

Primjena dubokih neuronskih mreža za predviđanje dnevne koncentracije lebdećih čestica u zraku

Petrić, Valentino

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Chemical Engineering and Technology / Sveučilište u Zagrebu, Fakultet kemijskog inženjerstva i tehnologije**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:149:596824>

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

Download date / Datum preuzimanja: **2024-07-28**



Repository / Repozitorij:

[Repository of Faculty of Chemical Engineering and Technology University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE
SVEUČILIŠNI DIPLOMSKI STUDIJ

Valentino Petrić

DIPLOMSKI RAD

Zagreb, srpanj 2022.

SVEUČILIŠTE U ZAGREBU
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE
SVEUČILIŠNI DIPLOMSKI STUDIJ

Valentino Petrić

DIPLOMSKI RAD

Primjena dubokih neuronskih mreža za predviđanje dnevne koncentracije
lebdećih čestica u zraku

Voditelj rada: doc. dr. sc. Željka Ujević Andrijić

Članovi ispitnog povjerenstva:

doc. dr. sc. Željka Ujević Andrijić

doc. dr. sc. Dajana Kučić Grgić

izv. prof. dr. sc. Gordana Peh nec

Ovaj diplomski rad izrađen je na istraživačkom institutu Know-Center, Graz, Austrija pod vodstvom doc. dr. sc. Marija Lovrića, te Zavoda za mjerenja i automatsko vođenje procesa Fakulteta kemijskog inženjerstva i tehnologije Sveučilišta u Zagrebu pod mentorstvom doc. dr. sc. Željke Ujević Andrijić

Prije svega zahvalio bih se svojim mentorima doc. dr. sc. Mariju Lovriću i doc. dr. sc. Željki Ujević Andrijić koji su svojim brojnim stručnim savjetima, znanjem i strpljivošću doprinijeli izradi ovog rada.

Ovim putem bih se zahvalio i svim partnerima Know-Centra za dostavu potrebnih podataka za izradu ovog rada: Michael Haberl, dipl.-ing. sa Instituta za ceste i promet na Tehničkom Sveučilištu u Grazu, Graz, Austrija; Gradskom uredu za promet, Graz, Austrija; Milana Vučković iz Europskog centra za vremenske prognoze, Reading, Ujedinjeno Kraljevstvo; Andreas Schopper, mag. iz Gradskog ureda za promatranje zraka, Graz, Austrija.

Na kraju veliko hvala i mojoj obitelji na beskrajnoj podršci kako tijekom izrade ovog rada tako i tijekom cijelog mog akademskog obrazovanja.

„It's not the size of the dog in the fight, it's the size of the fight in the dog.“

~ Mark Twain

Sažetak

Cilj ovog rada je razviti modele za procjenu koncentracije lebdećih čestica u zraku pomoću metoda umjetne inteligencije. U radu su primijenjene metoda nasumičnih šuma i metoda neuronskih mreža koristeći programski jezik Python i pripadajuće knjižnice. Mjerni podaci korišteni u ovom radu odnose se na austrijski grad Graz čiji su meteorološki podaci mjereni na pet postaja u gradu: Sjever, Istok, Zapad, Jug i Don Bosco. Uz meteorološke podatke na postajama je mjerena i koncentracija lebdećih čestica u zraku. Uz lokalne meteorološke podatke korišteni su i satelitski podaci preuzeti sa stranice *Climate Data Store*. Koristeći lokalne i satelitske podatke razvijeni su modeli za svih pet postaja te je ispitivano koji podaci će pridonijeti boljoj kvaliteti modela. Rezultati su pokazali da su modeli razvijeni sa lokalnim meteorološkim podacima nešto točniji u usporedbi s modelima razvijenih sa satelitskim meteorološkim podacima. U ovom radu se ispitivalo i kako podaci za promet mjereni na šesnaest prometnica u gradu utječu na koncentraciju lebdećih čestica. Uvođenjem prometa kao dodatne ulazne varijable za razvoj modela postignuto je poboljšanje kvalitete modela posebno na skupu s lokalnim podacima. U konačnici su se svi dostupni podaci spojili u dva skupa, pri čemu su prvi skup činili lokalni i satelitski meteorološki podaci, a drugi skup su činili lokalni i satelitski meteorološki podaci kojima je pridodan i promet. Kod algoritma nasumičnih šuma skup sa svim podacima, tj. skup sa lokalnim i satelitskim meteorološkim podacima i prometom se pokazao najboljim, dok je kod algoritma neuronskih mreža taj skup bio nešto lošiji od skupa s lokalnim meteorološkim podacima i prometom. Razvijeni modeli pokazali su zadovoljavajuću točnost te se u budućnosti mogu primijeniti za predviđanje koncentracije lebdećih čestica na novim skupovima podataka.

