UNIVERZITET U NIŠU ELEKTRONSKI FAKULTET

Katedra za računarstvo

Implementacija resenja kod detekcije laznih profila na drustvenim mrezama

Profesor: Student:

### Prof. Dr. Bratislav Predić Ivan Milojković, br. indeksa 1687

Niš, 2024.godine

# Sadrzaj

Uvod ....3

Digitalna forenzika……………………………………………………………………………[4](#_TOC_250000)

Istrazivanje problema 5

Predlozeno resenje ………………………………………………...........................................7

Implementacija……………………………………………………………………………….10

Evaluacija 16

Zakljucak 21

Literatura 22

Uvod

U danasnjoj generaciji, drustveni zivot svih nas postao je povezan sa online drustvenim mrezama svih vrsta. Ovakve platform koje dnevno koriste milijarde korisnika sirom sveta, su napravile drasticne promene u nacinu na koji vodimo I zivimo nase zivote. Prijateljstva I odrzavanje kontakata sa prijateljima, kao I pracenje njihovih aktivnosti, su na jedan korak od nas. Medjutim, sa naglom ekspanzijom I rastom drustvenih mreza, pojavili su se mnogi problemi. Jedan od problema, odnosno pojava, sa kojima se se vecina osoba suocava jesu lazni profili I lazna predstavljanja.U ovom seminarskom cu predstaviti jedan od nacina kako je moguce da se ovaj problem razresi. Ovo resenje koristi tehniku klasifikacije Random Forest Classifier da bi klasifikovao profile u dve glavne grupe: lazne I prave. Posto je ovo metoda automatskog otkrivanja, moze se jednostavno upotrebiti na drustvenim mrezama koje imaju milione razlicitih profila.

**Digitalna forenzika**

Digitalna forenzika predstavlja sistematičan pristup istraživanju digitalnih uređaja, podataka i sistema radi prikupljanja, analize i interpretacije digitalnih dokaza. Ova disciplina je kljucna u istrazivanju kriminalnih aktivnosti koje ukljucuju racunarstvo I informatiku. Ona obuhvata sirok spektar aktivnosti, sve od prikupljanja podataka sa racunara, mobilnih uredjaja I mreznih sistema, pa do analize istih.

Cilj digitalne forenzike je da pronadje I sacuva digitalne dokaze. Ovo ukljucuje striktno pridrzavanje metoda koje osiguravaju integritet podataka, kao I preciznu dokumentaciju svakog koraka u procesu prikupljanja I analize dokaza. Digitalni forenzicari koriste specijalizovane alate I tehnike za povracaj obrisanih datoteka, dekodiranje sifrovanih podataka I analiziranje mreznih komunikacija.

Postoje nekoliko glavnih oblasti digitalne forenzike:

* Forenzika racunara: Fokusira se na istrazivanje podataka sa racunara, ukljucujuci hard diskove, SSD-ove I druge vrste skladista. Vrsi se analiza operativnog sistema, datoteka, logova I drugih podataka koji mogu biti od znacaja za krajnji cilj analize.
* Mrezna forenzika: Kod ove oblasti, fokus je na proucavanje mreznog saobracaja I aktivnosti na mrezi. Ovde se vrse analize od mrezni logova, preko prikupljenih podataka sa mreznih uredjaja poput rutera, pa sve do dostupnih informacija o napadima na mrezu.
* Forenzika mobilnih uredjaja: Bavi se prikupljanjem I analizi podataka sa mobilnih telefona, tableta I ostalih prenosivih uredjaja. Ovo moze ukljucivati tekstualne poruke, istoriju poziva, aplikacija, fotografije, kao I GPS podatke uredjaja.
* Forenzika interneta: Ovde je fokus stavljen na prikupljanje I analizu podataka sa interneta, ukljucujuci drustvene mreze, e-postu, web pretrazivanja kao I online transakcije. Cilj je identifikovati aktivnosti na internetu koje mogu biti relevantne za istragu.
* Forenzika podataka u oblaku/cloudu: Proucava podatke sacuvane na udaljenim serverima I u cloud okruzenjima. Ova oblast je sve vaznija s obzirom na rastucu potrebu cloud usluga.

Digitalna forenzika je jako bitna oblast racunarstva I infromatika, jer omogucuje pracenje digitalnih tragova I otkrivanje dokaza koji bi inace bili nevidljivi. Koristeci napredne tehnologije I metode, digitalni forenzicari mogu rekonstruisati dogadjaje, identifikovati ucesnike I pruziti kljucne dokaze ukoliko je u pitanju vodjenje krivicnih postupaka.

**Drustvene mreze**

Drustvena mreza je, ukratko receno, websajt na kojem svaki korisnik ima profil I moze ostati u kontaktu sa prijateljima, deliti svoja iskustva, I upoznavati, odnosno slati poruke drugim korisnicima. Drustvene mreze (OSN) koje trenutno koristimo koriste Web 2.0 tehnologiju, koja dozvoljava korisnicima da medjusobno interaguju jedni sa drugima. Veoma se brzo size I menjaju nacin na koji ljudi ostaju u kontaktu jedni sa drugima.

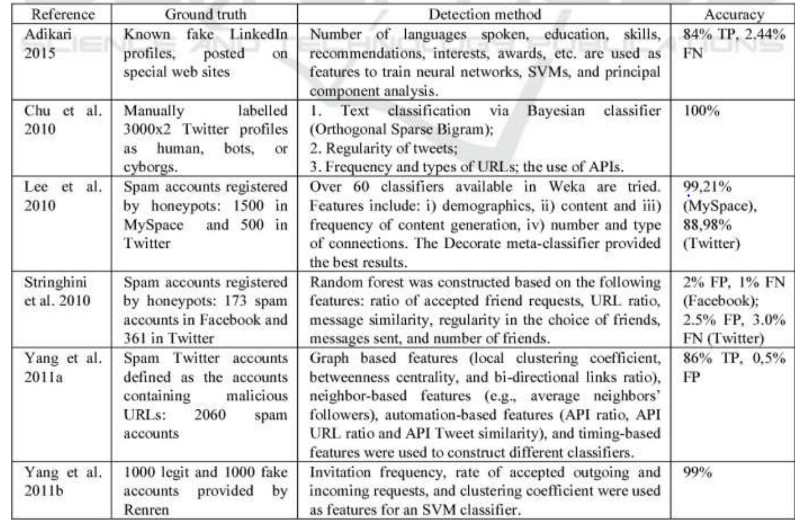
One su nastale jos davne 1997. Godine ([www.sixdegress.com](http://www.sixdegress.com)), i tokom svoje burne istorije dozivele su veliki broj znacajnih promena. Naucnici I dalje izucavaju uticaj koje ove drustvene mreze imaju na njihove korisnike. U ovom radu cemo se fokusirati na one negativne. Preveliki problem sa drustvenim mrezama jesu fake profili, kojih je svakog dana sve vise I vise… Do danasnjeg dana niko nije dosao na neku sveobuhvatnu ideju koja bi mogla resiti ovaj problem. Zbog svega toga sam se odlucio da fokus mog rada iz ovog predmeta bude na pronalazenju resenja za automatsku detekciju fake profila na drustvenim mrezama, kako bi sajtovima bilo lakse da se stvore jedan bezbedniji ekosistem za prave korisnike I time poboljsaju korisnicko iskustvo.

