstvenijim riječima. Takve leksičke stavke često su manje uobičajene, apstraktne i sofisticirane, što upućuje na šire leksičko, a samim time i šire jezično znanje. Drugim riječima, studenti ocijenjeni nižim ocjenama imaju siromašniji vokabular, vjerojatno zbog manje izloženosti čitanju kompleksnijih tekstova na stranome jeziku, pa samim time koriste kraće i češće riječi koje se usvajaju u prvim fazama učenja stranoga jezika, što potvrđuje rezultate studija [20-21] u kojima su ocjenjivači višim ocjenama nagrađivali eseje sa složenijim i opskurnijim vokabularom.

3) Period – točka

Točka je interpunkcijski znak koji se koristi za okončavanje misli i rečenica. Taj interpunkcijski znak studenti najlakše svladaju bez obzira radi li se o materinskom ili stranom jeziku jer je njegovo korištenje, barem kada se radi o funkciji okončavanja misli i rečenica, univerzalno. LIWC bilježi sve točke u tekstu, uključujući točke za decimalne brojeve i kratice, stoga smo uklonili sve nepotrebne točke, odnosno one koje ne označavaju kraj rečenica, o čemu smo govorili i u poglavlju o broju riječi u rečenici.

Po završetku računalne analize provedene programskim alatom LIWC-om, proučili smo vrijednosti sirovih podataka za interpunkcijski znak točke, usporedili ih s ocjenama naših ispitanika i postavili granične vrijednosti za korištenje ove kategorije, što je prikazano u tablici 5.3.

Tab 5.3. Granične vrijednosti ocjena za interpunkcijski znak točke

Ocjena	Granične vrijednosti
Dovoljan (2)	7,34 – 9,49
Dobar (3)	5,96 – 7,33
Vrlo dobar (4)	4,88 – 5,95
Izvrstan (5)	3,34 – 4,87

Prema graničnim vrijednostima i rezultatima naše računalne analize, studenti ocijenjeni višim ocjenama koriste manje točaka, odnosno njihovi eseji, jednakoga broja riječi kao i svi ostali, sastoje se od manje duljih rečenica, dok se eseji studenata ocijenjenih nižim ocjenama sastoje od više kraćih rečenica. Drugim riječima, studenti ocijenjeni višim ocjenama koriste kompleksnije rečenice za iskazivanje svojih misli, a studenti ocijenjeni nižim ocjenama nemaju dovoljno znanja da misli povežu u veće cjeline koristeći odnosne rečenice, veznike i tranzicijske izraze pa koriste jednostavne kratke rečenice, što je u skladu s našim rezultatima kategorije *broj riječi po rečenici*.

4) Personal pronouns – osobne zamjenice

Unatoč tome što su kratke i većini ljudi neprimjetne riječi, zamjenice mogu puno otkriti o načinu na koji ljudi percipiraju i pristupaju stvarima, no, prema istraživanju [15], mogu pretkazati i nečiji uspjeh. Naime, na temelju analize 25 000 studentskih eseja, Pennebaker i suradnici [15] otkrili su negativnu korelaciju uporabe osobnih zamjenica i ocjena studenata, odnosno oni studenti koji su koristili puno osobnih zamjenica kasnije su polučili slabiji uspjeh od njihovih kolega koji su koristili manje osobnih zamjenica. To nas je istraživanje inspiriralo da analiziramo uporabu osobnih zamjenica kod naših ispitanika.

Tab 5.4. Granične vrijednosti ocjena za osobne zamjenice

Ocjena	Granične vrijednosti
Dovoljan (2)	13,94 – 17,65
Dobar (3)	13,00 – 13,93
Vrlo dobar (4)	10,25 – 12,99
Izvrstan (5)	6,83 – 10,24

U tablici 5.4. prikazali smo granične vrijednosti uporabe osobnih zamjenica za svaku ocjenu. Prema rezultatima naše računalne analize i graničnih vrijednosti iz tablice 5.4., razvidno je da je upotreba osobnih zamjenica obrnuto proporcionalna ocjeni, što je u skladu s rezultatima istraživanja Pennebakera i suradnika [15]. Zamjenice su, po definiciji, vrlo društvene riječi, a uporaba osobnih zamjenica ukazuje na to da pisac (govornik) stavlja sebe i/ili druge ljude u fokus razmišljanja koji se definira kao dinamičan stil razmišljanja [23]. Nadalje, veća upotreba osobnih zamjenica ukazuje na narativan stil govorenja i pisanja, što uključuje iznošenje osobnih

podataka i priča [11], a to negativno korelira s uspjehom jer studenti ocijenjeni višim ocjenama vjerojatno smatraju da osobnim pričama nije mjesto u formalnome načinu izražavanja poput pisanja eseja.