Ključne riječi: lebdeće čestice, strojno učenje, Graz, zaštita okoliša, meteorološki podaci, promet

Summary

The goal of this study is the application of artificial intelligence methods to develop models for estimating the concentration of particle matter in the air. Random forest regression and neural networks models were developed using the Python programming language and its libraries. Meteorological data used in this study is related to the Austrian city of Graz measured at five stations, namely: North, East, West, South and Don Bosco. In addition to meteorological data, the concentration of particle matter in the air is also measured at the stations. Additionally, satellite data from the Climate Data store is also used. Using these data, models were developed for all the stations and it was observed which data would contribute more to the final quality of the model. It has been proven that models with locally acquired meteorological data are better than models with satellite meteorological data. Furthermore, traffic data obtained by measuring traffic on sixteen roads in the city is also used in this study. Traffic data has been added to the set with local and satellite meteorological data. It has been proven that traffic as an additional variable improved the quality of the model's assessments. This increase was more pronounced in the set with local data. In the end, all the above data is combined into two sets, where the first set consisted of local and satellite meteorological data, and the second set consisted of local and satellite meteorological data to which traffic was also added. With the random forest algorithm, the model developed using the set with all the data, i.e. the local and satellite meteorological with traffic, turned out to be the best. However, with the neural network algorithm, the model using all data show slightly worse results in comparison with neural network model developed on set with local meteorological data and traffic. The developed models showed satisfactory accuracy and can be applied to estimating the concentration of particle matter in the air using the new data set.

Key words: particulate matter, machine learning, Graz, environmental protection, meteorological data, traffic

SADRŽAJ

1	Uvod.....	1
2	Teorijski uvod.....	2
2.1	Onečišćenje zraka.....	2
2.2	Glavni izvori onečišćenja u gradovima.....	4
2.3	Lebdeće čestice.....	5
2.4	Strojno učenje.....	6
2.4.1	Algoritam nasumičnih šuma.....	7
2.4.2	Neuronske mreže.....	8
2.4.3	Optimizacija modela.....	10
2.4.4	Inženjerstvo značajki.....	11
2.4.5	Mjere ocjene kvalitete modela.....	12
2.5	Graz.....	13
2.6	Kvaliteta zraka u Grazu.....	14
2.7	Python i njegove knjižnice.....	14
3	Materijali i metode.....	16
3.1	Prikupljanje podataka.....	16
3.2	Meteorološki podaci.....	18
3.2.1	Vjetar.....	18
3.2.2	Temperatura zraka.....	20
3.2.3	Relativna vlažnost zraka.....	21
3.2.4	Tlak zraka.....	22
3.2.5	Oborine i radijacija.....	23
3.2.6	Satelitski podaci.....	24
3.3	Promet.....	25
3.3.1	Istok.....	25
3.3.2	Don Bosco.....	26
3.3.3	Sjever.....	27
3.3.4	Zapad.....	28
3.3.5	Jug.....	29
3.4	Temporalni podaci.....	29
3.5	Koncentracija lebdećih čestica.....	30
3.6	Inženjerstvo značajki.....	31
4	Ekperimentalni dio.....	32
4.1	Odabir hiperparametara i značajki.....	32

4.2	Treniranje modela	35
5	Rezultati i rasprava	37
5.1	Rezultati optimizacije modela.....	37
5.2	Satelitski i lokalni meteorološki podaci.....	40
5.3	Doprinos prometa modelu za procjenu lebdećih čestica.....	44
5.4	Usporedba modela razvijenog sa svim dostupnim podacima	47
5.5	Odabir najboljeg algoritma i skupa podataka.....	50
6	Zaključak.....	51
7	Popis oznaka i kratica	52
8	Literatura.....	53
9	Životopis	56