**Istrazivanje problema**

Brojne metodologije za prepoznavanje fake profila zavise od analize pojedinacnih profila na drustvenim mrezama, sa ciljem identifikovanja karakteristika ili njihovih kombinacija koje bi pomogle u razlikovanju legitimnih I laznih profila. Konkretno, razlicite karakteristike se izvlace iz profila I objava, a zatim se koriste algoritmi masinskog ucenja kako bi se izgradio klasifikator sposoban za prepoznavanje laznih profila.  
  
 Na primer, Nazir (2010) [1] opisuje fantomske profile u aplikacijama za drustvene igre. Clanak analizira poznatu online igru “Fighters club”, poznatu po tome sto pruza podsticaje I prednosti u igri korisnicima koji pozivaju svoje prijatelje da im se pridruze. On tvrdi, da dajuci takve podsticaje, igra motivise igrace da sa namerom kreiraju lazne profile. Unosenjem laznih profila, korisnik bi povecao motivacioni faktor za sebe.

Adikari I Dutta (2014) [2] opisuju prepoznavanje laznih profila na Linkedin-u. Rad pokazuje da se lazni profili mogu prepoznati sa 84% tacnosti. Tehnike koje su koriscene, su neuronske mreze, SVM-ovi, I analizu glavnih komponenti. Podaci koji su bili korisceni jesu broj jezika koje korisnik govori, obrazovanje, vestine, preporuke, interesi I nagrade. Informacije sa profila za koje se zna da su lazni su koriscene kao baza.

Chu I saradnici (2010) [3] se bave razlikovanjem Twitter naloga koje vode ljudi I botovi. Kao deo formulacije problema detekcije, identifikacija spam naloga je ostvarena uz pomoc OSB (Orthogonal Sparse Bigram) tekst klasifikatora koji koristi parove reci kao karakteristike.



Slika 1. Razlicita istrazivanja I radovi na temu detekcija laznih profila na drustvenim mrezama

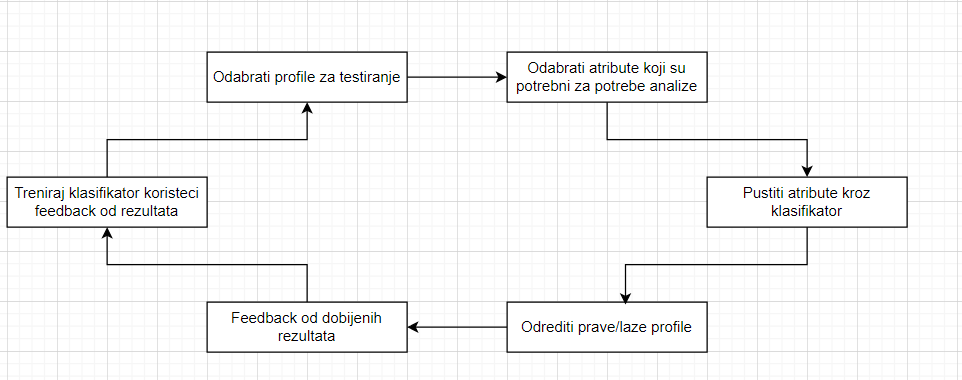
**Predlozeno resenje**

Svaki profil na drustvenoj mrezi sadrzi mnogo informacija kao sto su pol, broj prijatelja, broj komentara, obrazovanje, posao itd… Neke od tih informacija su privatne, a neke su javne. Posto privatne infromacije nisu dosupne, za ovo resenje su koriscene samo javne informacije kako bismo pomocu njih utvrdili lazne profile. Medjutim, ako bi ovakvo resenje koristile drustvene mreze, tada bi one mogle koristiti privatne informacije profila za detekciju, bez krsenja privatnosti korisnika. Ove informacije smo uzeli kao karakteristike profila za klasifikaciju laznih I stvarnih profila. Koraci su sledeci:

1. Prvo se biraju sve karakteristike nad kojima ce se upotrebiti algoritam klasifikacije. Treba voditi racuna o pravilnom izboru karakteristika, kao sto su karakterisike koje ne treba da zavise jedna od druge I treba odabrati one karakteristike koje mogu povecati efikasnost klasifikacije.
2. Nakon pravilnog izbora atribura, potreban je skup podataka prethodno identifikovanih laznih I stvarnih profila za obuku algoritma klasifikacije.
3. Atributi izabrani u prvom koraku moraju biti izdvojeni iz profila (laznih I stvarnih). Iz skupa, 80% oba profila, koristi se za pripremu skupa podataka za obuku, dok se 20% oba profila koristi za pripremu skupa podataka za testiranje. Efikasnost algoritma za klasifikaciju odredjujemo koristeci skup podataka za obuku koji sadrzi 922 profila I skup podataka za testiranje koji sadrzi 240 profila.
4. Nakon pripreme skupa podataka za obuku I testiranje, skup podataka za obuku se unosi u algoritam klasifikacije. Algoritam uci iz podataka za obuku I ocekuje se da pruzi tacne klase za skup podataka za testiranje.
5. Nivoi iz skupa podataka za testiranje se uklanjaju I prepustaju odredjivanju testiranom klasifikatoru. Efikasnost klasifikatora se izracunava deljenjem broja tacnih predikacija sa ukupnim brojem predikcija. Iskorisceni su tri algoritma za klasifikaciju I uporedili smo efikasnost klasifikacije ovih algoritama.

