Zaprimljen: 29.5.2025.
Odobren za objavu: 23.7.2025.
Izvorni znanstveni rad

EMPIRIJSKA STUDIJA POSTUPAKA STROJNOG UČENJA ZA PREPOZNAVANJE MALICIOZNIH NAPADA

Antonio Carević[1], Mario Dudjak[2]
[1] Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet Elektrotehnike, računarstva i informacijskih tehnologija Osijek, Kneza Trpimira 2B, 31000 Osijek, Hrvatska
ePošta: acarevic@etfos.hr
[1] Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet Elektrotehnike, računarstva i informacijskih tehnologija Osijek, Kneza Trpimira 2B, 31000 Osijek, Hrvatska
ePošta: mdudjak@etfos.hr

Sažetak: Cilj rada je definirati tok učenja algoritama klasifikacije iz skupova podataka koji opisuju različite vrste malicioznih napada i odrediti pojedinačne procedure unutar tog toka. Primjenom definiranog toka dobiveni su rezultati koji pokazuju visoku kvalitetu klasifikacijskih modela u prepoznavanju malicioznih napada, što potvrđuje njegovu primjenjivost u području kibernetičke sigurnosti, posebno u sustavima za detekciju upada. Korišteni algoritmi strojnog učenja su: naivni Bayesov algoritam, k-najbližih susjeda, stablo odluke, nasumična šuma i logistička regresija. Tijekom odabira značajki korišteni su filtri s Pearsonovim koeficijentom korelacije, zajedničkom informacijom i ANOVA F-vrijednosti te omotač slijedna pretraga unaprijed. Za obradu neuravnoteženih skupova podataka primijenjeni su postupci nasumičnog preuzorkovanja i poduzorkovanja. Najbolje rezultate postigao je algoritam stablo odluke s F1 mjerom od 1.0 na većini skupova podataka, dok je naivni Bayesov algoritam imao znatno slabije performanse, s F1 vrijednostima u rasponu od 0.12 do 0.98. Tehnike odabira značajki uglavnom su poboljšale performanse, pri čemu se posebno istaknuo omotač. Među postupcima za smanjenje neuravnoteženosti podataka, nasumično preuzorkovanje dosljedno je poboljšalo performanse svih algoritama, dok je poduzorkovanje dovelo do značajnog smanjenja performansi kod pojedinih algoritama, uz pad F1 mjere i do 0.22. Predloženi tok učenja omogućuje sustavno vrednovanje utjecaja različitih metoda predobrade podataka i algoritama klasifikacije, čime doprinosi boljem razumijevanju procesa detekcije malicioznih napada u neuravnoteženim i heterogenim podatkovnim skupovima te može poslužiti kao temelj za razvoj učinkovitijih sustava kibernetičke obrane u stvarnim okruženjima.

Ključne riječi: klasifikacija, maliciozni napadi, neuravnoteženi skup podataka, odabir značajki, strojno učenje

1. Uvod

Brz razvoj Interneta omogućio je jednostavan pristup velikoj količini informacija, što ih istovremeno čini izloženima brojnim sigurnosnim prijetnjama. Jedna od glavnih prijetnji su maliciozni napadi, koji se definiraju kao programi čija je svrha probijanje obrambenih sustava računala te ugrožavanje povjerljivosti, integriteta i dostupnosti podataka (Sharp, 2017). Postoji mnogo načina na koje se maliciozni napadi mogu izvesti, a neki od najčešćih medija uključuju zaražene elektroničke poruke te kompromitirane internetske stranice (Ahsan i sur., 2022). Iako su vrste malicioznih napada brojne, ovaj se rad fokusira na neke od najpoznatijih i najčešćih, uključujući: viruse, crve, reklamne programe, ucjenjivačke programe, neželjenu poštu (engl. *spam*), trojanske konje, špijunske programe, napade distribuiranim

uskraćivanjem usluga (engl. *Distributed Denial of Service*, DDOS) te ubrizgavanje zlonamjernih URL-ova. Tradicionalni sustavi obrane ne mogu pratiti sve brži razvoj i sve veću složenost ovih prijetnji. Nedostatak informacija o novim vrstama napada dodatno ograničava učinkovitost klasičnih pristupa. Kao moguće rješenje nameće se primjena strojnog učenja. Strojno učenje bavi se razvojem računalnih algoritama koji na temelju empirijskih podataka izgrađuju modele sposobne za prepoznavanje obrazaca i zaključivanje (Tsiakos i Chalkias, 2023). Zbog te sposobnosti, takvi su modeli osobito prikladni za prepoznavanje različitih vrsta malicioznih napada, uključujući i novonastale napade o kojima postoje ograničeni podaci. Na temelju znanja stečenog iz postojećih primjera napada, modeli mogu naučiti prepoznavati i nove prijetnje. Nadzirano strojno učenje predstavlja oblik strojnog učenja koji se pokazao najprikladnijim za detekciju malicioznih aktivnosti. Kod ovog pristupa algoritmi uče iz unaprijed označenih podataka, pri čemu je poznat ishod za svaki skup ulaznih podataka (Shaukat i sur., 2020). Poseban oblik nadziranog strojnog učenja koji se široko primjenjuje u detekciji napada je klasifikacija.