5) Pronoun *I* – zamjenica *ja*

Jedna od podvrsta kategorije osobnih zamjenica jest zamjenica *ja* i njene padežne izvedenice poput *meni*, *sa mnom* i sl. Zbog velike je potrebe za istraživanjem samo te osobne zamjenice, ona izdvojena u posebnu kategoriju u posljednjoj verziji LIWC-a. Istraživanje Pennebakera i suradnika [15] ukazalo je na negativnu korelaciju uporabe osobne zamjenice *ja* i uspjeha studenata, odnosno studenti koji su više koristili zamjenicu *ja* polučili su lošiji uspjeh od studenata koji su manje bili fokusirani na sebe, stoga smo ju odlučili uvrstiti u našu analizu postavljajući granične vrijednosti za uporabu zamjenice *ja* prikazane u tablici 5.5.

Tab 5.5. Granične vrijednosti ocjena za zamjenicu *ja*

Ocjena	Granične vrijednosti
Dovoljan (2)	11,96 – 14,85
Dobar (3)	10,16 – 11,95
Vrlo dobar (4)	8,00 – 10,15
Izvrstan (5)	5,51 – 7,99

Prema graničnim vrijednostima, temeljenima na računalnoj analizi, vidljiva je obrnuta proporcionalnost ocjena studenata i upotrebe osobne zamjenice *ja*, što potvrđuje prethodno istraživanje [15]. Ukoliko se složimo s hipotezom Pennebakera i suradnika [15] da je upotreba osobne zamjenice *ja* povezana s dinamičkim stilom razmišljanja, a samim time i lošijim uspjehom, iz našega bismo istraživanja mogli zaključiti da studenti koji koriste osobne zamjenice i zamjenicu *ja* imaju razvijen dinamički stil razmišljanja, što pozitivno korelira s nižim ocjenama na kolegiju. Nadalje, neka su istraživanja [24-25] dokazala povezanost upotrebe osobnom zamjenice *ja* sa statusom govornika/autora, što se može primijeniti i u interpretaciji naših rezultata. Naime, moguće je da su studenti koji su manje koristili zamjenicu *ja* samouvjereniji i više fokusirani na obavljanje zadatka (pisanje eseja), dok studenti ocijenjeni nižim ocjenama više pribjegavaju jednostavnijem načinu izražavanja i korištenju aktivnih rečenica koje obiluju

subjektom *ja*, a manje pasiviziranim bezličnim konstrukcijama, čije se usvajanje odvija u kasnijim fazama učenja stranoga jezika.

6) Article – član

Članovi su dijelovi govora čija je funkcija identificiranje imenice, što je i razlog za proučavanje distribucije članova jer pokazuju sposobnost kategoriziranja stvari. Za razliku od visoke upotrebe zamjenica, koje su odlika dinamičkoga stila razmišljanja, velika upotreba imenica i članova odlika je kategoričkoga stila razmišljanja [11]. Prethodna su istraživanja [14, 15, 26] dokazala postojanje pozitivne korelacije korištenja većega broja članova i imenica i polučenoga uspjeha studenata, što nas je inspiriralo da proučimo isto.

Tab 5.6. Granične vrijednosti ocjena za članove

Ocjena	Granične vrijednosti
Dovoljan (2)	0,00 – 2,00
Dobar (3)	2,01 – 2,89
Vrlo dobar (4)	2,90 – 3,50
Izvrstan (5)	3,51 – 8,40

Prema rezultatima naše računalne analize i postavljenim graničnim vrijednostima za članove, prikazanima u tablici 5.6., evidentno je da su studenti ocijenjeni višim ocjenama više koristili članove i imenice od studenata ocijenjenih nižim ocjenama, što potvrđuje prethodna istraživanja [14, 15, 26]. Nadalje, uzimajući u obzir ove rezultate, ali i one iz kategorije osobnih zamjenica, možemo potvrditi hipotezu da studente ocijenjene nižim ocjenama karakterizira dinamičan stil razmišljanja, što se u govoru i pismu odražava većom upotrebom osobnih zamjenica, dok studente ocijenjene višim ocjenama karakterizira kategoričan stil razmišljanja, veća upotreba članova i imenica i formalniji način izražavanja bez prevelikog iznošenja osobnih priča i podataka [11].

7) Leisure – slobodno vrijeme

Kategorija koja je vrlo zanimljiva za analizu jest *slobodno vrijeme*. Ova kategorija uključuje spominjanje riječi koje se mogu povezati sa slobodnim vremenom poput različitih vrsta igara

(Nintendo, kartanje) i sportova (nogomet, skijanje) do hobija (pjevanje, televizija) i načina zabavljanja (shopping centri, alkohol, karaoke). S obzirom da je studentima dana uputa da se u esejima predstave spominjući, ako i koliko to žele, između ostaloga, i svoj način provođenja slobodnoga vremena, proučavanje je ove jezične kategorije i više nego opravdano, a postavljene se granične vrijednosti nalaze u tablici 5.7.