Predlozeni okvir na slici 2. prikazuje niz procesa koje je potrebno pratiti za kontinuirano otkrivanje laznih profila uz aktivno ucenje iz povratnih informacija koje daje algoritam za klasifikaciju. Ovako nesto se moze implementirati od strane kompanija koje su vlasnici tih mreza.

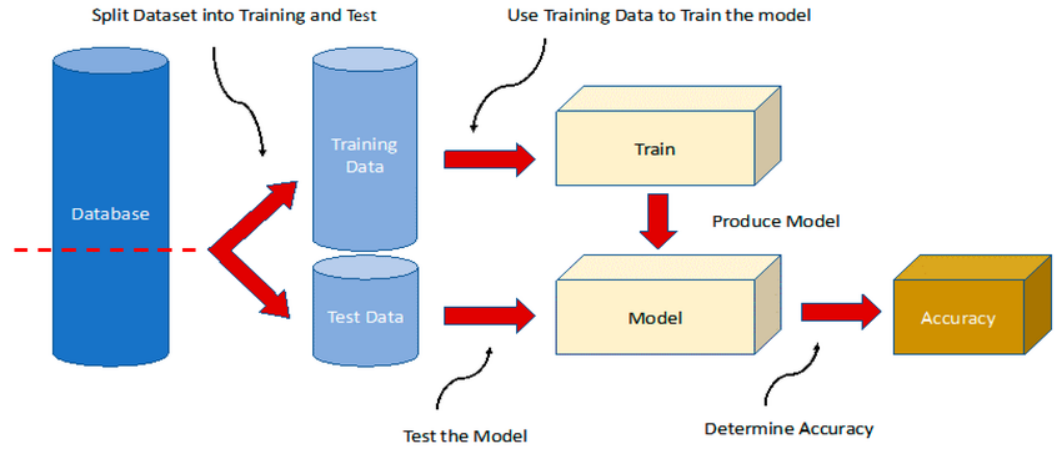
1. Proces detekcije pocinje sa selekcijom profila koji bi se testirali.
2. Nakon izbora profila, biraju se odgovarajuci atributi, tj. karakteristike na koje se primenjuju algoritam za klasifikaciju.
3. Izdvojeni atributi se prosledjuju treniranom klasifikatoru. Klasifikator se redovno trenira kako se novi trening podaci ubacuju u klasifikator.
4. Klasifikator odredjuje da li je profil lazan ili pravi.
5. Klasifikator mozda na pocetku nije 100% tacan u klasifikaciji profila.
6. Proces se ponavlja I vremenom, trenirani podaci se povecavaju I klasifikator postaje sve precizniji u predvidjanju laznih profila.



Slika 2. Niz procesa tokom detekcije laznih profila

**Klasifikacija**

Klasifikacija u mašinskom učenju je proces razvrstavanja podataka u unapred definisane kategorije ili klase. Ova tehnika spada u nadgledano učenje (supervised learning), gde algoritam uči iz skupa podataka koji već ima poznate oznake (labels) kako bi mogao pravilno da klasifikuje nove, nepoznate podatke.



Slika 3. Pristup kod izgradnje modela klasifikacije

Ova tehnika koristi algoritam ucenja da identifikuje model koji najvise odgovara vezi izmedju skupa atributa I oznake klase u skupu podataka za trening.

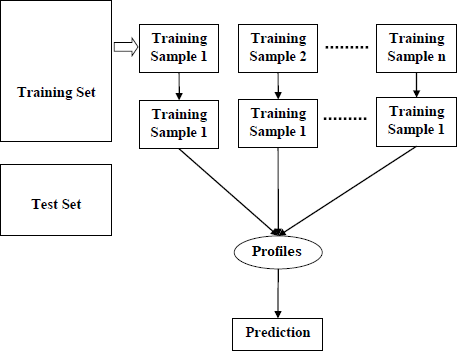
Slika 3. Prikazuje opsti pristup za izgradnju modela klasifikacije. Model koji generise algoritam ucenja treba da odgovara ulaznim podacima I da ispravno predvidja oznake klasa u test skupu sa velikom preciznoscu. Kljucni cilj algoritma za ucenje je izgradnja modela sa dobrom sposobnoscu generalizacije. Klasifikator koji je implementiran za klasifikaciju u ovom radu jeste Random Forest.

**Random Forest**

Random forest je jedan od najkoriscenijih algoritama masinskog ucenja. Moze se koristiti I za regresiju I za klasifikaciju. Ali se najcesce koristi za problem klasifikacije. Kao sto znamo u realnom svetu suma se sastoji od velikog broja drveta, I sto vise drveta u sumi, to je suma gusca. Slicno tome, ovaj algoritam kreira stabla odluke na uzrocima podataka, a zatim dobija predikciju za svakog od njih I na kraju bira najbolje resenja putem glasanja. To je ensemble metoda koja dalje bolje rezultate od pojedinacnog stable odluka jer smanjuje prekomerno prilagodjavanje (over-fitting) preko nalazenja srednje vrednosti rezultata (averaging the result).

Mozemo razumeti kako radi Random Forest algoritam pomocu sledecih koraka:

1. Na pocetku, zapocinjemo sa selekcijom, odnosno odabirom slucajnih (random) uzoraka iz datog dataseta.
2. Dalje, ovaj algoritam ce kreirati stablo odluke za svaki uzorak. Nakon toga ce uzeti rezultate predikcije iz svakog stable odluke.
3. U ovom koraku, glasanje ce biti obavljati za svaki predvidjen rezultat.
4. Na kraju se odabere onaj rezultat koji ima najvise glasova, kao konacni.

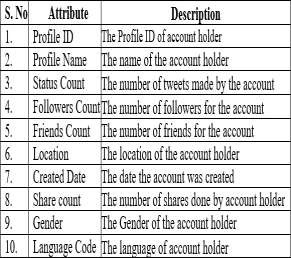


Slika 4. Koraci kod algoritma

**Implementacija**

Za potrebe naseg rada, bio nam je potreban dataset sa laznim I stvarnim profilima korisnika. Razliciti atributi ukljuceni u ovom datasetu su broj prijatelja, pratioci, pol, broj osoba koje korisnici prate, itd… Dataset je podeljen u trening I test skup podataka. Algoritmi klasifikacije su trenirani koriscenjem trening dataseta I test dataseta kako bi se nasla efikasnost algoritma. Iz datih dataset-ova, 80% korisnika (I iz laznih I pravih profila), bice korisceni kako bi se pripremili trening podaci, I 20% za test podatke.