Većina dosadašnjih radova na ovu temu fokusira se na jedan specifičan problem i prikazuje ostvarene rezultate, no često nedostaje opsežna eksperimentalna analiza koja bi omogućila dublji uvid u prednosti i nedostatke pojedinih algoritama. Rijetko se pritom navodi koji je algoritam učinkovitiji za određene vrste napada, a tek neznatan broj radova obrađuje problem neuravnoteženosti skupova podataka. Taj je problem prisutan u gotovo svim dostupnim skupovima koji opisuju maliciozne aktivnosti, a njegovo rješavanje omogućuje preciznije rezultate i bolji uvid u stvarne performanse klasifikacijskih algoritama.

Cilj ovog rada je definirati prikladan tok učenja algoritama klasifikacije iz skupova podataka koji opisuju različite oblike malicioznih napada te odrediti odgovarajuće postupke unutar tog toka. Odabirom specifičnih algoritama strojnog učenja izgradit će se modeli za detekciju malicioznih aktivnosti. Nadalje, primjenom tehnika predobrade u svrhu odabira značajki pokušat će se poboljšati performanse modela te prikazati utjecaj odabira značajki, aspekt koji je u postojećim radovima često zanemaren. Također, primjenom metoda uzorkovanja nastojat će se ublažiti neželjeni učinak problema neuravnoteženosti klasa.

Struktura rada organizirana je na sljedeći način. U drugom poglavlju prikazan je pregled relevantne literature te su objašnjeni najznačajniji oblici malicioznih napada i postojeći pristupi njihovom prepoznavanju. Treće poglavlje donosi opis eksperimentalnih postavki i korištene metodologije. U četvrtom poglavlju izneseni su rezultati analize te njihova interpretacija. Konačno, peto poglavlje sadrži zaključke rada i prijedloge za buduća istraživanja.

2. Pregled literature

Zahvaljujući sposobnosti prepoznavanja i prilagodbe novim vrstama napada, strojno učenje sve češće se primjenjuje u obrambenim kibernetičkim sustavima. U radu (Martínez Torres i sur., 2019) analizirani su najčešći oblici malicioznih napada, uključujući spam, mrežnu krađu identiteta (engl. *phishing*) i zlonamjerne programe. Kao najučinkovitiji algoritmi istaknuti su naivni Bayesov algoritam, stroj potpornih vektora, stabla odluke, k-najbližih susjeda, neuronske mreže i nasumične šume. Međutim, navedeno istraživanje ne uzima u obzir problem neuravnoteženih podataka, što predstavlja značajno ograničenje. U radu (D'hooge i sur., 2019) fokus je stavljen na DDoS napade, pri čemu su algoritmi temeljeni na stablima odluke postigli najbolje rezultate, dok je algoritam k-najbližih susjeda identificiran kao alternativno rješenje. Za razliku od ovog rada, naše istraživanje obuhvaća širi spektar napada i omogućuje usporedbu učinkovitosti različitih algoritama.

Primjena neuronskih mreža za upravljanje sigurnosnim uređajima istražena je u radu (Fraley i Cannady, 2017), gdje su postignuti obećavajući rezultati. S druge strane, rad (Ahsan i sur., 2022) prikazuje pregled različitih tehnika u području kibernetičke sigurnosti, ističući kao učinkovite algoritme naivni Bayes, logističku regresiju i stabla odluke. Ipak, ni u jednom od ta dva rada nije provedena detaljna evaluacija performansi niti usporedba većeg broja algoritama. Rad (Shaukat i sur., 2020) opisuje napredne tehnike strojnog učenja za unaprjeđenje kibernetičke sigurnosti, pri čemu su istaknuti algoritmi stabla odluke, nasumične šume, k-najbližih susjeda, neuronske mreže i naivni Bayes. Međutim, nedostatak ovog rada ogleda se u nedovoljnoj analizi procesa učenja iz podataka te neobrazloženom odabiru algoritama i tehnika predobrade. Zajednički nedostatak svih navedenih istraživanja jest ograničeno razmatranje utjecaja tehnika odabira značajki. U radu (Walling i Lodh, 2024) autori analiziraju jednu tehniku odabira značajki i njen utjecaj na prepoznavanje malicioznih napada, pri čemu je korišten algoritam nasumične šume. Međutim, rad ne uključuje usporedbu s drugim tehnikama koje bi potencijalno mogle dati bolje rezultate. Slično tome, rad (Kocher i Kumar, 2021) prikazuje utjecaj odabira značajki na algoritme k-najbližih susjeda, nasumične šume, logističke regresije i naivnog Bayesa, ali ne provodi međusobnu usporedbu korištenih tehnika predobrade, čime izostaje jasna procjena njihove učinkovitosti. Za razliku od navedenih radova, istraživanje (Yin i sur., 2023) obuhvaća prepoznavanje šireg spektra malicioznih napada korištenjem algoritama logističke regresije, stroja potpornih vektora, stabla odluke i nasumične šume. Autori su fokus stavili na smanjenje lažno pozitivnih rezultata i poboljšanje ukupnih performansi, ali bez primjene tehnika odabira značajki, što ograničava optimizaciju modela. U radu (Yin i sur., 2023) problem prepoznavanja malicioznih napada adresiran je korištenjem dubokih neuronskih mreža i nasumičnih šuma. Nedostatak ovog istraživanja jest korištenje samo jednog skupa podataka, čime je ograničena generalizacija modela na različite vrste napada. Rad (Sarhan i sur., 2021) prikazuje utjecaj tehnika predobrade na algoritme nasumične šume i dubokih neuronskih mreža. Iako su postignuti relevantni rezultati, rad ne definira cjelovit proces učenja koji uključuje sve faze primjene algoritama na različitim vrstama malicioznih napada, što predstavlja ključnu razliku u odnosu na ovu studiju.