Tab 5.7. Granične vrijednosti ocjena za slobodno vrijeme

Ocjena	Granične vrijednosti
Dovoljan (2)	2,01 – 3,00
Dobar (3)	3,01 – 7,83
Vrlo dobar (4)	1,01 – 2,00
Izvrstan (5)	0,00 – 1,00

Iz tablice 5.7. razvidno je da su studenti ocijenjeni višim ocjenama manje koristili vokabular vezan uz slobodno vrijeme od studenata ocijenjenih nižim ocjenama. Ono što je zanimljivo jest da su spomenuti vokabular najviše koristili studenti ocijenjeni ocjenom *dobar*, a ne *dovoljan* kako bi se moglo pretpostaviti. Ovaj je rezultat moguće objasniti time da su studenti ocijenjeni ocjenom *dovoljan* možda smatrali da će se njihov lošiji uspjeh interpretirati kroz prizmu prevelikog posvećivanja vremena hobijima i općenito slobodnom vremenu, a ne učenju, dok studenti ocijenjeni ocjenom *dobar* nisu bili time toliko opterećeni. S druge strane, moguće je da studenti ocijenjeni višim ocjenama (vrlo dobar i izvrstan) pisanje eseja shvaćaju kao formalan vid izražavanja u kojemu se pojedinac više orijentira predstavljanju sebe kroz neke formalnije aktivnosti (akademski i poslovni interesi), a ne slobodno vrijeme.

S obzirom da nismo pronašli prethodna istraživanja koja su se bavila korelacijom upotrebe vokabulara vezanog uz slobodno vrijeme i uspjehom studenata, naši rezultati mogu poslužiti kao inspiracija za provođenje sličnoga istraživanja.

8) Achievement – postignuća

Kategorija *postignuća* uključuje riječi vezane uz uspjeh poput *realizirati*, *usavršiti*, *vještina*, *izazov*, *napredovati*, itd. Slično kao i kod prethodne kategorije, programski alat LIWC prebrojava riječi vezane uz postignuća i izračunava postotak korištenja tih riječi. S obzirom na relativnu povezanost kategorija *slobodnoga vremena* i *postignuća*, a uzimajući u obzir uputu

danu studentima, činilo nam se logičnim proučiti povezanost korištenja vokabulara vezanog uz postignuća i ocjene studenata.

Tab 5.8. Granične vrijednosti ocjena za postignuća

Ocjena	Granične vrijednosti
Dovoljan (2)	2,89 – 3,44
Dobar (3)	2,15 – 2,88
Vrlo dobar (4)	3,45 – 6,88
Izvrstan (5)	0,00 – 2,14

Prema rezultatima računalne analize, postavljene su granične vrijednosti prikazane u tablici 5.8. iz koje je vidljivo da su studenti ocijenjeni ocjenom *izvrstan* najmanje koristili vokabular vezan uz postignuća, dok su ga najviše koristili studenti ocijenjeni ocjenom *vrlo dobar*. Čest je slučaj da su uspješne osobe svjesne svog uspjeha, stoga često nemaju potrebu to dodatno naglašavati. S druge strane, studente ocijenjene ocjenom *vrlo dobar* ocjenjivači ne smatraju jednako uspješnima kao njihove kolege ocijenjene najvišom ocjenom, čega su i oni vjerojatno svjesni. Jednako tako, često je vrlo tanka granica između vrlo dobrih i izvrsnih studenata pa je moguće da su vrlo dobri studenti toga svjesni i dodatnim naglašavanjem svoga uspjeha pokušavaju „prijeći“ u kategoriju izvrsnih studenata, odnosno možda osjećaju potrebu eksplicitno naglasiti svoj uspjeh koji nije nužno i objektivno uspjeh.

Jednako kao i kod prethodne kategorije, ni za ovu kategoriju nismo pronašli prethodna istraživanja te naši rezultati mogu poslužiti kao hipoteza za buduća slična istraživanja.

5.1. Samoanaliza

Nakon provedene računalne analize programskim alatom LIWC-om, analize sirovih podataka dobivenih spomenutom analizom i postavljanja graničnih vrijednosti za osam odabranih kategorija, odlučili smo analizirati svoju uspješnost u procjenjivanju uspjeha, odnosno postignutih ocjena studenata na kolegiju *Engleski jezik*. Proučavane su jezične kategorije međusobno neovisne, ali istovremeno i ovisne u smislu što ukupni rezultat svih osam proučavanih kategorija definira konačnu ocjenu studenta. Tijekom provođenja istraživanja i postavljanja graničnih vrijednosti, bilo je zanimljivo promatrati kako granične vrijednosti samo jedne od osam kategorija mogu podići ili spustiti konačnu ocjenu. Primjerice, dva studenta imaju

jednake konačne ocjene, a pri analizi se razlikuju u samo dvjema kategorijama. U takvom slučaju, za očekivati je kako prosjek teško može biti isti, štoviše, može se dogoditi i različitost ocjena. Osim toga, postoji i različitost pri konačnim ocjenama. Kod nekih studenata ocjena postignuta na dvjema kontrolnim zadaćama i ukupna ocjena iz kolegija nisu iste. Stoga je trebalo vrlo studiozno proučiti sirove podatke računalne analize, ocjene postignute na dvjema kontrolnim zadaćama i ukupne ocjene iz kolegija kako bi se postavile što preciznije granične vrijednosti za svaku ocjenu i jezičnu kategoriju, a to je zahtijevalo dosta vremena. Uzimajući u obzir različite grupacije studentskih ocjena (ocjene iz dviju kontrolnih zadaća i konačne ocjene iz kolegija), pri analizi smo načinili tri kategorije – *pogođena ocjena*, *nepogođena ocjena* i *pogođena okvirna ocjena*.