Tabela ispod prikazuje attribute koji ce biti iskoriscenji kako bi se pronasli I preoznali lazni profile.



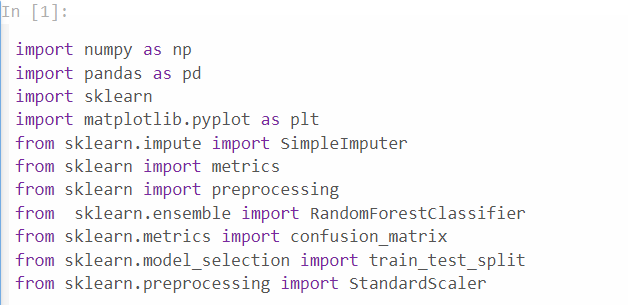
Slika 5. Atributi koji ce biti korisceni

Kao sto je receno, izvrsicemo ekstrakciju, sledecih atributa iz dataset-a:

* id – Jedinstveni identifikator korisnika drustvene mreze
* listed\_count - Atribut koji predstavlja koliko ljudi su dodali korisnika
* favourites\_count : Predstavlja koliko je puta odredjeni post oznacen kao omiljeni
* lang\_code : Jezik preko koga korisnik interaguje sa GUI drustvene mreze
* friends\_count – Broj prijatelja
* followers\_count – Broj ljudi koje korisnik prati

Interaktivno okruzenje koje je iskorisceno za potrebe rada jeste Jupyter Notebook, a prrogramski jezik koji je iskoriscen jeste Python.

Na samom pocetku uzovimo odredjene biblioteke koje cemo koristiti:



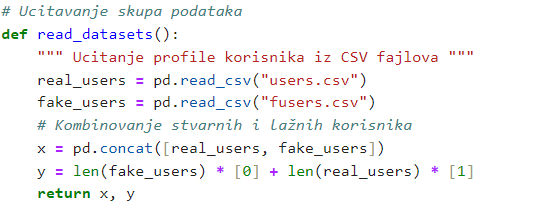
Slika 6. Importovanje biblioteka

Numpy predstavlja biblioteku u Pythonu koja pruza podrsku za efikasno maniupulisanje viseimenzionalnim nizovima I matricama.

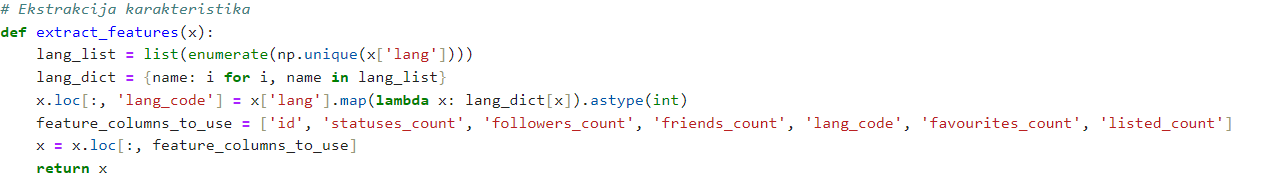
Pandas predstavlja biblioteku u Pythonu koja se koristi za manipulaciju I analizu podataka, kao sto su citanje I pisanje podataka, ciscenje podataka, agregacija I grupisanje…

Matplotlib je popularna Python biblioteka koja skuzi za vizuelizaciju podataka.

Sklearn biblioteka se koristi kod masinskog ucenja u Pythonu. Pruza jednostavne I efikasne alate za analizu podataka.

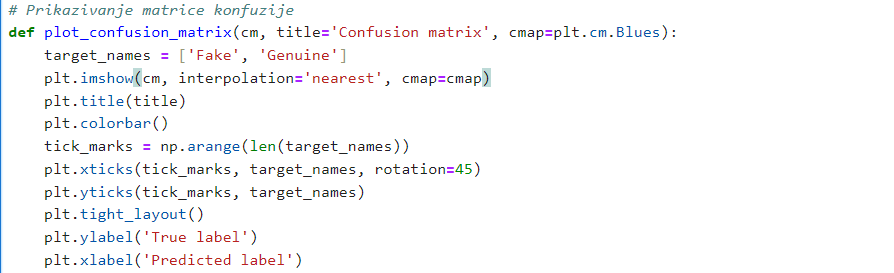
  
Slika 7. Metoda za ucitavanje podataka

Dalje, imamo metodu koja sluzi za ucitavanja skupa podataka, iz uploadovanih dataset-ova na Jupyter Notebook-u. Oba dataset-a su u .csv formatu. Preko biblioteke pandas ucitavamo te podatke I nakon toga vrsimo njihovu kombinaciju.

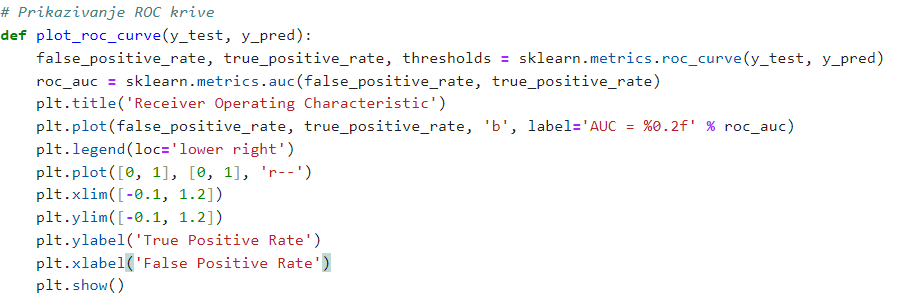


Slika 8. Metoda za ekstrakciju feature-a

Sledece metoda se koristi za ekstrakciju karakteristika, ranije navedenih. Svi feature-I su konvertovani u integer tipove.

 Slika 8. Metoda za kreiranje i prikazivanje matrice konfuzije

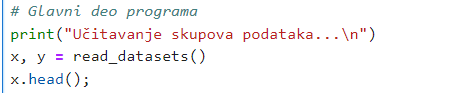
Metoda plot\_confusion\_matrix vrsi vizuelizaciju rezultata predikcije koriscenjem matrice konfuzije, koja nam dalje pomaze da otkrijemo tacnost (accuracy) klasifikatora.