Tablica 1. Skupovi podataka korišteni za potrebe eksperimentalne analize

Naziv	**Vrsta napada**	**Broj instanci**	**Broj značajki**	**Broj klasa**	**Stupanj neuravnoteženosti**
DARPA	DoS	4 554 344	4	2	1.51
KDD99	DoS, ubrizgavanje URL-a, mrežna krađa identiteta i drugi	494 020	42	23	3530.99
NSL-KDD	DoS, ubrizgavanje URL-a, mrežna krađa identiteta i drugi	148 517	42	40	37.27
KYOTO	DDoS	303 849	24	3	50644.17
Malware	Virus, crv, trojanski konj	100 000	35	2	1
DREBIN	Različiti maliciozni napadi operacijskog sustava Android	15 036	215	2	1.70

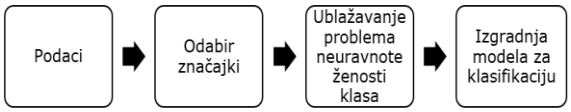
ISCXIDS2012	DoS, ubrizgavanje URL-a, mrežna krađa identiteta i drugi	171380	21	2	44.39
CICIDS2017	DDoS	225 745	79	2	1.31
DS2OS	Ucjenjivački, reklamni i špijunski programi	357 952	13	8	225.23
IMPACT	DoS, mrežna krađa identiteta i drugi	19 940	9	20	11.44
UNSW-NB15	DDoS, ubrizgavanje URL-a, mrežna krađa identiteta i drugi	257 673	45	2	1.77
CIC-DDOS2019	DDoS	300 000	88	19	787.62

3. Postavke i metodologija eksperimentalne analize

U radu je korišteno pet klasifikacijskih algoritama strojnog učenja. Odabrani algoritmi su: k-najbližih susjeda, stablo odluke, nasumična šuma, logistička regresija i naivni Bayesov algoritam (Kotsiantis i sur., 2007). Ovi algoritmi su odabrani kao prevladavajući u dostupnoj literaturi jer su često korišteni u svrhe treniranja modela strojnog učenja te sadrže razne vrste malicioznih napada. Dvanaest skupova podataka korišteno je za vrednovanje odabranih algoritama, a informacije o svakom skupu vidljive su u tablici 1. Skupovi podataka preuzeti su s javno dostupnih repozitorija UCI (Aha, 1987) i Kaggle (Goldbloom i Kaggle, 2010) te s internetskih stranica Instituta za kibernetičku sigurnost Sveučilišta u New Brunswicku (https://www.unb.ca/cic/datasets/ids.html) i grupe za istraživanje inteligentne sigurnosti Sveučilišta UNSW Sydney (https://research.unsw.edu.au/projects/unsw-nb15-dataset).

Svaki skup podataka podijeljen je u omjeru 80:20% na skup za treniranje i skup za testiranje. Za svaki hiperparametar klasifikatora definirane su različite vrijednosti čije su se kombinacije vrednovale pretraživanjem po mreži na podskupu za treniranje kako bi se pronašla optimalna. Nedostajuće vrijednosti su izbrisane, a kategoričke i ordinalne značajke kodirane su pomoću tehnike kodiranja oznaka te su normalizirane u raspon [0,1]. Kako bi se smanjio veliki broj redundantnih i nepotrebnih značajki, upotrijebljene su procedure odabira značajki. Metode koje su korištene u ovu svrhu su: filtri s Pearsonovim koeficijentom korelacije, zajedničkom informacijom i ANOVA F-vrijednosti te SFS omotač (Venkatesh i Anuradha, 2019). Prilikom upotrebe SFS omotača skup za treniranje dodatno je podijeljen na skup za treniranje i skup za validaciju u omjeru 65:35%. Nadalje, za ublažavanje problema neuravnoteženosti klasa primijenjene su metode nasumičnog preuzorkovanja i nasumičnog poduzorkovanja (Dudjak, 2022).