a) Kategorija *pogođena ocjena*

U ovu su kategoriju svrstani svi oni studenti koji su imali iste ocjene na dvjema kontrolnim zadaćama i konačnu ocjenu iz kolegija, a točno smo im odredili ocjenu. Također, ovdje su svrstani i oni studenti koji su imali različite ocjene iz dviju kontrolnih zadaća od konačne ocjene iz kolegija, a točno smo im odredili jednu od tih dviju vrsta ocjena. Naime, konačna se ocjena iz kolegija temelji na ocjeni dviju kontrolnih zadaća, aktivnome sudjelovanju na nastavi, domaćim zadaćama i usmenome ispitu, stoga se kod nekih studenata, uglavnom onih kod kojih je ocjena iz dviju kontrolnih zadaća bila na granici dviju ocjena, konačna ocjena razlikuje od one postignute na dvjema kontrolnim zadaćama zbog uspjeha postignutim na drugim spomenutim elementima koji se ocjenjuju.

b) Kategorija *nepogođena ocjena*

U ovu su kategoriju svrstani oni studenti kojima nismo točno odredili ocjenu. Ono što je zanimljivo jest da se radi o velikom odstupanju od ocjene koju smo mi odredili i ocjene kojom su studenti stvarno ocijenjeni i to na način da ocjene variraju za dvije ili tri razine. Primjerice, našom je analizom određena ocjena *dovoljan*, a student je ocijenjen ocjenom *vrlo dobar*.

c) Kategorija *pogođena okvirna ocjena*

Studenti nesvrstani niti u jednu od prethodnih dviju kategorija svrstani su u ovu kategoriju na način da smo ih podijelili u dvije skupine *viših* i *nižih* ocjena. U skupinu *viših* ocjena uvrstili smo studente koji su ocijenjeni ocjenama *vrlo dobar* i *izvrstan*, dok

smo u skupini *nižih* ocjena uvrstili studente ocijenjene ocjenama *dovoljan* i *dobar*. Primjerice, ukoliko je našom analizom svih osam jezičnih kategorija izračunat prosjek 4,35, što je ocjena *vrlo dobar*, a student je zapravo ocijenjen ocjenom *izvrstan*, brojali smo to kao pogodenu okvirnu *višu* ocjenu. Drugim riječima, ukoliko je našom analizom svih osam jezičnih kategorija izračunat prosjek 4,35, a student je ocijenjen ocjenama *dovoljan* ili *dobar*, to smo brojali u prethodnu kategoriju nepogođenih ocjena.

U tablicama 5.9. do 5.11. prikazali smo uspješnost određenih ocjena našom analizom i postotak podudarnosti sa stvarima podacima/ocjenama kojima su studenti ocijenjeni na kolegiju.

Tab 5.9. Pogodena ocjena

Ukupan broj studenata	Broj pogodjenih ocjena	Postotak uspješnosti
268	167	62,31 %

U tablici 5.9. prikazana je prva kategorija *pogođena ocjena* iz čega je vidljivo da smo kod 167 studenata i studentica točno odredili ocjenu iz dviju kontrolnih zadaća koja je jednaka konačnoj ocjeni iz kolegija, što je relativno visokih 62,31 %.

Tab 5.10. Nepogođena ocjena

Ukupan broj studenata	Broj nepogođenih ocjena	Postotak neuspješnosti
268	66	24,63 %

U tablici 5.10. prikazana je druga kategorija *nepogođena ocjena*. Kod 66 studenata nismo točno odredili ocjenu. Međutim, kod 32 od tih 66 studenata veliko je odstupanje u određenoj i stvarnoj ocjeni. Točnije, kod svih je 32 studenata isti slučaj – na temelju rezultata analiza studentu je određena ocjena *vrlo dobar*, a na više je kolegija engleskoga jezika student postigao ocjenu *dovoljan*, što je veliko odstupanje.

Tab 5.11. Pogođena okvirna ocjena

Ukupan broj studenata	Broj pogođenih ocjena	Postotak uspješnosti
268	35	13,06 %

Tablica 5.11. prikazuje postotak iz kategorije *okvirno pogođena ocjena*. Kod ove je kategorije specifično to da su studenti često na granici dviju ocjena. Ponekada ulaze u raspon koji je postavljen, no češće su za decimalu ili dvije udaljene od okvirne ocjene koja im se određuje.