1 UVOD

Onečišćenje zraka predstavlja sve veći problem današnjice jer šteti zdravlju ljudi i okolišu. Onečišćenje zraka je izraženije u urbanističkim središtima s mnogo automobila i industrijskih postrojenja, mada i ruralna središta bilježe sve lošiju kvalitetu zraka. U posljednje vrijeme posljedice onečišćenja zraka su sve vidljivije te se procjenjuje da godišnje sedam milijuna ljudi umire upravo zbog loše kvalitete zraka¹. Kvaliteta zraka ovisi ponajprije o koncentraciji štetnih tvari u zraku. Štetne tvari koje se pojavljuju u zraku su: ugljikov dioksid CO₂, sumporovi oksidi SO_x, dušikovi oksidi NO_x, ugljikov monoksid CO, hlapljive organske komponente, toksični metali, klorfluorugljici, amonijak i lebdeće čestice. Lebdeće čestice smatraju se najopasnijim onečišćenjem u zraku, a to su čestice koje čovjek svakodnevno udiše te tako utječu na krvožilni i respiratorni sustav². Lebdeće čestice u zraku rezultat su antropogenih i prirodnih aktivnosti. Utjecaj iz antropogenih aktivnosti je veći od prirodnih utjecaja, ali se ni prirodni utjecaji ne smiju zanemariti. Glavni izvori lebdećih čestica su promet i nepotpuno izgaranje fosilnih goriva³. Zbog sve većeg utjecaja koje onečišćenje zraka ima na svakodnevni život u literaturi se može naći dosta radova na tu temu. Između ostalih metoda u literaturi se može naći i primjena metoda umjetne inteligencije za procjenu koncentracije lebdećih čestica u zraku. Umjetna inteligencija pokazala se izvrsnim alatom u rješavanju ovako složenih matematičkih problema s velikim brojem praćenih veličina odnosno nezavisnih varijabli. U radu Lovrića i suradnika⁴ dokazano je kako meteorološki i temporalni podaci utječu na kvalitetu zraka, stoga je u ovom radu postavljena hipoteza da će satelitski podaci doprinijeti poboljšanju primijenjenih algoritama. Još jedna od hipoteza koja će se pokušati potvrditi je da promet i količina automobila imaju utjecaj na kvalitetu zraka u gradovima, tj. da podaci za promet poboljšavaju kvalitetu modela za predviđanje lebdećih čestica. Za ovaj diplomski rad za primjenu predložene metodologije uzet je austrijski grad Graz za koji su bili javno dostupni navedeni meteorološki podaci na osnovu kojih je bilo moguće provesti ovo istraživanje. U radu su za razvoj modela za procjenu koncentracije lebdećih čestica primijenjena dva algoritma umjetne inteligencije, a to su algoritam nasumičnih šuma i algoritam neuronskih mreža.

2 TEORIJSKI UVOD

2.1 ONEČIŠĆENJE ZRAKA

Pojam onečišćenja zraka odnosi se na prisutnost tvari u atmosferi koje su štetne po zdravlje ljudi i drugih živih bića ili uzrokuju promjenu u klimi. Danas je prisutan veliki broj onečišćivača zraka u atmosferi, koji dolaze u tri oblika: kao plinovi, čestice i biološke molekule. Pod plinove onečišćivača zraka spadaju: ugljikov monoksid, amonijak, sumporov dioksid, dušikovi oksidi, metan, ugljični monoksid i klorfluorugljici. Čestice koje se nalaze u zraku mogu biti organskog ili anorganskog podrijetla⁵. Prisutnost štetnih tvari u zraku kod ljudi može izazvati alergije, bolesti i u nekim slučajevima čak i smrt. Bolesti koje se povezuju s kvalitetom zraka su: respiratorne bolesti, bolesti srca, kronična opstruktivna plućna bolest, moždani udar i rak pluća⁶. Provedena istraživanja ukazuju da onečišćenje zraka utječe čak i na koncentraciju i na kvocijent inteligencije⁷. Uz to, dokazano je i da štetne tvari u zraku utječu na psihičko zdravlje ljudi, pa je tako u gradovima s onečišćenim zrakom veća stopa ljudi koji boluju od depresije⁷. Štetan utjecaj onečišćenja zraka može se vidjeti i u prirodi u obliku klimatskih promjena, oštećenja ozona, kiselih kiša itd. Vanjsko onečišćenje zraka povezuje se s 4,21 milijuna smrtnih slučajeva godišnje i to ga čini jednim od najvećih svjetskih uzročnika smrti⁸. Prema istraživanjima WHO (Svjetska zdravstvena organizacija), 90% svjetske populacije u većoj ili manjoj mjeri udiše štetne tvari iz zraka⁹. Procjenjuje se da onečišćenje zraka košta svjetsko gospodarstvo 5 milijardi dolara godišnje [10]. Kvaliteta zraka regulira se nacionalnim i međunarodnim propisima i na mjestima gdje se oni provode vidljivo je znatno poboljšanje u kvaliteti zraka. Jedan od uspješnijih međunarodnih propisa je Helsinški protokol koji je smanjio emisije sumpora u atmosferi. Popis tvari koje utječu na onečišćenje zraka:

- Ugljikov dioksid, CO₂
- Sumporovi oksidi, SO_x
- Dušikovi oksidi, NO_x
- Ugljikov monoksid, CO
- Hlapljive organske komponente (acetona, octena kiselina, butanal, etanol, formaldehid, metilen klorid i ugljični disulfid)
- Sitne čestice
- Toksični metali (olovo i živa)
- Klorfluorugljici
- Amonijak

Izvori onečišćenja dijele se na ljudske i prirodne. Ljudski izvori onečišćenja imaju znatno veći utjecaj na kvalitetu zraka od onih koji potječu iz prirode. Onečišćenje uzrokovano ljudskim djelovanjem najviše se odnosi na izgaranje fosilnih goriva. Neki od ljudskih izvora onečišćenja su: termoelektrane koje za pogon koriste ugljen, biomasu (drvo, sijeno i otpad s usjeva) ¹⁰, proizvodni pogoni kao npr. tvornice, spaljivanje otpada, pogrešno zbrinjavanje otpada, požari te isto tako i grijanje domova i poslovnih prostora. Uz sve navedeno tu su još motorna vozila, vlakovi, brodovi, zrakoplovi. Na kvalitetu zraka još utječu i isparavanja boja i lakova te drugih otapala. Međutim, bitno je spomenuti i onečišćenja koja su dio prirode, a to su: pustinjske oluje koje donose sitne čestice prašine vjetrom, metan koji probavom proizvode životinje, raspad radioaktivnog radona koji se nalazi u Zemljinoj kori. Veliko onečišćenje zraka uzrokuju šumski požari koji u trenutku izgaranja uzrokuju 75% ukupnog onečišćenja zraka. Neke vrste biljaka će tijekom ljetnih mjeseci emitirati značajne količine hlapljivih organskih spojeva koji u doticaju s NO_x, SO₂ i ugljikovim spojevima stvaraju maglu¹¹. Postoje mjesta, tzv. vruće točke (eng. *Hotspots*) u kojima se ljudi izlažu većim emisijama onečišćenja zraka i koja dodatno utječu na zdravlje. Takva mjesta vrlo su česta u urbanim područjima gdje je veliki broj automobila i kamiona te industrijskih postrojenja. Ovdje je bitno spomenuti dizelsku čađu koja ulazi u pluća i krvotok. Gusto naseljena područja kao velike metropole karakterizira veliko onečišćenje zraka. Takva područja posebno su izražena u zemljama u razvoju gdje gradovi doživljavaju brzi rast, a zakoni za očuvanje okoliša su slabi ili ne postoje. Urbanizacija dovodi do većeg onečišćenja zraka pa tako i prerane smrti. Međutim, kvaliteta zraka nije loša samo u gradovima koji se nalaze u zemljama u razvoju, neki gradovi u razvijenim zemljama također imaju lošu kvalitetu zraka. Gradovi s lošom kvalitetom zraka su Los Angeles (SAD) i Rim (Italija) te Peking (Kina) ¹². U tablici 1. prikazano je 5 gradova s najvećim onečišćenjem u 2021. godini.

Tablica 1. Pet gradova s najvećim onečišćenjem zraka u 2021. godini prema koncentraciji PM 2.5 čestica u zraku ($\mu\text{g}/\text{m}^3$) ¹³

Najonečišćeniji gradovi na svijetu u 2021.	Godišnji prosjek PM 2.5
Bhiwadi, Indija	106,2
Ghaziabad, Indija	102
Hotan, Kina	101,5
Delhi, Indija	96,4
Jaunpur, Indija	95,3

Austrija se prema kvaliteti zraka nalazi na 83. mjestu od 118 država, te je u prošloj godini najonečišćeniji grad bio Neudorf bei Staats, a najčišći grad je bio Ebensee. S druge strane, Hrvatska npr. se nalazi na 35. mjestu po kvaliteti zraka, a u 2021. godini najonečišćeniji grad bio je Zagreb, a najčišći Polaca, Croatia¹³.