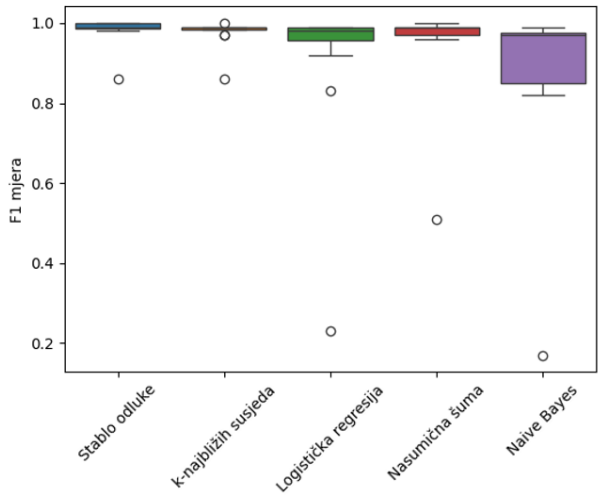
Računalo na kojem je proveden eksperiment opremljeno je sa Ryzen 5 5600 procesorom koji radi na 3.50 GHz, 16 GB RAM-a, AMD Radeon RX 6700 XT grafičku karticu sa 12 GB VRAM memorije te 1 TB SSD memorije. Cjelokupni eksperiment je ponovljen deset puta pri čemu su skupovi podataka različito podijeljeni s ciljem ublažavanja utjecaja stohastičnosti korištenih algoritama na dobivene rezultate. Veličine koje se koriste za evaluaciju dobivenih modela su: točnost, F1 mjera, stopa stvarno pozitivan rezultat (engl. *True Positive Rate*, TPR) i stopa stvarno negativan rezultat (engl. *True Negative Rate*, TNR). Konačne vrijednosti za svaku mjeru dobivene su kao prosječna vrijednost svih vrijednosti iz deset iteracija. Tok eksperimenta prikazan je na slici 1.



Slika 1. Koraci učenja iz skupova podataka koji opisuju maliciozne napade

4. Rezultati i rasprava

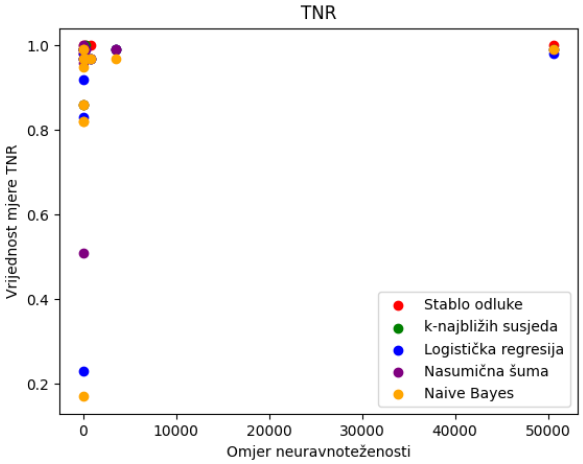
Dijagram pravokutnika (engl. *box plot*) na slici 2. sažeto prikazuje ostvarene vrijednosti F1 mjere na korištenim skupovima podataka. Vrijednosti F1 mjere algoritama stablo odluke, k-najbližih susjeda, nasumična šuma i logistička regresija usko su grupirane oko vrijednosti 1.0. To je pokazatelj kako svi ovi algoritmi ostvaruju visoke performanse za sve skupove podataka. Među njima se može izdvojiti algoritam stablo odluke zbog kontinuirano postignute vrijednosti 1.0 za F1 mjeru. Ostala tri algoritma povremeno ostvaruju lošije performanse na što ukazuje postojanje stršećih vrijednosti. Performanse naivnog Bayesovog algoritma primjetno su slabije kod gotovo svih skupova podataka. To je vidljivo na slici 2, gdje su vrijednosti F1 mjere naivnog Bayesovog algoritma u rasponu [0.80, 0.98] dok su rasponi ostalih algoritama znatno manji i bliže vrijednosti 1.0.



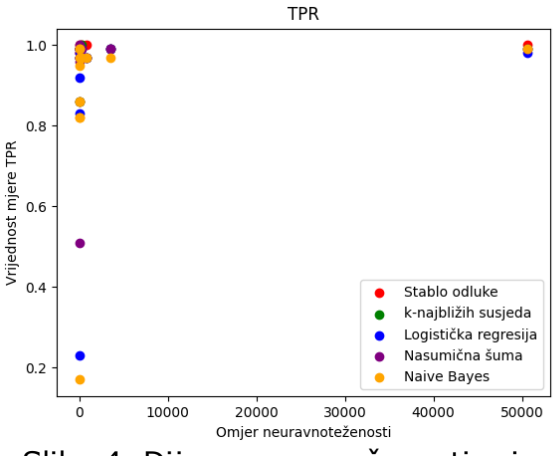
Slika 2. Dijagram pravokutnika vrijednosti F1 mjere

Manje vrijednosti točnosti na pojedinim skupovima podataka također upućuju na slabije performanse ovog algoritma. Na skupu NSL-KDD točnost naivnog Bayesovog algoritma iznosi 0.83, na skupu DREBIN 0.82, dok ostali algoritmi ostvaruju vrijednosti u rasponu [0.97, 0.99].

