

Strojno učenje i daljinska istraživanja za mapiranje prometnica u Splitsko-dalmatinskoj županiji

Knežić Buhovac, Selena; Šerić, Ljiljana; Ivanda, Antonia

Source / Izvornik: **Zajednički temelji 2023. - uniSTem : deseti skup mladih istraživača iz područja građevinarstva i srodnih tehničkih znanosti, Split, 14.-17. rujna, 2023. : zbornik radova, 2023, 62 - 67**

Conference paper / Rad u zborniku

Publication status / Verzija rada: **Published version / Objavljena verzija rada (izdavačev PDF)**

<https://doi.org/10.31534/10.ZT.2023.22>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:123:498504>

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

Download date / Datum preuzimanja: **2024-07-10**



Repository / Repozitorij:

[FCEAG Repository - Repository of the Faculty of Civil Engineering, Architecture and Geodesy, University of Split](#)



UNIVERSITY OF SPLIT

DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI



UDRUGA
HRVATSKIH
GRAĐEVINSKIH
FAKULTETA



<https://doi.org/10.31534/10.ZT.2023.22>

STROJNO UČENJE I DALJINSKA ISTRAŽIVANJA ZA MAPIRANJE PROMETNICA U SPLITSKO-DALMATINSKOJ ŽUPANIJI

Selena Knežić Buhovac¹, Ljiljana Šerić², Antonia Ivanda²

(1) Sveučilište u Mostaru, Fakultet strojarstva, računarstva i elektrotehnike, Bosna i Hercegovina, selena.knezic@fsre.sum.ba

(2) Sveučilište u Splitu, Fakultet elektrotehnike, strojarstva i brodogradnje, R. Hrvatska, {ljiljana, asenta00}@fesb.hr

Sažetak

U ovom istraživačkom radu prikazana je metodologija za tematsko mapiranje prometnica u Splitsko-dalmatinskoj županiji. U svrhu ostvarivanja cilja korišten je satelitski snimak jedne scene Splitsko-dalmatinske županije Sentinel-2 satelita i podaci o položaju prometnica iz OpenStreetMap (OSM) projekta. Uzimajući u obzir činjenicu da su podaci iz OSM projekta neslužbeni i mogu biti nepouzdan izvor podataka, ovo istraživanje je usmjereno na modeliranje, mapiranje i evaluaciju uspješnosti predviđanja cesta primjenjujući algoritme strojnog učenja. Mjere evaluacije pokazale su da je prometnice moguće prepoznati iz satelitskih podataka korištenjem nelinearnih modela strojnog učenja. Budući da ovaj rad pruža osnovu koristeći značajke jednog piksela, u daljnjim istraživanjima može se istražiti primjena naprednih algoritama koji uzimaju u obzir i okolne piksele, i njihov kontekst kako bi se postigla još veća preciznost mapiranja.

Ključne riječi: daljinska istraživanja, strojno učenje, klasifikacija

MACHINE LEARNING AND REMOTE SENSING FOR ROAD MAPPING INSPLIT-DALMATIA COUNTY

Abstract

This research paper presents a methodology for thematic mapping of roads in Split-Dalmatia County. To achieve the goal, a satellite image of a scene in Split-Dalmatia County from the Sentinel-2 satellite and road position data from the OpenStreetMap (OSM) project were used. Considering that the data from the OSM project is unofficial and can be an unreliable data source, this research focuses on modeling, mapping, and evaluating the success of road predictions using machine learning algorithms. Evaluation measures have shown that roads can be identified from satellite data using non-linear machine learning models. Since this paper provides a foundation using single-pixel features, further research can explore the application of advanced algorithms that take into account neighboring pixels and their context to achieve even greater mapping accuracy.