2.2 GLAVNI IZVORI ONEČIŠĆENJA U GRADOVIMA

Postoji pet kategorija glavnih izvora onečišćenja u gradovima i to su redom: promet, industrija, izgaranje fosilnih goriva za grijanje, prirodni izvori kao što su prašina i morska sol i na kraju nespecifično onečišćenje koje uzrokuje čovjek¹⁴. Promet uzrokuje veliki dio onečišćenja u gradovima, budući da se tijekom vožnje iz vozila ispuštaju razni štetni plinovi i sitne čestice. Tijekom izgaranja goriva i raspadom maziva nastaju sitne organske i anorganske čestice, a uz to vozila proizvode sitne čestice i kočenjem te trošenjem guma¹⁵. Nastale čestice talože se na cestama te se uz pomoć ostalih vozila u prometu suspendiraju s prašinom i ostalim nosivim materijalima. Industrija proizvodi heterogena onečišćenja koja su povezana s izgaranjem nafte i ugljena u elektranama, emisijskim plinovima iz proizvodnje u industrijama kao što su petrokemijska, metalurgija, kemijska, farmaceutska i IT (proizvodnja hardvera). Veliki dio industrijskog onečišćenja čine i luke koje se nalaze u velikim gradovima poput Hamburga¹⁶. Izgaranje fosilnih goriva za potrebe grijanja kućanstava također je bitan izvor onečišćenja u gradovima. Goriva koja se koriste za grijanje su drvo, ugljen i prirodni plin. Prirodni plin se uz grijanje koristi i za kuhanje te je dodatan izvor onečišćenja. U srednjoj Europi je značajno onečišćenje iz ovih izvora zato što se za grijanje pretežno koristi drvo i ugljen. Onečišćenje iz prirode uključuje prašinu tla i morsku sol. Prašina nastaje u područjima s velikim stijenama iz Zemljine kore i tla. Kod ovog onečišćenja bitnu ulogu ima vjetar jer on suspendira čestice prašine s česticama nastalim na prometnicama. Čestice morske soli u zraku prisutne su u područjima gdje je u blizini morska obala, a osim što utječu na kvalitetu zraka, izazivaju i koroziju. Također, jedan od izvora soli u zraku može biti posipanje prometnica solju kako bi se smanjilo ledište vode. U nespecifična onečišćenja koje je uzrokovao čovjek spadaju sekundarne čestice koje nastaju reakcijom primarnih čestica onečišćenja (NO_2 , NH_3 , SO_2) i nemetanskih hlapljivih organskih spojeva NMVOC. Nastale sekundarne čestice su organski ili anorganski aerosoli. Anorganski aerosoli nastaju iz poljoprivrednog NH_3 koji reagira s SO_2 , SO_2 nastaje tijekom transporta i proizvodnje električne energije. Anorganski nitratni aerosoli nastaju reakcijom NH_3 i NO_x . Anorganski aerosoli nastali na ovaj način povezuju se s transportom na duge udaljenosti jer se sulfati i

nitriti zadržavaju u atmosferi 3 do 9 dana¹⁷. Organski aerosoli mogu biti izravno emitirani u atmosferu i to izgaranjem drva, biomase, ugljena i organskih tvari. Također, organski aerosoli mogu nastati kombinacijom s drugim sekundarnim česticama ili reakcijom s NMVOC.