Tab 5.12. Ukupni podaci analize

Ukupan broj studenata	Ukupno pogođena ocjena i pogođena okvirna ocjena	Postotak uspješnosti
268	202	75,37 %

Tablica 5.12. prikazuje ukupnu analizu uspješno određenih ocjena. Kod 202 od 268 studenata i studentica točno smo ili okvirno odredili ocjenu iz kolegija, odnosno u 75,37 % smo slučajeva točno ili okvirno odredili ocjenu, što je vrlo visok postotak uzimajući u obzir vanjske čimbenike, poput toga da ocjena nije mjerilo znanja, da neki studenti rade tijekom studiranja i nemaju se vremena posvetiti učenju i realizirati svoj potencijal, nego prihvaćaju ocjene koje nisu u skladu s njihovim intelektualnim sposobnostima ili mogućnost da im je zadani esej napisao netko drugi.

Od 66 studenata kod kojih nismo točno ili okvirno odredili ocjenu, kod 10 je studenata vidljiva tendencija prema stvarnoj ocjeni, no prosjek je bio za jednu decimalu veći ili manji, stoga nisu ušli u kategoriju točno određenih ocjena. Nadalje, kod 32 je studenata specifično da im je, na temelju računalne analize, određena ocjena *vrlo dobar*, a na više su kolegija engleskoga jezika ocijenjeni ocjenom *dovoljan*, što je veliko odstupanje. U ovakvim se slučajevima dovodi u pitanje ispravnost postupka pisanja eseja. Naime, eseji su zadani kao domaća zadaća za koju su se studenti mogli koristiti raspoloživim pomagalima poput online i tiskanih rječnika. Nažalost, neki su priliku pisanja eseja u slobodnome vremenu uz sva dostupna sredstva mogli iskoristiti i

na negativan način koristeći se plagiranjem, odnosno postoji mogućnost da su neki studenti zamolili drugu osobu da im napiše esej i time utjecali na rezultate istraživanja. Nadalje, u nekoliko je slučajeva samoanalizom utvrđeno kako pojedini studenti lošije ocijenjeni imaju čak bolji prosjek od kolega koji su na kolegiju postigli visoke ocjene. U našem slučaju imamo čak 24 takvih studenata, što na uzorku od 268 ispitanika čini 8,96 %, a da nema takvih slučajeva, postotak točno određenih ocjena bio bi nešto manji od 90%.

S obzirom da smo populaciju ispitanika u poglavlju 3.1. podijelili prema spolu i smjeru studija, analizirali smo uspješnost određenih ocjena prema tim dvjema kategorijama i prikazali u tablicama 5.13. i 5.14.

Tab 5.13. Samoanaliza po spolu

M/Ž	Broj ispitanika (N)	Pogođena ocjena (N)	Postotak pogođenosti (%)
Studentice	45	39	86,67 %
Studenti	223	163	73,09 %

Tablica 5.13. prikazuje analizu točno određenih ocjena prema spolu ispitanika. Iako je broj studentica i studenata neujednačen, u skladu je s postotkom upisanih studentica i studenata na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija Osijek. Iz tablice 5.13. razvidno je kako je postotak točno određenih ocjena kod studentica veći, negoli kod studenata. Ovome rezultatu pridonosi činjenica da se veliko odstupanje točno određenih ocjena koje smo spominjali u prethodnome poglavlju (slučajevi kada je računalnom analizom određena ocjena *vrlo dobar*, a student je ocijenjen ocjenom *dovoljan*) odnosi isključivo na studente. Ako prihvatimo mogućnost da je odstupanje posljedica plagiranja, možemo zaključiti da su studenti nepošteniji i skloniji plagiranju ovoga tipa od studentica.

Tab 5.14. Samoanaliza po smjeru studija

Smjer	Broj ispitanika (N)	Pogođena ocjena (N)	Postotak pogođenosti (%)
Elektroenergetika	27	22	81,48 %
Komunikacija i informatika	19	18	94,74 %

Računarstvo	164	132	80,49 %
Elektroenergetika (stručni studij)	12	7	58,33 %
Automatika	10	4	40,00 %
Informatika	36	19	52,78 %

U tablici 5.14. prikazana je samoanaliza točno određenih ocjena prema smjerovima studiranja naših ispitanika. *Elektrotehnika, Komunikacije i informatika* i *Računarstvo* smjerovi su preddiplomskoga sveučilišnoga studija, dok su *Elektroenergetika, Automatika* i *Informatika* smjerovi na preddiplomskome stručnome studiju. Prema izračunu prosječne vrijednosti triju postotaka pogođenosti, vidimo da je prosječna vrijednosti točne određenosti ocjena na preddiplomskome sveučilišnome studiju 85,57 %, dok na preddiplomskome stručnome studiju ona iznosi samo 50,37 %, što je velika razlika. Ukoliko i dalje zadržimo hipotezu da su se neki studenti poslužili plagiranjem i predali esej koji je netko drugi napisao, iz ovih bismo rezultata mogli zaključiti da su studenti preddiplomskoga stručnoga studija skloniji plagiranju od studenata preddiplomskoga sveučilišnoga studija. S obzirom da se kolegiji engleskoga jezika na preddiplomskome stručnome studiju održavaju na prvoj godini studija, a na preddiplomskome sveučilišnome studiju na drugoj i trećoj godini studija, postoji mogućnost da studenti preddiplomskoga stručnoga studija još nisu usvojili akademska pravila ponašanja i nultu toleranciju za plagiranje, što bi objasnilo ovakav rezultat.