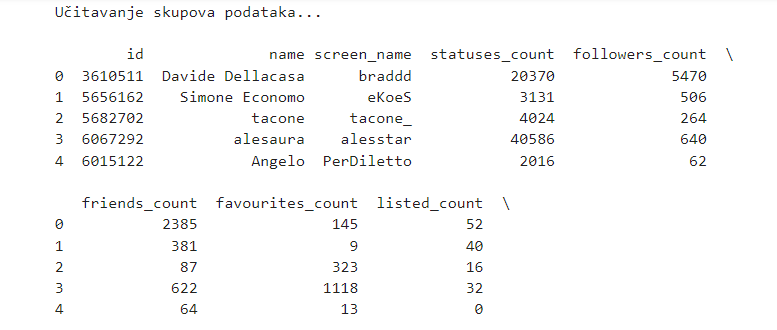
Slika 9. Metoda za kreiranje I prikazivanje ROC krive

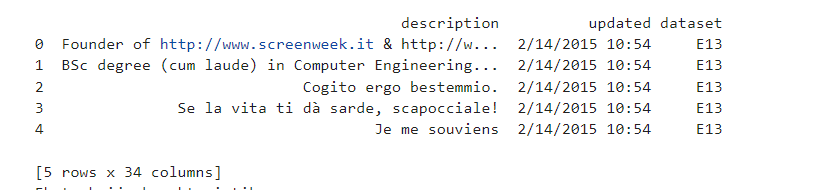
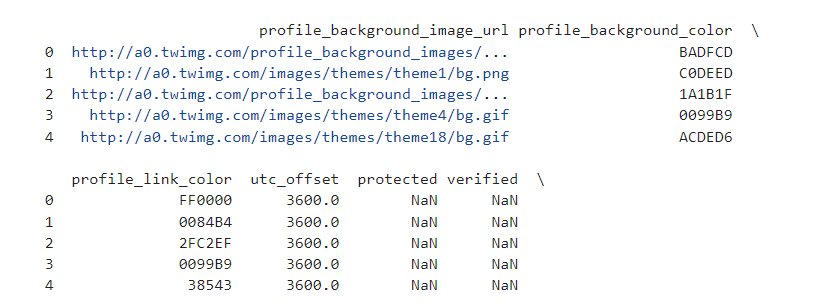
Metoda plot\_roc\_curve, u stvari prikazuje performance klasifikacionog modela pomocu grafa na svim pragovima klasifikacije. Ova kriva prikazuje dva parametra: True Positive Rate I False Positive Rate.

U glavnom delu programa, zovemo metodu read\_datasets(), I iz rezultata te metode pozivamo metodu head() koja nam vraca prvih 5 redova podataka.



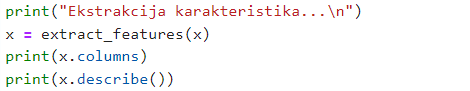
Slika 10. Ucitavanje podataka



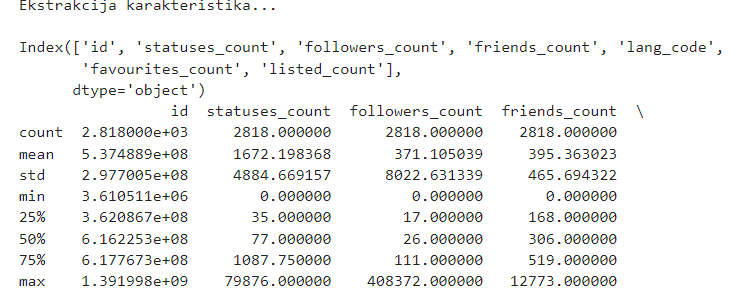
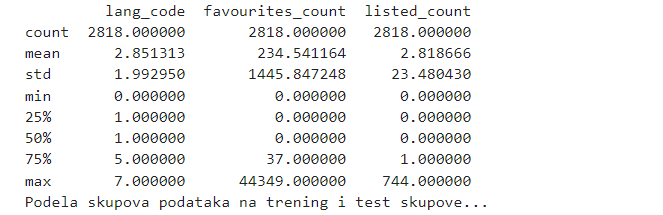
  


Slika 11. Ucitani podaci

U sledecem koraku imamo output nakon pozivanja metode extract\_features(x).



Slika 12. Ekstrakcija karakteristika

Slika 13. Ekstraktovani podaci

Dalje cemo implementirati algoritam Random Forest ili algoritam slučajnih šuma iz Scikit Learn biblioteke u naš algoritam. Random Forests su ensemble metoda učenja koja se koristi za klasifikaciju, regresiju i druge zadatke. To je ensemble metoda koja kombinuje više pojedinačnih stabala odlučivanja kako bi poboljšala performanse modela i smanjila problem prekomernog prilagođavanja (overfitting).

Random Forests ispravljaju problem prenaučenosti koji je karakterističan za stabla odlučivanja.

Inicijalizovaćemo objekat 'clf' koji će sadržati model Random Forest-a i potom ćemo podesiti parametre kao što su broj stabala ili (out of bag score) za taj objekat. Nakon toga, objekat će biti uparen sa našim skupom za obuku kako bismo implementirali Random Forest na našem skupu podataka.

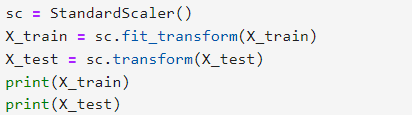
Podela skupa podataka na test I train, gde ce 20% podataka biti kod test skupa, dok ce se ostalih 80% podataka koristiti za treniranje:



Slika 14. Podela podataka na test I trening

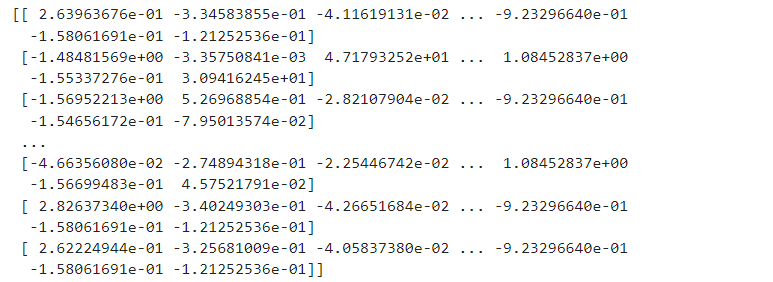
Skaliranje osobina (Feature scaling) podataka je tehnika koja standardizuje nezavisne osobine prisutne u podacima u fiksnom opsegu. Ako se skaliranje osobina ne obavi, algoritam mašinskog učenja ima tendenciju da većim vrednostima pridaje veću važnost, smatrajući ih većim, dok manje vrednosti smatra manjim, bez obzira na jedinicu vrednosti. Skaliranje osobina se vrši na nezavisnim osobinama u skupu podataka, tj. X\_train i X\_test.