Zbog neuravnoteženosti podataka, model bolje klasificira većinsku klasu, postižući veće vrijednosti točnosti i TNR, dok se manjinska klasa slabije prepoznaje, što smanjuje F1 i TPR, što je prikazano na slikama 3 i 4. Ovaj trend je posebice vidljiv za algoritam naivnog Bayesa. Na NSL-KDD skupu podatak TNR za ovaj algoritam iznosi 0.93, dok mjere F1 i TPR imaju vrijednosti 0.82 i 0.83.



Slika 3. Dijagram raspršenosti mjere TNR i omjera neuravnoteženosti

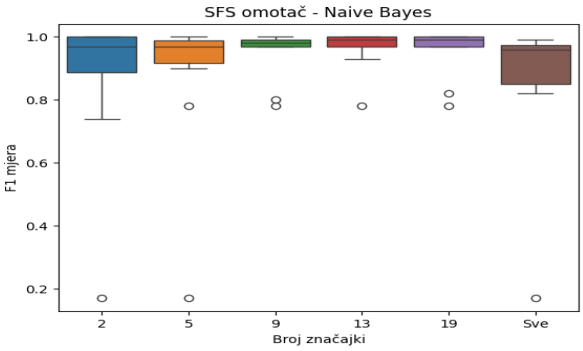


Slika 4. Dijagram raspršenosti mjere TPR i omjera neuravnoteženosti

4.1. Analiza učinka odabira značajki

Tehnike odabira značajki korištene su za poboljšanje performansi modela uklanjanjem irelevantnih značajki. Filtri

zasnovani na Pearsonovoj korelaciji, zajedničkoj informaciji i ANOVA F vrijednosti uglavnom su smanjili broj značajki, ali nisu znatno poboljšale performanse; posebice je Pearsonov filter loše radio na DREBIN skupu zbog velikog broja značajki. Vrijednosti F1 mjere svih algoritama su u padu, a najveći pad primjetan je kod logističke regresije. F1 mjera tog algoritma iznosi 0.90 nakon primjene filtera. Filter zasnovan na zajedničkoj informaciji poboljšao je performanse algoritma k-najbližih susjeda za 0.01, dok ANOVA filter nije imao značajan doprinos. S druge strane, omotač SFS značajno je poboljšao performanse svih algoritama, a posebice naivni Bayesov algoritam koji je postigao prihvatljive rezultate čak i na skupovima gdje je prije imao slabije rezultate. Detaljne performanse SFS metode za naivni Bayesov algoritam prikazane su na slici 5, dok su rezultati za sve algoritme i skupove podataka dostupni u tablici 2. Na dnu tablice izvedeni su rangovi Friedmanova testa za višestruku usporedbu (Derrac i sur., 2011), gdje manje vrijednosti sugeriraju bolju izvedbu uspoređenih algoritama. Dodatno, *post-hoc* statističke procedure ovog testa (poput primjerice, Nemenyove, Holmove, Shafferove te Bergmannove) ukazuju na to da dva najbolja algoritma (stablo odluke i algoritam k-najbližih susjeda) statistički značajno nadmašuju logističku regresiju i naivni Bayesov algoritam, uz razinu značajnosti od 0.05.



Slika 5. Dijagram pravokutnika F1 mjere naivnog Bayesovog algoritma za različit broj značajki odabranih SFS omotačem

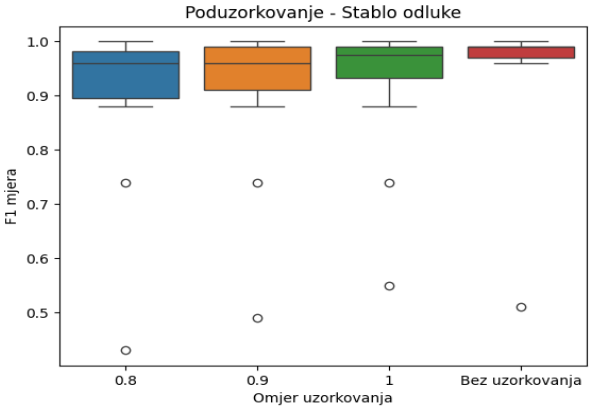
Tablica 2. Rezultati F1 mjere nakon primjene SFS omotača