2.3 LEBDEĆE ČESTICE

Krute čestice ili tekuće tvari koje su suspendirane u zraku nazivaju se atmosferske čestice ili skraćeno PM (eng. *Particulate Matter*), a postoji još i naziv atmosferski aerosoli. Pojam aerosoli odnosi se na smjesu čestica i zraka¹⁷. Izvori ovih čestica u zraku mogu biti prirodnog ili ljudskog podrijetla, a same čestice utječu na klimu i oborine te na taj način i na ljudski organizam. Vrste čestica koje se suspendiraju sa zrakom su:

- grube čestice koje se mogu udahnuti, označene kao PM10 i čiji je promjer 10 μm ili manje
- fine čestice PM2.5 čiji je promjer 2.5 μm ili manje
- ultra fine čestice promjera 100 nm ili manje
- čađa

Međunarodna agencija za istraživanje raka (IARC) i Svjetska zdravstvena organizacija (WHO) označavaju PM kao glavnog uzročnika karcinoma¹⁸. Najštetnije čestice za ljudsko zdravlje su ultra fine čestice zbog svoje sposobnosti da prodiru duboko u pluća, krvotok i mozak. Neke od bolesti koje uzrokuju ove vrste čestica u organizmu su srčani udar i respiratorne bolesti. U radu objavljenom u 2013. godini dokazano je da s povećanjem koncentracije PM10 čestica u zraku od 10 $\mu\text{g}/\text{m}^3$ stopa raka pluća raste za 22 % te da PM2.5 povećava stopu raka pluća za 18% kod rasta koncentracije od 5 $\mu\text{g}/\text{m}^3$ ¹⁹. Čestice koje su prisutne u zraku šesti su uzročnik prerane smrti u svijetu. Kao što je ranije rečeno, čestice u atmosferi potječu i iz prirodnih izvora kao što su: vulkani, pustinjske oluje, požari, biljke i valovi na moru. Prisutnost čestica u atmosferi utječe i na klimu i to tako da mijenja količinu dolaznog Sunčevog zračenja i povratak zračenja u atmosferu. Postoje dva efekta koja opisuju promjenu klime zbog prisustva čestica u zraku, a to su:

- izravni
- neizravni

Kod izravnih efekata čestice reagiraju sa zračenjem i to tako da apsorbiraju ili raspršuju dolazno zračenje. Neizravni efekt se postiže kada se oko čestica u zraku formira kapljica koja reagira sa zračenjem. Čestice koje su nastale iz ljudskih izvora su manje pa su i brojnije od čestica iz prirodnih izvora. Zbog svoje veličine čestice nastale iz ljudskih izvora su i opasnije po ljudsko zdravlje jer prodiru dulje u čovjekov organizam.