Skaliranje osobina uvek treba da se obavi posle podele skupa za obuku i skupa za testiranje. Razlog za to je što, ako se skaliranje osobina obavi pre podele, informacije mogu procuriti unutar modela i algoritam može da predvidi vrednosti u testnom setu unapred, što može dovesti do netačnih rezultata tačnosti.



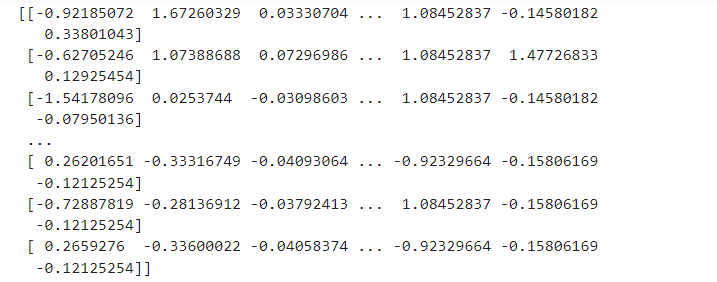
Slika 15. Standardizacija podataka

Output za X\_train:



Slika 16. Standardizovani treining podaci

Output za X\_test:

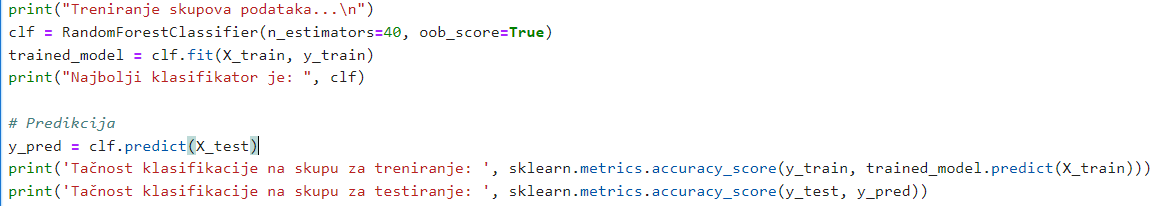


Slika 17. Standardizovani test podaci

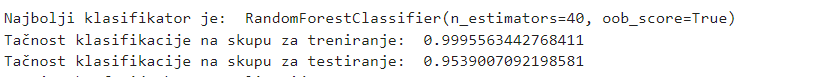
Tačnosti klasifikacije pomažu nam da odredimo da li naš model ima prenaučenost.

Prenaučenost se javlja kada naš model mašinskog učenja previše tačno odgovara određenom skupu podataka (u ovom slučaju skupu za obuku) i može zbog toga dopustiti da se prilagodi dodatnim podacima ili predvidi buduće opservacije (skup za testiranje).

Možemo proveriti prenaučenost poređenjem tačnosti klasifikacije na skupu za obuku sa tačnošću na skupu za testiranje. Ako model pokazuje izuzetno dobre rezultate na skupu za obuku, ali loše na skupu za testiranje, to je siguran znak prenaučenosti, ali to ovde nije slučaj.



Slika 18. Metode za poziv Random Forest klasifikatora I predikcije za test I trening podatke

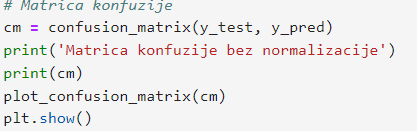


Slika 19. Rezultati izracunavanja

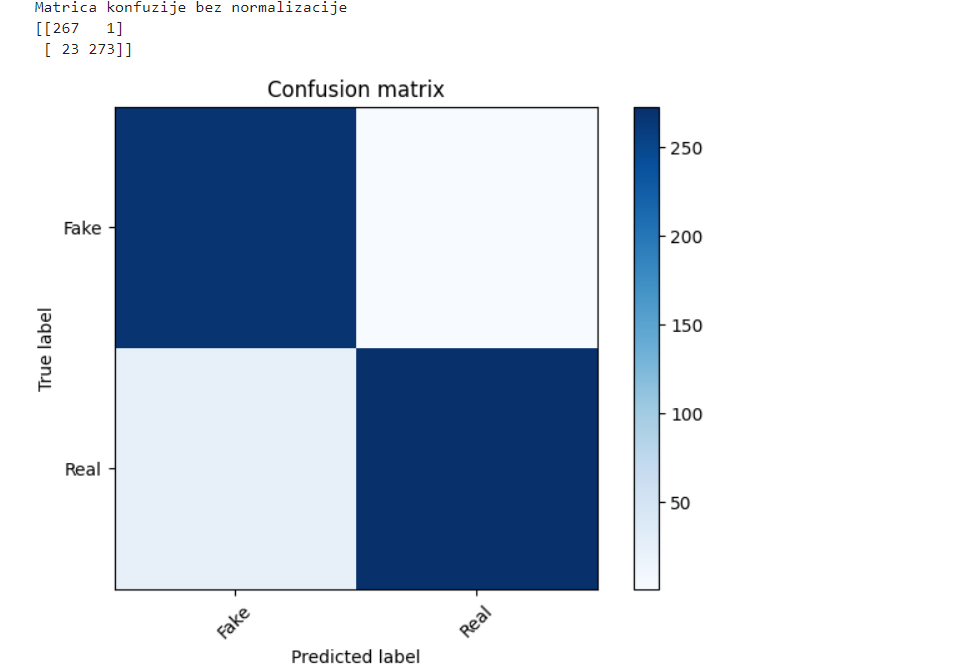
**Evaluacija**

Kao sto je receno u prethnodnim recenicama, matrica konfuzije je tabela koja se često koristi za opisivanje performansi klasifikacionog modela (ili "klasifikatora") na skupu testnih podataka za koje su poznate stvarne vrednosti.