Keywords: remote sensing, machine learning, classification

1. Uvod

Daljinska istraživanja (engl. *Remote Sensing*) opisana su mnogim definicijama u brojnim literaturi. Kada iščitamo i usporedimo neke od definicija, dolazimo do jednog zaključka, a to je da su daljinska istraživanja prikupljanja informacija na daljinu [1]. Službena definicija bi bila da su daljinska istraživanja proces dobivanja podataka o kopnenim i vodenim površinama Zemlje pomoću senzora smještenih na zrakoplovima ili satelitima. Ti senzori hvataju elektromagnetsku energiju koju emitira ili reflektira promatrana površina [2]. Istraživanje daljinskog očitavanja sve više stavlja naglasak na klasifikaciju pomoću tehnika strojnog učenja. Algoritmi strojnog učenja imaju sposobnost modeliranja složenih uzoraka klasa, mogu raditi s različitim ulaznim podacima i ne polaze od pretpostavki o distribuciji podataka (tj. neovisni su o parametrima). Općenito, brojna istraživanja su pokazala da ove metode često postižu veću točnost u usporedbi s tradicionalnim klasifikatorima koji se oslanjaju na parametre, posebno u slučajevima kada se radi s kompleksnim podacima koji imaju veliki broj značajki, odnosno prediktorskih varijabli [3]. Daljinska istraživanja pružaju mnoštvo podataka o Zemlji i njezinoj površini. S druge strane, strojno učenje predstavlja snažan alat za modeliranje i tematsko mapiranje tih podataka. Kombinacija ovih dvaju tehnologija omogućuje nam da iskoristimo bogatstvo podataka dobivenih daljinskim očitavanjem i primijenimo sofisticirane algoritme strojnog učenja kako bismo stvorili detaljne i precizne tematske karte. Strojno učenje omogućuje nam modeliranje složenih uzoraka klasa i rad s raznolikim ulaznim podacima, što doprinosi većoj točnosti u odnosu na tradicionalne metode. Ova kombinacija otvara vrata novim mogućnostima u istraživanju i razumijevanju Zemlje te pruža vrijedne informacije za odlučivanje u različitim područjima kao što su upravljanje okolišem ili npr. urbanizacija. U ovom radu bit će predstavljena metodologija za tematsko mapiranje prometnica Splitsko-dalmatinske županije. Podaci za strojno učenje dobiveni su iz izvora OpenStreetMap (OSM), koji nisu službeni i stoga se smatraju nepouzdanim. Bit će prikazana metodologija modeliranja, mapiranja evaluacije uspješnosti predviđanja cesta u Splitsko-dalmatinskoj županiji pomoću različitih algoritama strojnog učenja.

2. Metodologija

2.1. Područje istraživanja

Republika Hrvatska je podijeljena na 21 županiju. Površinom najveća je Splitsko-dalmatinska županija koja je i fokus ovog istraživanja. Smještena u središnjem dijelu Jadranske obale, prostire se na površini od 14.106,40 km² [4]. Kopneni dio ove županije obuhvaća površinu od 4.523,64 km² i vodeća je u Republici Hrvatskoj po dužini županijskih i lokalnih cesta. Splitsko-dalmatinska županija ima razgranatu cestovnu mrežu koja uključuje autoceste, državne ceste, županijske ceste i lokalne ceste. Udio županijskih cesta županije u Republici Hrvatskoj iznosi otprilike 8,6 % (814,6 km od ukupno 9.486,7 km), dok je udio lokalnih cesta oko 10,5 % (925,3 km od ukupno 8.787 km) [5].

2.2. Set podataka

Za potrebe ovoga rada korištena je snimka scene Splitsko-dalmatinske županije preuzete sa Sentinel hub-a, a snimljene pomoću satelita Sentinel-2. Snimka je preuzeta na datum kada je bilo manje oblaka kako bi bila što bolja vidljivost cesta. Podaci o prometnicama u Splitsko-dalmatinskoj županiji preuzete su s OpenStreetMaps (OSM). Dio gdje postoje ceste bit će označen brojem 1, a gdje se ne prepoznaje cesta brojem 0.