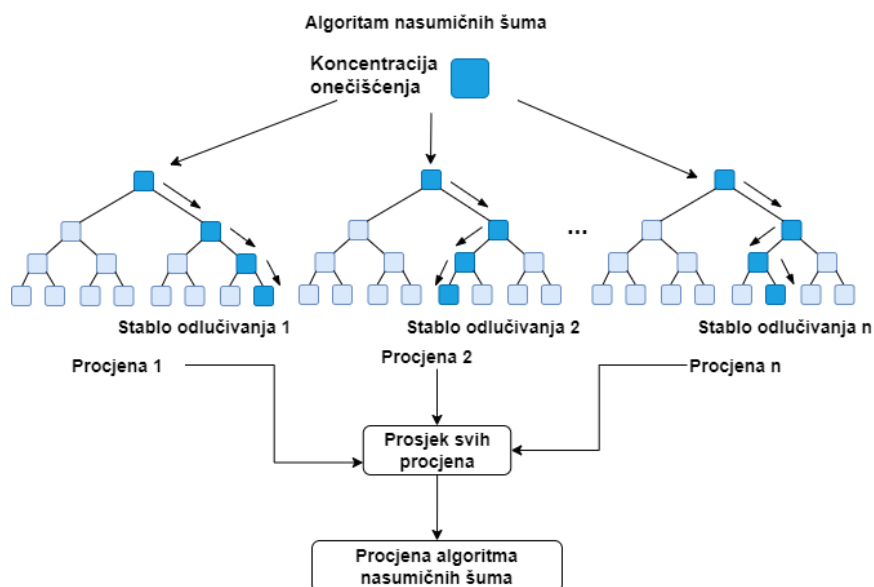
2.4 STROJNO UČENJE

Strojno učenje je programiranje računala na način da optimiziraju neki kriterij uspješnosti temeljem podatkovnih primjera ili prethodnog iskustva. Strojno učenje dio je veće cjeline koja se naziva umjetna inteligencija. Postoje tri načina strojnog učenja, a to su nadzirano, nenadzirano i podržano/ojačano učenje. Kod nadziranog učenja podaci su parovi gdje x predstavlja vrijednosti ulaza, a y vrijednosti izlaza te je potrebno pronaći vezu između ulaza i izlaza. U nadziranom učenju postoje dva pristupa učenju: regresija i klasifikacija. Kada je izlazna vrijednost diskretna, tj. nebrojčana vrijednost koristi se klasifikacija, a kada je izlaz kontinuirana, tj. brojčana vrijednost koristi se regresija²⁰. U ovom radu primijenit će se dvije nelinearne metode strojnog učenja, a to su metoda slučajnih šuma (engl. *random forests*, RF) i umjetne neuronske mreže (engl. *neural networks*, NN). U nenadziranom učenju podaci su dani bez ciljanih vrijednosti, tj. potrebno je pronaći neku pravilnost u podacima. Postoje tri pristupa nenadziranom učenju, a oni su: grupiranje (engl. *clustering*), procjena gustoće (engl. *density estimation*) i smanjenje dimenzionalnosti (engl. *dimensionality reduction*). Treći način strojnog učenja je podržano ili ojačano učenje (engl. *reinforcement learning*) gdje se učenje optimalne strategije odvija na temelju pokušaja s odgođenom nagradom (Lovrić, 2021). Postoji 8 koraka u procesu strojnog učenja, a oni su²¹:

- **Priprema podataka i preliminarna analiza** - prije početka razvoja modela potrebno se upoznati s podacima i razumjeti ih
- **Podjela podataka na set za učenje i set za ispitivanje/vrednovanje** – ponekad se ovaj korak može i preskočiti kada se radi o nenadziranom učenju
- **Ekstrakcija značajki** – odabrati na koji će se način modelu prikazati vektori podataka
- **Redukcija dimenzionalnosti** – ovaj korak je potreban kada je prisutan velik broj značajki
- **Odabir modela** – odabir vrste modela koji će se koristiti za strojno učenje
- **Učenje modela** – model uči iz zadanih podataka, a ovisno o složenosti algoritma ovaj proces može trajati kraće ili duže
- **Vrednovanje modela** – provjera modela
- **Dijagnostika i korigiranje modela** – ako model ne predviđa dobro potrebno ga je modificirati kako bi se dobili bolji rezultati

2.4.1 Algoritam nasumičnih šuma

Algoritam nasumičnih šuma (engl. *Random Forest*, RF) sastoji se od mnoštva stabala odlučivanja. Tijekom treniranja RF algoritma stabla uče i to po nekoliko stotina stabala odjednom. U slučajevima kada se radi o klasifikacijskim modelima, odnosno kod modela u kojima je prisutna binarna klasifikacija (1 ili 0) sva stabla sudjeluju u glasanju. Kada se radi o regresijskim modelima uzimaju se u obzir sva stabla i određuje se njihova srednja vrijednost. Kvaliteta modela testira se na neovisnom skupu podataka, a modeli se poboljšavaju optimalnim odabirom tzv. hiperparametara o čemu će biti govora u idućim poglavljima. Ovaj algoritam se može koristiti za klasifikacijske i regresijske probleme i to mu je velika prednost. Prednost RF algoritma je i ta da sprječava pretreniranje modela zato što se procjena modela temelji na prosjeku ili većinskom glasanju. RF algoritam daje dobra predviđanja čak i kada u skupovima podataka nema određenih vrijednosti te je svako stablo neovisno jedno o drugom pa je prisutno svojstvo paralelizacije. Dimenzionalnost se ne uzima u obzir kod ovih algoritama jer svako stablo dobiva samo dio atributa pa je prostor značajki smanjen. Uz sve prednosti ovaj algoritam posjeduje i nekoliko nedostataka, a oni su: znatno veća složenost kada se uspoređuju sa stablima odluke te samim time i duže vrijeme koje je potrebno za treniranje modela. Naime, kako bi algoritam kreirao pretpostavku svako stablo treba donijeti odluku i oblikovati izlaz²². Na slici 1. prikazana je shema algoritma nasumičnih šuma. Na slici je vidljivo kako svako od stabala donosi svoju procjenu i zatim se računa prosjek procjena i dobiva se konačna vrijednost procjene izlazne vrijednosti algoritma. Svako od stabala odlučuje neovisno jedno o drugome i formira svoju procjenu.