6. ZAKLJUČAK

Internacionalizacija i sve veći broj studenata koji upisuju i završavaju studijske programe doveli su do potrebe za profiliranjem budućih studenata, odnosno potencijalnih djelatnika. Brojni su čimbenici koji utječu na studentov uspjeh tijekom studiranja poput lošije financijske obiteljske situacije i potrebe studenta da radi tijekom studiranja, nezainteresiranost studenta za ocjene, ali i veća motiviranost studenta za postizanje boljih ocjena zbog ostvarivanja prava na stipendiju i sl. Zbog nabrojanih i sličnih čimbenika, neki studenti ne ostvaruju svoj potencijal, dok drugi postižu bolje teoretske rezultate (ocjene) uz smanjeno razumijevanje i nedostatak praktičnoga znanja. Stoga su uprave visokih učilišta i poslodavci svjesni da ocjene koje je učenik/student postigao nisu nužno mjerilo njegova znanja zbog čega se stvara potreba za drugačijim načinom vrednovanja učenika/studenta.

Cilj je ovoga diplomskoga rada bio izučiti korelaciju studentskih ocjena i načina njihova izražavanja. U istraživanju je sudjelovalo 268 studentica i studenata Fakulteta elektrotehnike, računarstva i informacijskih tehnologija Osijek. Zadatak je ispitanika bio da napišu esej od 200 do 250 riječi na engleskome jeziku i, služeći se dostupnim online i tiskanim rječnicima, predstave sebe, svoje interese, hobije i ambicije. Svi su ispitanici dobrovoljno sudjelovali u istraživanju. Prikupljeni su se eseji računalno analizirali programskim alatom LIWC-om nakon čega su proučeni sirovi podaci i određene granične vrijednosti za osam odabranih kategorija (*broj riječi po rečenici, riječi duže od 6 grafema, interpunkcijski znak točka, zamjenica j., osobne zamjenice, članovi, riječi vezane uz slobodno vrijeme i riječi vezane uz postignuća*). Granične su vrijednosti određene na temelju ocjena koje su ispitanici postigli na kolegijima engleskoga jezika, a koje su ustupile profesorice engleskoga jezika.

Na temelju računalne i usporedne analize studentskih ocjena, uočili smo da su studenti koji su postigli bolje rezultate koristili opskurniji i složeniji vokabular i duže rečenice, što je i očekivano budući da je za pretpostaviti da studenti koji imaju više znanja iz engleskoga jezika mogu koristiti kohezijska sredstva i detaljnija objašnjenja od svojih kolega koji imaju manje jezičnoga znanja. Nadalje, potvrdili smo hipoteze prijašnjih istraživanja [11, 15, 23] da studente ocijenjene nižim ocjenama karakterizira dinamičan stil razmišljanja koji se u jeziku održava većom upotrebom osobnih zamjenica, dok studente ocijenjene višim ocjenama karakterizira kategoričan stil razmišljanja koji se održava većom upotrebom članova i imenica. Kod analize kategorija koje sadrže vokabular vezan uz *slobodno vrijeme* i *postignuća*, zabilježili smo

zanimljivosti. Naime, studenti ocijenjeni ocjenom *dobar* najviše su se fokusirali na predstavljanje sebe kroz svoje slobodno vrijeme. Moguće je da se studenti ocijenjeni ocjenom *dovoljan* nisu željeli predstavljati kroz svoje slobodno vrijeme kako se njihov lošiji uspjeh iz kolegija ne bi povezivao s njihovim prekomjernim provođenjem vremena na hobijima. S druge strane, studenti ocijenjeni ocjenom *vrlo dobar* najviše su se fokusirali na predstavljanje sebe kroz svoje uspjehe i postignuća za razliku od studenata ocijenjenih ocjenom *izvrstan* koji su vjerojatno svjesni svoga uspjeha da ga ne moraju toliko eksplicitno naglašavati.

Po završenim analizama, načinili smo samoanalizu kako bismo procijenili uspjeh postavljenih graničnih vrijednosti i podudaranje sa stvarnim ocjenama studenata. Kod 202 od 268 studenata postoji podudaranje u ocjeni određenoj našim analizama i stvarno dobivenoj ocjeni na kolegiju. Drugim riječima, u 75,37 % slučajeva točno smo ili okvirno odredili ocjenu studenta, dok u 24,63 % ili kod 66 studenata nismo uspjeli u točnome određivanju ocjene. Međutim, kod 32 od 66 studenata radi se o velikom odstupanju (određena ocjena *vrlo dobar*, a stvarna *dovoljan*) zbog čega postoji velika mogućnost plagiranja domaće zadaće. Uspješnije je bilo određivanje ocjena kod studentica (86,67 %), negoli kod studenata (73,09 %). Također, određivanje je bilo uspješnije za studente koji studiraju na preddiplomskome sveučilišnome studiju (85,57 %), negoli za studente preddiplomskoga stručnoga studija (50,37 %). Dakle, ako prihvatimo mogućnost da su neki studenti plagirali svoju domaću zadaću, možemo zaključiti da su to vjerojatno bili studenti muškoga spola koji studiraju na preddiplomskome stručnome studiju.