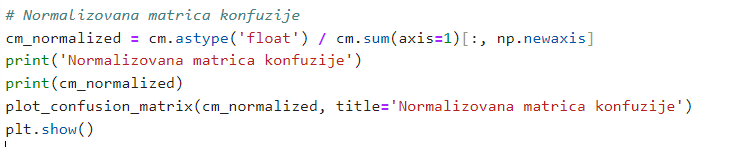
Prvo cemo prikazati matricu konfuzije bez normalizacije:



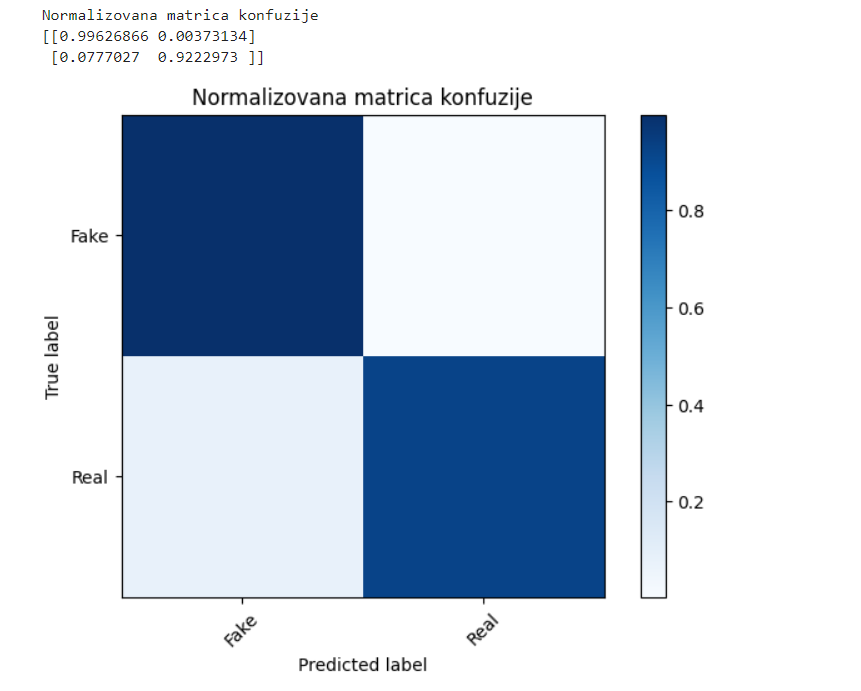
Slika 20. Metode za kreiranje matrice konfuzije bez normalizacije



Slika 21. Matrica konfuzije bez normalizacije

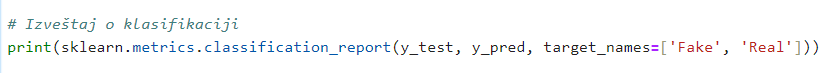
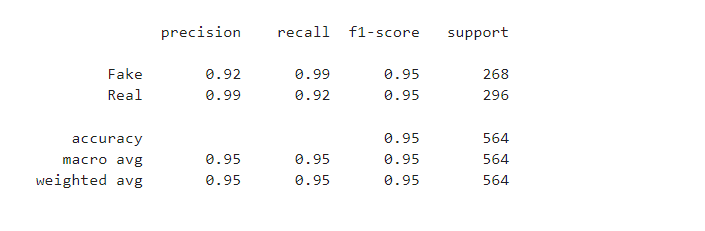


Slika 22. Metode za kreiranje normalizovane matrice konfuzije



Slika 23. Normalizovana matrica konfuzije

Nakon toga cemo kreirati izvestaj o klasifikaciji, pomocu metoda iz Sklearn biblioteke.

Slika 24. Izvestaj o klasifikaciji

Na osnovu brojača svake sekcije, možemo izračunati preciznost (precision) i odziv (recall) za svaku oznaku:

Preciznost je mera tačnosti koja se odnosi na tačnost predikcije za određenu klasnu oznaku. Definiše se kao preciznost = TP / (TP + FP), gde su TP (True Positive) tačno predviđeni pozitivni slučajevi, a FP (False Positive) su netačno predviđeni pozitivni slučajevi.

Odziv je stopa pravilno predviđenih pozitivnih slučajeva i definiše se kao odziv = TP / (TP + FN), gde su FN (False Negative) slučajevi koji su netačno klasifikovani kao negativni.

Na osnovu preciznosti i odziva za svaku klasu, možemo izračunati F1 skor.

F1 skor je harmonijska srednja vrednost preciznosti i odziva, gde F1 skor dostiže najbolju vrednost od 1 (savršena preciznost i odziv) do najgore vrednosti od 0. Ovo je dobar način da se pokaže da klasifikator ima dobru vrednost kako za odziv tako i za preciznost.

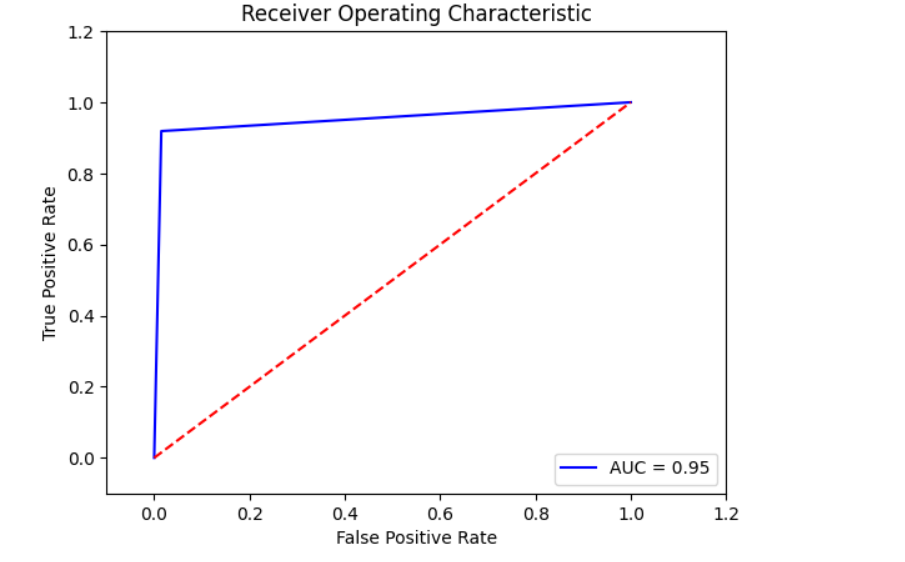
Na kraju, prosečna tačnost za ovaj klasifikator se računa kao prosečna vrednost F1 skora za obe oznake, što u našem slučaju iznosi 0.95.

Ukratko, ovi pojmovi i metrike pomažu nam da procenimo koliko dobro naš klasifikator radi na identifikaciji svake klase, kao i ukupnu tačnost modela.