2.2.1. Sentinel-2

Sentinel-2 sateliti za promatranje Zemlje opremljeni su sustavima za multispektralno snimanje (MSI) i prikupljaju optičke slike. Snimke visoke prostorne rezolucije (do 10 metara) putem 12 spektralnih frekvencijskih kanala, uključujući vidljive, infracrvene i termalne raspone spektra, a ponavlja snimanje cijele Zemljine površine svakih pet dana. Sentinel-2A i Sentinel-2B su sateliti koji imaju isti instrument MSI te se nalaze u istoj sinkronoj orbiti Sunca, ali su postavljeni na suprotnim stranama [6]. Ovi sateliti su u vlasništvu i upravljaju Europske svemirske agencije (ESA), a proizvedeni su od strane konzorcija na čelu s Airbus Defence and Space (Airbus DS) [7]. Njihova misija podržava različite usluge i primjene, uključujući praćenje poljoprivrede, upravljanje katastrofama i klasifikaciju pokrova/zemljišta [8]. Kroz Copernicus program Europske unije podaci o zemljinoj površini prikupljeni ovim, ali i drugim satelitima, otvoreni su za korištenje cijeloj zajednici. Snimke s ovog instrumenta dostupne su kroz servis Sentinel Hub u geo-referenciranom formatu.

2.2.2. OpenStreetMap (OSM)

OpenStreetMap (OSM), osnovan 2004. godine, razvio se iz lokalnog projekta u globalnu kartu koja ima široku prmjenu i sadrži visokokvalitetne podatke. Trenutno je OSM snažan konkurent Google kartama. Oslanja se na online suradnju korisnika koji koriste ručne GPS uređaje, snimke visoke razlučivosti iz zraka i individualno prostorno kognitivno znanje kao temeljnu geografsku referencu za stvaranje, uređivanje, upravljanje i održavanje geografskih informacija. Dostupan je svim korisnicima putem interneta diljem svijeta [8].

Podaci iz oba korištena izvora svedeni su na istu rezoluciju. Iščitavanjem pomoću skripte napisane u Python programskom jeziku kreiran je set podataka u kojem je svaki piksel zemljine površine veličine 11 m x 15 m opisan s vrijednostima reflektanci u 12 frekvencijskih područja detektiranih s multispektralnim instrumentom Sentinel-2 satelita i vrijednosti 0 ili 1 koja govori nalazi i se na tom pikselu nalazi prometnica ili ne.

2.3. Algoritmi

Klasifikacija je proces koji se koristi za grupiranje podataka u različite klase prema određenim kriterijima. Drugim riječima, klasifikacija je način generaliziranja podataka prema različitim karakteristikama [9]. U ovom radu cilj nam je kreirati klasifikator koji će temeljem reflektanci u 12 frekvencijskih područja (eksplanatorne varijable) predvidjeti prisutnost prometnice na zemljinoj površini. Set podataka koji je opisan u prijašnjoj sekciji podijeljen je na dva podseta: trening (80 %) i test (20 %) podaci. Kako bismo odabrali reprezentativne pristupe klasifikaciji

temeljene na strojnom učenju i vidjeli koji algoritmi daju najbolje rezultate u našem slučaju, istražili smo linearne, probabilističke i nelinearne pristupe. Za potrebe ovoga istraživanja koristili smo sljedeće klasifikacijske algoritme:

- *Naivni Bayesov klasifikator (engl. Naïve Bayes)* - radi na principu uvjetne vjerojatnosti i temelji se na Bayesovom teoremu pretpostavljajući neovisnost između eksplanatornih varijabli te izračunavajući distribuciju vjerojatnosti po klasama ishoda [10]. Pripada probabilističkim algoritmima strojnog učenja.
- *Algoritam stabla odluke (engl. Decision Tree)* - predstavnik algoritma koji logički i nadzirano uči. Strukturiran je kao stablo gdje svaki unutarnji čvor predstavlja pitanje koje razdvaja podatke, a čvorovi listovi predstavljaju odluke [11]. Pripada nelinearnim klasifikatorima.
- *Algoritam slučajnih stabala (engl. Random Forest)* - također jedan od nelinearnih klasifikatora. Koristi tehniku ansambla učenja gdje se stvara više stabala odlučivanja, što često dovodi do boljih rezultata u usporedbi s pojedinačnim stablima odlučivanja [12].
- *Višeslojni perceptron (engl. Multilayer perceptron)* - je najpoznatiji i najčešće korišten oblik neuronske mreže, nelinearni klasifikator. U većini slučajeva, signali se prenose unutar mreže samo u jednom smjeru - od ulaza prema izlazu. Nema povratnih petlji, što znači da izlaz svakog neurona ne utječe na sam neuron. Ova arhitektura se naziva prosljeđivanje unaprijed (eng. feedforward) [13].
- *Logistička regresija (engl. Logistic Regression)* - jedan od najčešće korištenih linearnih klasifikatora. Riječ je o algoritmu koji pretpostavlja linearnu vezu između objašnjavajućih varijabli i klasa rezultata. Ovaj algoritam odabire prag za granicu odlučivanja [14].
- *Linearna analiza diskriminanti (engl. Linear Discriminant Analysis)* - koristi linearnu kombinaciju značajki ili objašnjavajućih varijabli kako bi podatke razvrstao u različite klase [15].
- *Metode potpornih vektora (engl. Support Vector Machines (SVM))* – temelji se na principima statističke teorije učenja i konveksne optimizacije [16]. SVM ima svrhu smanjivanja gornje granice generalizacijske pogreške tako što maksimizira razmak između odvajajuće hiperplohe i podataka [17].

Za svaki od spomenutih klasifikatora korištena je Python biblioteka scikit-learn.

3. Rezultati

Nakon implementacije iznad opisanih algoritama, treniranja i testiranja podataka, koristili smo evaluacijske mjere koje se najčešće koriste kod klasifikacije. Konkretno, korištene su metrike *Precision, Recall, F-score, Balanced Accuracy, Accuracy*. Rezultat su prikazani u Tablici 1.

Tablica 1. Rezultati korištenih klasifikacijskih algoritama

Algoritam	Precision	Recall	F-score	Balanced accuracy	Accuracy
<i>Naivni Bayesov klasifikator</i>	0.03434	0.75023	0.06568	0.71496	0.68075
<i>Stablo odlučivanja</i>	0.22842	0.26141	0.24380	0.62400	0.97574
<i>Algoritam slučajnih stabala</i>	0.801169	0.06383	0.11825	0.53179	0.98575
<i>Logistička regresija</i>	0.17372	0.01910	0.03442	0.508862	0.98396
<i>Linearna analiza diskriminanti</i>	0.27995	0.38863	0.32546	0.68672	0.97590
<i>Višeslojni perceptron</i>	0.57142	0.17707	0.27036	0.58752	0.98570
<i>Metode potpornih vektora</i>	0.10344	0.00279	0.00544	0.50121	0.98472

4. Diskusija

Ako pogledamo rezultate evaluacijskih mjera, primjećujemo veliku razliku između vrijednosti točnosti (engl. *accuracy*) i balansirane točnosti (engl. *balanced accuracy*). Iako različite mjere izdvajaju različite algoritme, uzimajući u obzir balansiranu točnost, Naivni Bayesov algoritam dao je najbolje rezultate za detekciju cesta na satelitskim snimkama. To ne znači nužno da ostali algoritmi lošije rade, nego samo da je u ovakvom konkretnom slučaju dobro koristiti Naivni Bayesov algoritam. Važno je napomenuti da učinkovitost algoritama strojnog učenja može varirati ovisno o mnogim faktorima, uključujući skup podataka, odabir značajki, parametre modela i ostale čimbenike. Stoga, najbolji algoritam za određenu zadaću može se razlikovati ovisno o kontekstu i specifičnim uvjetima istraživanja. U budućem radu istražiti će se složeniji algoritmi koji osim značajki jednog piksela uzimaju u obzir i stanje okolnih piksela i kontekst s ciljem postizanja bolje preciznosti. Također, ispitat će se tehnike balansiranja skupa podataka jer su mjere pokazale da nebalansiranost između podataka o cestama i ostalih podataka pravi velik problem algoritmima.