6.1. Ograničenja istraživanja

Ovo istraživanje ima svoja ograničenja u tome što mu je ograničen korpus, odnosno što je rađeno na relativno homogenoj skupini ispitanika koji studiraju na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija Osijek, stoga se rezultati istraživanja ne mogu generalno primijeniti. Drugo je ograničenje ovoga istraživanja u načinu zadavanja domaće zadaće koja je bila podložna plagiranju, stoga ne možemo sa sigurnošću utvrditi identitet autora domaće zadaće, a samim time i relevantnost istraživanja.

6.2. Prijedlozi za buduća istraživanja i primjenjivost rezultata

Buduća istraživanja u području računalne lingvistike mogu se izravno nastaviti na naše istraživanje na način da se načini korpus s drugim ispitanicima koji će imati drugačije i/ili slične akademske afilijacije. Također, osim osam analiziranih jezičnih kategorija, naredna istraživanja mogu uključiti i druge jezične kategorije koje nudi programski alat LIWC. Rezultati se ovoga i budućih srodnih istraživanja mogu primijeniti u profiliranju potencijalnih kandidata za radno mjesto, istraživanje, stipendiju i sl. budući da smo istraživanjem postavili temelje i granične vrijednosti za jednostavno profiliranje i predviđanje uspjeha kandidata na temelju njegova pisanoga izražavanja na engleskome jeziku.

LITERATURA

- [1] E. Ferguson, A. Sanders, F. O’Hehir, F. & D. James. Predictive validity of personal statements and the role of the five-factor model of personality in relation to medical training. *Journal of Occupational and Organizational Psychology*, 73, 321-344, 2000.
- [2] L. GlenMaye & M. Oakes. Assessing suitability of MSW applicants through objective scoring of personal statements. *Journal of Social Work Education*, 39, 67-82, 2002.
- [3] S. Walfish & J. Moreira. Relative weighting of admission variables in marriage and family therapy graduate programs. *American Journal of Family Therapy*, 33, 395-402, 2005.
- [4] M. W. Rabow, J. R. Wrubel & N. Rachel. Promise of professionalism: Personal mission statements among a national cohort of medical students. *Annals of Family Medicine*, 7, 336-342, 2009.
- [5] R. Turner & S. Nicholson. Reasons selectors give for accepting and rejecting medical applicants before interview. *Medical Education*, 45, 298-307, 2011.
- [6] Y. Attali & J. Burstein. Automated essay scoring with e-rater V.2. *Journal of Technology, Learning, and Assessment*, 4(3), 2006. Dostupno na <https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1650> (13. srpnja 2018.)
- [7] D. S. McNamara, S. A. Crossley & R. Roscoe. Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45 (2), 499-515, 2013.
- [8] S. Elliott. Intellimetric: From Here to Validity. U M. Shermis & J. Burstein (ur.) *Automated essay scoring: A cross-disciplinary perspective* (str. 71-86). Hillsdale, NJ: Lawrence Erlbaum Associates, 2003.
- [9] J. Burstein, M. Chodorow & C. Leacock. Automated essay evaluation: The Criterion online writing system. *AI Magazine*, 25, 27-36, 2004.
- [10] P. Deane. On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18, 7-24, 2013.
- [11] R. E. Nisbett, K. Peng, I. Choi & A. Norenzayan. Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108, 91–310, 2001.
- [12] S. Jarvis, L. Grant, D. Bikowski & D. Ferris. Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, 12 (4), 377-403, 2003.
- [13] A. C. Graesser, D. S. McNamara, M. M. Louwerse & Z. Cai. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36, 193-202, 2004.

- [14] R. L. Robinson, R. Navea & W. Ickes. Predicting final course performance from students' written self-introductions; A LIWC analysis. *Journal of Language and Social Psychology*, 32, 481-491, 2013.
- [15] J. W. Pennebaker, C. K. Chung, J. Frazee, G. M. Lavergne & D. I. Beaver. *When Small Words Foretell Academic Success: The Case of College Admissions Essays*. PloS ONE, 9(12), 2014. Dostupno na <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4281205/> (24. srpnja 2018.)
- [16] R. H. Haswell. Documenting improvement in college writing: A longitudinal approach. *Written Communication*, 17, 307-352, 2000.
- [17] S. A. Crossley, J. Weston, S. T. McLain Sullivan & D. S. McNamara. The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28 (3), 282-311, 2011.
- [18] B. L. Fredrickson. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, 56, 218-226, 2001.
- [19] J. W. Pennebaker, M. E. Francis & R. J. Booth. *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah: Lawrence Erlbaum Associates, 2001.
- [20] S. A. Crossley & D. S. McNamara. Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35, 115-136, 2012.
- [21] L. K. Varner, R. D. Roscoe & D. S. McNamara. (2013). Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis. *Journal of Writing Research*, 5 (1), 35-59, 2013.
- [22] S. A. Crossley, R. Roscoe & D. S. McNamara. What is successful writing? An investigation into the multi ways writers can write successful essays. *Written Communication*, 31 (2), 184-214, 2014.
- [23] Y. R. Tausczik & J. W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29 (1), 24-54, 2010.
- [24] J. W. Pennebaker. *The Secret Life of Pronouns: What Our Words Say About Us*. New York: Bloomsbury Press, 2013.
- [25] E. Kacwicz, J. W. Pennebaker, M. Davis, M. Jeon & A. Graesser. Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33 (2), 125-143, 2014.
- [26] F. Heylighen, J. M. Dewaele. (2002) Variation in the contextuality of language: an empirical measure. *Foundations of Science*, 6, 293-340, 2002.