Naziv	Stablo odluke	k-najbližih susjeda	Logistička regresija	Nasumična šuma	Naivni Bayesov
DARPA	0.99 ± 0.00 (-0.00)	0.99 ± 0.00 (-0.00)	0.83 ± 0.01 (-0.00)	0.99 ± 0.00 (-0.00)	0.91 ± 0.01 (-0.04)
KDD99	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (+0.01)	0.99 ± 0.01 (+0.01)	0.99 ± 0.01 (-0.00)	0.98 ± 0.01 (+0.01)
NSL-KDD	0.99 ± 0.01 (-0.00)	1.00 ± 0.00 (+0.01)	0.99 ± 0.01 (+0.02)	0.99 ± 0.01 (-0.00)	0.88 ± 0.01 (+0.06)
KYOTO	1.00 ± 0.00 (-0.00)	1.00 ± 0.00 (+0.01)	0.99 ± 0.01 (+0.01)	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (+0.01)
Malware	1.00 ± 0.00 (-0.00)	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (-0.00)	1.00 ± 0.00 (+0.03)
DREBIN	0.99 ± 0.01 (+0.01)	0.98 ± 0.01 (-0.01)	0.98 ± 0.01 (-0.00)	0.98 ± 0.01 (+0.01)	0.95 ± 0.01 (+0.13)
ISCXIDS2012	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (+0.01)
CICIDS2017	1.00 ± 0.00 (+0.01)	1.00 ± 0.00 (+0.01)	0.99 ± 0.01 (-0.00)	1.00 ± 0.00 (+0.01)	0.99 ± 0.01 (-0.00)
DS2OS	1.00 ± 0.00 (-0.00)	1.00 ± 0.00 (+0.00)	0.98 ± 0.01 (-0.02)	1.00 ± 0.00 (+0.01)	0.97 ± 0.01 (-0.00)
IMPACT	0.87 ± 0.01 (+0.01)	0.88 ± 0.01 (+0.01)	0.22 ± 0.01 (-0.01)	0.54 ± 0.01 (+0.03)	0.16 ± 0.01 (-0.01)
UNSW-NB15	1.00 ± 0.00 (+0.02)	1.00 ± 0.00 (+0.03)	0.95 ± 0.01 (+0.03)	1.00 ± 0.00 (+0.04)	1.00 ± 0.00 (+0.14)

CIC-DDOS2019	1.00 ± 0.00 (-0.00)	0.99 ± 0.01 (+0.02)	0.98 ± 0.01 (+0.01)	0.98 ± 0.01 (+0.01)	0.96 ± 0.01 (+0.09)
FR	2.125	2.125	3.875	2.75	4.125

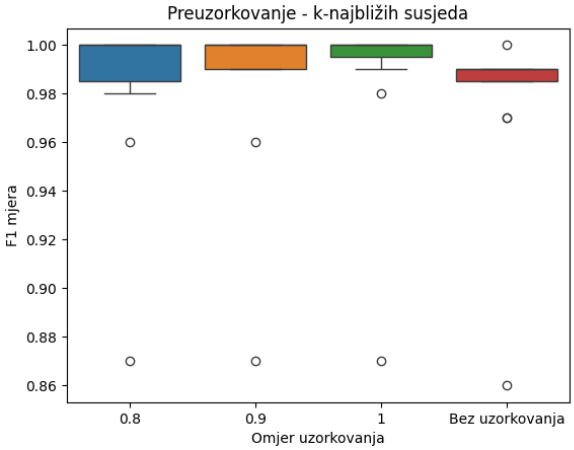
4.2. Analiza učinka uzorkovanja

Tehnike uzorkovanja imaju ključnu ulogu u prepoznavanju malicioznih napada zbog neuravnoteženosti većine skupova podataka gdje su primjeri manjinske klase često najvažniji za prepoznavanje (Krawczyk, 2016). Za rješavanje ovog problema koriste se nasumično preuzorkovanje i poduzorkovanje. Poduzorkovanje je uglavnom smanjilo performanse zbog dodatnog smanjenja već malog broja primjera. Isto se može primijetiti na KYOTO skupu podataka. Algoritmi k-najbližih susjeda, logistička regresija i nasumična šuma bilježe pad F1 vrijednosti za 0.75 te su njihove ostvarene vrijednosti 0.22, 0.25 i 0.24. Jedino kod skupova DARPA i CICIDS2017 se zamjećuje da je izjednačavanje kardinalnosti klasa poboljšalo rezultate, te algoritmi kod ovih skupova ostvaruju F1 vrijednosti oko 1.00. Suprotno tome, na skupu NSL-KDD primjetan je značajan pad performansi od oko 0.51, zbog velikog broja klasa s malim brojem primjera. Stablo odluke jedini je algoritam koji je zadržao vrijednosti F1 mjere veće od 0.9 kod većine skupova podataka, što je vidljivo na slici 6.