Literatura

- [1] Campbell, J. B., Wynne, R. H.: Introduction to remote sensing, pp. 31 – 55, 285 – 286, 339–356, 2011, doi: 10.3390/rs5010282
- [2] Mather, P., Tso B. : Classification Methods for Remotely Sensed Data (2nd ed.), CRC Press, 2016, doi: 10.1201/9781420090741
- [3] Maxwell, Aaron E., Timothy A. Warner, Fang F. : Implementation of machine-learning classification in remote sensing: An applied review, International journal of remote sensing , pp. 2784 - 2817, 2018, doi: 10.1080/01431161.2018.1433343
- [4] SD Županija-Regionalne karakteristike Splitsko-dalmatinske županije, www.dalmacija.hr/zupanija/informacije, 21.02.2022.
- [5] SD Ceste-Županijska uprava za ceste Split, www.zuc-split.hr, 10.07.2023.
- [6] Ivanda, A. : Primjena strojnog učenja u promatranju Zemlje daljinskim istraživanjima, Kvalifikacijski ispit, 2022, www.fesb.unist.hr/wp-content/uploads/2022/06/A.Ivanda-Kvalifikacijski-rad.pdf
- [7] Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y., Ranagalage, M.: Sentinel-2 Data for Land Cover/Use Mapping: A Review, Remote Sens, pp. 2291, 2020, doi: 10.3390/rs12142291
- [8] Zhang, Yingjia, et al. : Density and diversity of OpenStreetMap road networks in China, Journal of Urban Management pp. 135 - 146, 2015, doi: 10.1016/j.jum.2015.10.001
- [9] Kumar, R., Rajesh V.: Classification algorithms for data mining: A survey. International Journal of Innovations in Engineering and Technology (IJJET), pp.7-14, 2012
- [10] Singh, A., N. Halgamuge, M, Lakshmganthan R. : Impact of different data types on classifier performance of random forest, naive bayes, and k-nearest neighbors algorithms, International Journal of Advanced Computer Science and Applications, 2017, doi: 10.14569/ijacsa.2017.081201
- [11] Priyam, Anuja, et al. : Comparative analysis of decision tree classification algorithms International Journal of current engineering and technology, pp. 334 - 337, 2013
- [12] Breiman, L.: Random forests, Machine learning 45, pp. 5-32,2001
- [13] Popescu, M-C., et al. : Multilayer perceptron and neural networks, WSEAS Transactions on Circuits and Systems, pp. 579 - 588, 2009
- [14] Tolles, J., Meurer, W.J. : Logistic regression: relating patient characteristics to outcomes., Jama, pp. 533-534, 2016, doi: 10.1001/jama.2016.7653
- [15] Izenman, A.J. : Linear discriminant analysis, Modern multivariate statistical techniques: regression, classification, and manifold learning, Springer - New York, New York, pp. 237-280, 2013, doi: 10.1007/978-0-387-78189-1_8
- [16] Mammone, A., Turchi, M., Cristianini, N. : Support vector machines, Wiley Interdisciplinary Reviews: Computational Statistics, pp. 283-289, 2009, doi: 10.1002/wics.49
- [17] Amari, S., Wu S. : Improving support vector machine classifiers by modifying kernel functions, Neural Networks, pp. 783 - 789, 1999, doi:10.1016/s0893-6080(99)00032-5