SAŽETAK

Zadatak je ovoga diplomskoga rada bio izraditi kod koji će se moći koristiti za profiliranje kandidata i predviđanje njihova uspjeha. Korpus se istraživanja sastojao od 268 eseja studenata Fakulteta elektrotehnike, računarstva i informacijskih tehnologija Osijek. Eseji su napisani na engleskome jeziku, a u njima su se studenti trebali predstaviti kroz svoje hobije i ambicije. Eseji su računalno analizirani programskim alatom LIWC-om, čiji su se rezultati uspoređivali s ocjenama studenata postignutim na kolegijima engleskoga jezika. Rezultati su se koristili za postavljanje graničnih vrijednosti koje su unošene u izrađeni kod. Na temelju podataka analiza, programerski su određivane ocjene za svakog studenta, a uspješnost takvog određivanja ocjena, na temelju jezičnoga izražavanja studenta, bila je 75,37 %. Izrađeni se kod može koristiti za predviđanje uspjeha studenata i njihovo profiliranje.

Ključne riječi:

Računalna analiza, LIWC, eseji, kod, predviđanje uspjeha

ABSTRACT

This paper aimed to develop a code which might be used to profile candidates and predict their success rates. This research corpus was based on essays submitted by 268 students studying at the Faculty of Electrical Engineering, Computer Science and Information Technology Osijek. The essays were written in the English language and they were used as a tool for the students' introduction. The corpus was analyzed using a programming tool LIWC whose raw data were compared to the participants' grades acquired at the courses of the English language. The results were used to set margin values input in the developed code. Based on the research data, each student was predicted a course grade which was compared to an actual grade a student acquired. The success rate of such prediction was 75.37%. The developed code can be used for one's profiling and success prediction.

Keywords:

Computer analysis, LIWC, essays, code, grade prediction

ŽIVOTOPIS

Tomislav Rešicki rođen je 25. 03. 1992. godine u Uppsali (Švedska). Školovanje započinje u Osnovnoj školi dr. Franje Tuđmana. Pohađao je Prvu srednju školu Beli Manastir i stekao zvanje tehničar za računalstvo. Nakon toga, završio je preddiplomski sveučilišni studij računarstva u Osijeku. Potom je završio diplomski sveučilišni studij računarstvo, smjer Procesno računarstvo u Osijeku. Posjeduje znanje engleskog jezika kao i rad na računalu u različitim programskim jezicima i alatima (Word, Matlab, AutoCad, Visual Studio itd.). Tijekom školovanja sudjelovao je na natjecanjima u matematici (osnovna škola) i osnovama elektrotehnike (srednja škola) te na rukometnim natjecanjima (2. Hrvatska rukometna liga i 4. Mađarska rukometna liga). Dosadašnje radno iskustvo stekao je radeći preko studentskog servisa. Završio je tečaj programiranja za izradu web i mobilnih aplikacija.

PRILOZI

Kod

```
private void buttonOtvori_Click_1(object sender, EventArgs e)
{
    string strExcelPathName = null;
    OpenFileDialog openFileDialog = new OpenFileDialog();
    openFileDialog.Title = "Select file";
    openFileDialog.InitialDirectory = @"c:\\";
    openFileDialog.Filter = "Excel Sheet(*.xlsx)|*.xlsx|All Files(*.*)|*.*";
    openFileDialog.FilterIndex = 1;
    openFileDialog.RestoreDirectory = true;
    if (openFileDialog.ShowDialog() == DialogResult.OK)
    {
        if (openFileDialog.FileName != "")
        {
            strExcelPathName = openFileDialog.FileName;
            string connectionString = String.Format(@"Provider=Microsoft.ACE.OLEDB.12.0;Data
Source={0};Extended Properties=""Excel 8.0;HDR=YES;IMEX=1;""", openFileDialog.FileName);
            string query = String.Format("select * from [Sheet1$]");
            OleDbDataAdapter dataAdapter = new OleDbDataAdapter(query, connectionString);
            DataSet dataSet = new DataSet();
            dataAdapter.Fill(dataSet);
            DataTable dtView = dataSet.Tables[0];
            dataGridView1.DataSource = dtView;
        }
        else
        {
            MessageBox.Show("chose Excel sheet path..", "Information", MessageBoxButtons.OK,
MessageBoxIcon.Information);
        }
    }
    dataGridView1.Columns.Add("Ocjena", "Ocjena");
}
```