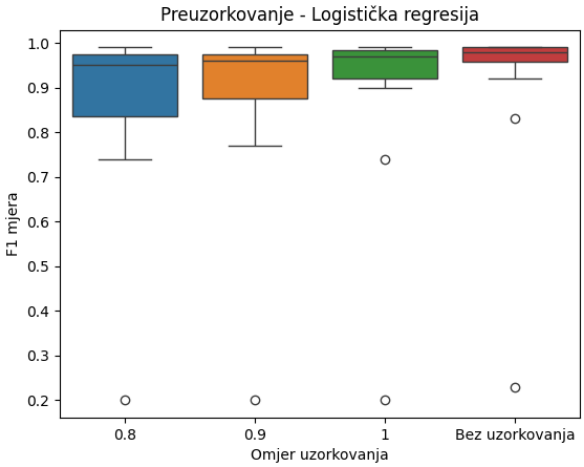
Slika 6. Dijagram pravokutnika F1 mjere algoritma stablo odluke za različite omjere uzorkovanja

Nasumično preuzorkovanje dosljedno je poboljšalo performanse svih algoritama na svim skupovima podataka, pri čemu stablo odluke na gotovo svim skupovima ostvaruje vrijednosti 1.0 za mjeru F1. Algoritam k-najbližih susjeda također bilježi značajna poboljšanja u odnosu na originalne skupove, kao što je prikazano na slici 7.



Slika 7. Dijagram pravokutnika F1 mjere algoritma k-najbližih susjeda za različite omjere uzorkovanja

S druge strane, kod logističke regresije i naivnog Bayes algoritma nasumično preuzorkovanje nije značajno doprinijelo performansama, pri čemu logistička regresija pokazuje varijabilne rezultate (slika 8).



Slika 8. Dijagram pravokutnika F1 mjere algoritma logistička regresija za različite omjere uzorkovanja

Algoritam naivni Bayes ima loše performansama na većini skupova podataka. Za Malware skup podataka vrijednost F1 mjere iznosi 1.0, na NSL-KDD skupu 0.51, a za IMPACT skupu podataka vrijednost je vrlo niskih 0.12. Iz navedenih brojeva primjećuje se nekonstantnost u prepoznavanju malicioznih napada, čak i nakon postignute ravnoteže u skupovima podataka.

5. Zaključak

U ovom radu prikazan je cjelovit pristup učenju algoritama klasifikacije na skupovima podataka s informacijama o malicioznim napadima, s ciljem ispitivanja mogućnosti primjene strojnog učenja u području kibernetičke sigurnosti. Znanstveni doprinos rada ogleda se u definiranju i evaluaciji toka učenja koji uključuje različite faze predobrade podataka, odabira značajki i uzorkovanja skupova podataka, čime se omogućuje sustavno vrednovanje utjecaja pojedinih metoda na performanse klasifikacijskih modela.
Rezultati eksperimentalne analize pokazuju da svi korišteni algoritmi ostvaruju zadovoljavajuće rezultate na većini skupova podataka, pri čemu se stablo odluke istaknulo kao najuspješniji algoritam. Uz stablo odluke, dobre rezultate pokazali su i algoritmi k-najbližih susjeda te nasumična šuma. Nasuprot tome, naivni Bayesov algoritam imao je najslabije performanse, što je posljedica njegove pretpostavke o međusobnoj neovisnosti značajki, uvjeta koji rijetko vrijedi za podatke o malicioznim napadima. Unatoč tome, upravo je kod ovog algoritma zabilježeno najveće poboljšanje u točnosti nakon primjene tehnika odabira značajki i uzorkovanja podataka. Od korištenih metoda za odabir značajki, SFS omotač pokazao se najuspješnijim, dok su i filter metode dale usporedive, ali nešto slabije rezultate. Kod rješavanja problema neuravnoteženosti skupova, učinkovitijom se pokazala tehnika nasumičnog preuzorkovanja.

Na temelju dobivenih rezultata može se zaključiti da strojno učenje ima značajan potencijal za unaprjeđenje kibernetičke sigurnosti. Predloženi tok učenja omogućuje primjenu klasifikacijskih modela visoke točnosti za prepoznavanje malicioznih aktivnosti u različitim podatkovnim okruženjima. Budući rad trebao bi se usmjeriti na razvoj kvalitetnijih i reprezentativnijih skupova podataka, uključivanje dodatnih izvora informacija kao što su mrežni dnevnici, vremenski obrasci i kontekstualni podaci, te primjenu i evaluaciju naprednijih modela dubokog učenja. Nadalje, potrebno je ispitati učinkovitost predloženog toka učenja u stvarnim, dinamičnim i distribuiranim okruženjima kibernetičke obrane, čime bi se dodatno potvrdila njegova praktična primjenjivost i robusnost.

6. Literatura

A., Goldbloom, Kaggle [online], San Francisco, 2010., dostupno na: https://www.kaggle.com/

Ahsan, M., Nygard, K. E., Gomes, R., Chowdhury, M. M., Rifat, N., & Connolly, J. F. (2022). Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning—A Review. Journal of Cybersecurity and Privacy, *2*(3), 527-555.