Zatim cemo izracunati I prikazati ROC krivu. ROC kriva (kriva radne karakteristike prijema) je grafikon koji prikazuje performanse klasifikacionog modela pri svim pragovima klasifikacije. Ovaj grafikon prikazuje dva parametra: True Positive Rate (stopa pravilno predviđenih pozitivnih slučajeva) i False Positive Rate (stopa netačno predviđenih pozitivnih slučajeva).

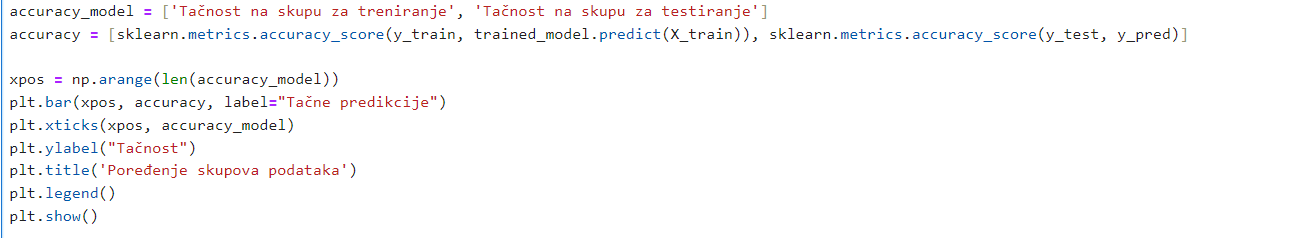
ROC krive obično imaju True Positive Rate (TPR) na Y osi i False Positive Rate (FPR) na X osi. To znači da je gornji levi ugao grafikona "idealna" tačka - FPR od nule i TPR od jedan. Ovo nije vrlo realistično, ali znači da je veća površina ispod krive (AUC - area under the curve) obično bolja.

ROC kriva je korisna za procenu performansi klasifikacionog modela jer omogućava vizualizaciju kako se model ponaša pri različitim pragovima za klasifikaciju pozitivnih i negativnih instanci.



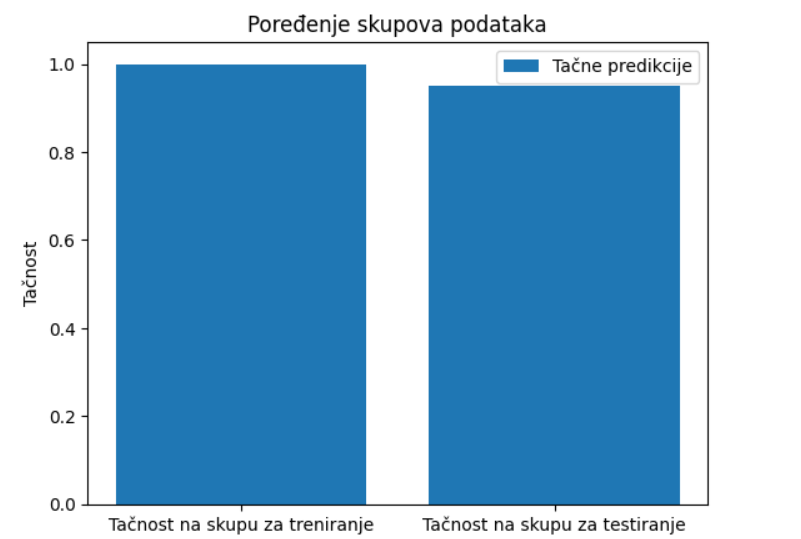
Slika 25. ROC kriva

Na kraju cemo uporediti preciznost (accuracy) kod skupa za treniranje I skupa za testiranje.



Slika 26. Uporedjivanje skupova

Kreirali smo stubičaste dijagrame da bismo napraviji vizueliaciju i uporedili tačnost klasifikacije na trening skupu sa tačnošću na test skupu. Stubičasti dijagrami su kreirani pomoću Matplotlib-a i možemo ponovo proveriti overfitting na ovaj način.



Slika 27. Poredjenje skupova podataka

**Zakljucak**

Na osnovu implementacije Random Forest klasifikatora za detekciju lažnih profila na osnovu podataka o korisnicima, možemo zaključiti da smo postigli visoku tačnost od oko 95%. Ovaj rezultat je pokazatelj da Random Forest klasifikator efikasno razlikuje između stvarnih i lažnih profila na osnovu karakteristika kao što su broj statusa, broj pratilaca, broj prijatelja i druge relevantne informacije.

Implementacija je pokazala da je model dobro generalizovan, jer je postigao visoku tačnost ne samo na skupu za obuku već i na nezavisnom skupu za testiranje. To znači da model nije samo zapamtio podatke iz skupa za obuku, već je uspeo da prepozna obrasce koji se mogu primeniti i na nove, neviđene podatke. Matrica konfuzije, koja prikazuje broj tačnih i netačnih predviđanja, kao i ROC kriva, koja ilustruje performanse modela na različitim pragovima klasifikacije, dodatno potvrđuju pouzdanost klasifikatora. Matrica konfuzije omogućava detaljan uvid u tačnost modela po klasama, dok ROC kriva pruža informacije o odnosu između tačnih pozitivnih i lažno pozitivnih stopa na različitim pragovima, čime se osigurava sveobuhvatno razumevanje performansi modela.

Ovaj pristup ima široku primenu u očuvanju sigurnosti i poverenja na online platformama, pružajući efikasan alat za detekciju potencijalno štetnih ili nepoželjnih korisničkih profila. S obzirom na sve veći broj lažnih i zlonamernih profila na društvenim mrežama i drugim online platformama, takvi algoritmi su od suštinskog značaja za održavanje integriteta i sigurnosti korisničkih baza podataka.Random Forest klasifikator nudi fleksibilnost i robustnost u radu sa različitim tipovima podataka, što ga čini idealnim za složene zadatke kao što je detekcija lažnih profila. Njegova sposobnost da se nosi sa velikim brojem karakteristika i da pruži informacije o važnosti svake karakteristike doprinosi njegovoj efikasnosti i interpretabilnosti.

**Literatura**

1. Nazir, Atif, Saqib Raza, Chen-Nee Chuah, Burkhard Schipper, and C. A. Davis. “Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications.” In *WOSN*. 2010.
2. Adikari, Shalinda, and Kaushik Dutta. “Identifying Fake Profiles in LinkedIn.” In *PACIS*, p. 278. 2014.
3. Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. “Who is tweeting on Twitter: human, bot, or cyborg?.” In Proceedings of the 26th annual computer security applications conference, pp. 21- 30. ACM, 2010.