Ahsan, M., Nygard, K. E., Gomes, R., Chowdhury, M. M., Rifat, N., & Connolly, J. F. (2022). Cybersecurity threats and their mitigation approaches using Machine Learning—A Review. Journal of Cybersecurity and Privacy, *2*(3), 527-555.

D., Aha, UCI Machine Learning Repository, California, 1987., dostupno na: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

Derrac, J., Garcia, S., Molina, D., Herrera, F. (2011). A practical tutorial on tthe use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol. Comput., 1, 3-18.

D'hooge, L., Wauters, T., Volckaert, B., & De Turck, F. (2019). In-depth comparative evaluation of supervised machine learning approaches for detection of cybersecurity threats. In 4th International Conference on Internet of Things, Big Data and Security (IoTBDS*)* (pp. 125-136).

Dudjak, M. (2022). Učenje iz neuravnoteženih podataka unaprijeđenim postupcima za odabir značajki, preuzorkovanje i izgradnju radijalnih neuronskih mreža [Disertacija, Sveučilište Josipa Jurja Strossmayera u Osijeku]

Fraley, J. B., & Cannady, J. (2017, March). The promise of machine learning in cybersecurity. In SoutheastCon 2017 (pp. 1-6). IEEE.

Kocher, G., & Kumar, G. (2021). Analysis of machine learning algorithms with feature selection for intrusion detection using UNSW-NB15 dataset. Available at SSRN 3784406.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, *160*(1), 3-24

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. Progress in artificial intelligence, *5*(4), 221-232

Martínez Torres, J., Iglesias Comesaña, C., & García-Nieto, P. J. (2019). Machine learning techniques applied to cybersecurity. International Journal of Machine Learning and Cybernetics, *10*(10), 2823-2836.

More, S., Idrissi, M., Mahmoud, H., & Asyhari, A. T. (2024). Enhanced intrusion detection systems performance with UNSW-NB15 data analysis. *Algorithms*, *17*(2), 64.

Sarhan, M., Layeghy, S., & Portmann, M. (2021). Feature analysis for machine learning-based IoT intrusion detection. *arXiv preprint arXiv:2108.12732*.

Sharp, R. (2017). An Introduction to Malware.

Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). Performance Comparison and Current Challenges of Using Machine Learning Techniques in Cybersecurity. *Energies*, *13*(10), 2509.

Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). Performance comparison and current challenges of using machine learning techniques in cybersecurity. Energies, *13*(10), 2509.

Tsiakos, C.-A. D., & Chalkias, C. (2023). Use of Machine Learning and Remote Sensing Techniques for Shoreline Monitoring: A Review of Recent Literature. Applied Sciences, *13*(5), 3268.

University of New Brunswick, dostupno na: https://www.unb.ca/cic/datasets/ids.html

UNSW Sydney, dostupno na: https://research.unsw.edu.au/projects/unsw-nb15-dataset

Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. Cybern. Inf. Technol, *19*(1), 3-26.

Walling, S., & Lodh, S. (2024). Enhancing IoT intrusion detection through machine learning with AN-SFS: a novel approach to high performing adaptive feature selection. Discover Internet of Things, *4*(1), 16.

Yin, Y., Jang-Jaccard, J., Xu, W., Singh, A., Zhu, J., Sabrina, F., & Kwak, J. (2023). IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset. *Journal of Big data*, *10*(1), 15.

.

AN EMPIRICAL STUDY OF MACHINE LEARNING TECHNIQUES FOR MALICIOUS ATTACK DETECTION

Abstract: The aim of this paper is to define the learning flow of classification algorithms from datasets describing various types of malicious attacks and to determine individual procedures within that flow. By applying the defined flow, results were obtained that demonstrate the high quality of classification models in detecting malicious attacks, confirming its applicability in the field of cybersecurity, especially in intrusion detection systems. The machine learning algorithms used include: Naive Bayes, k-Nearest Neighbors, Decision Tree, Random Forest, and Logistic Regression. During feature selection, filters with Pearson correlation coefficient, mutual information, and ANOVA F-value were used, as well as the sequential forward selection (SFS) wrapper. For processing imbalanced datasets, random oversampling and undersampling procedures were applied. The Decision Tree algorithm achieved the best results with an F1 score of 1.0 on most datasets, while the Naive Bayes algorithm showed significantly weaker performance, with F1 values ranging from 0.12 to 0.98. Feature selection techniques generally improved performance, with the SFS wrapper being particularly prominent. Among the procedures for reducing data imbalance, random oversampling consistently improved the performance of all algorithms, whereas undersampling led to a significant decrease in performance for some algorithms, with F1 score drops of up to 0.22. The proposed learning flow enables the systematic evaluation of the impact of different data preprocessing methods and classification algorithms, thereby contributing to a better understanding of the process of malicious attack detection in imbalanced and heterogeneous datasets, and can serve as a basis for the development of more effective cybersecurity defense systems in real-world environments.

Keywords: classification, malicious attacks, imbalanced dataset, feature selection, machine learning