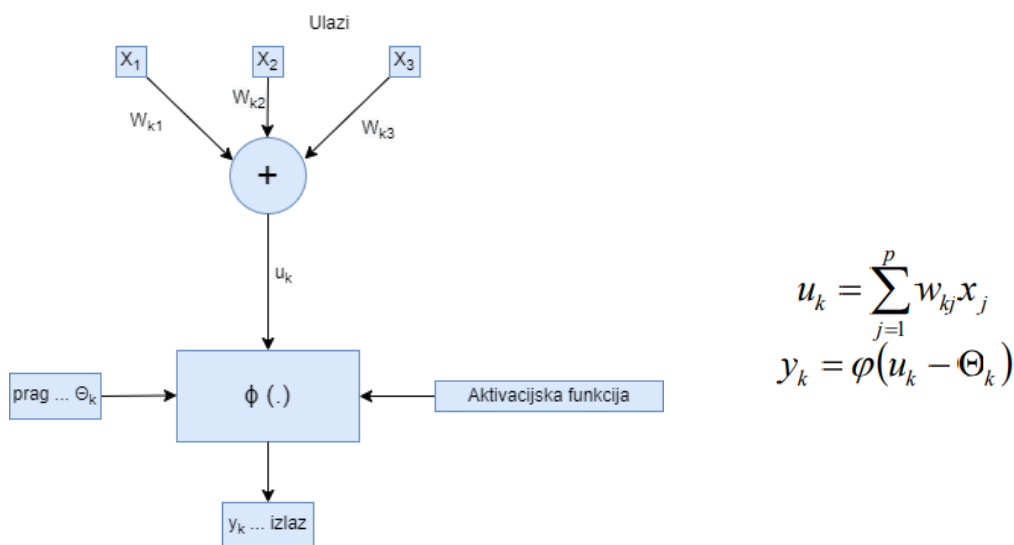
Slika 1. Algoritam nasumičnih šuma

2.4.2 Neuronske mreže

Umjetne neuronske mreže nastale su po uzoru na biološke neuronske mreže, a primjenjuju se u digitalnim računalima ili specijaliziranim sklopovima. Umjetna neuronska mreža predstavlja ogroman paralelni distributivni procesor koji je izvrstan u pamćenju znanja naučenog na iskustvu. Postoji sličnost između prirodne i umjetne neuronske mreže, a to je da se znanje stječe iskustvom, a veze između neurona se koriste za spremanje informacija. Zbog jednostavnosti, u nastavku teksta, umjetne neuronske mreže nazivat će se neuronske mreže. Svojstva neuronskih mreža su: nelinearnost, učenje s nadzorom (potrebni ulaz i izlaz), adaptivnost, tolerancija na grešku itd.

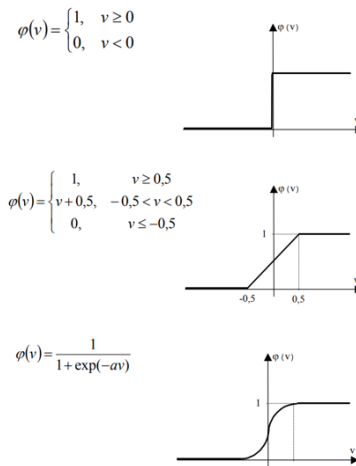
Neuron u mreži sastoji se od nekoliko elemenata, a oni su²³:

- Skup sinapsi, tj. skup ulaza od kojih svaki ima svoju težinu kojom pridonosi modelu
- Sumator koji se koristi za računanje linearnih kombinacija ulaza
- Nelinearna aktivacijska funkcija koja zadržava izlaz u rasponu od 0 do 1



Slika 2. Model neurona u neuronskoj mreži

Na slici 2. vidljivo je da je svaki neuron u neuronskoj mreži izgrađen od nekoliko ulaza koji dolaze sa svojim težinama, nakon toga otežani ulaze prolaze kroz sumator označenim simbolom +. Nakon sumiranja na ukupnu vrijednost se primjenjuje aktivacijska funkcija i prag koji se može prikazati i kao ulaz sa iznosom 1 i težinom Θ_k . Postoji nekoliko vrsta aktivacijskih funkcija koje se primjenjuju, a one su: aktivacijska funkcija tipa prag, aktivacijska funkcija linearna po odsječcima i sigmoidna aktivacijska funkcija. Na slici 3. prikazane su sve funkcije navedenim redom²³.

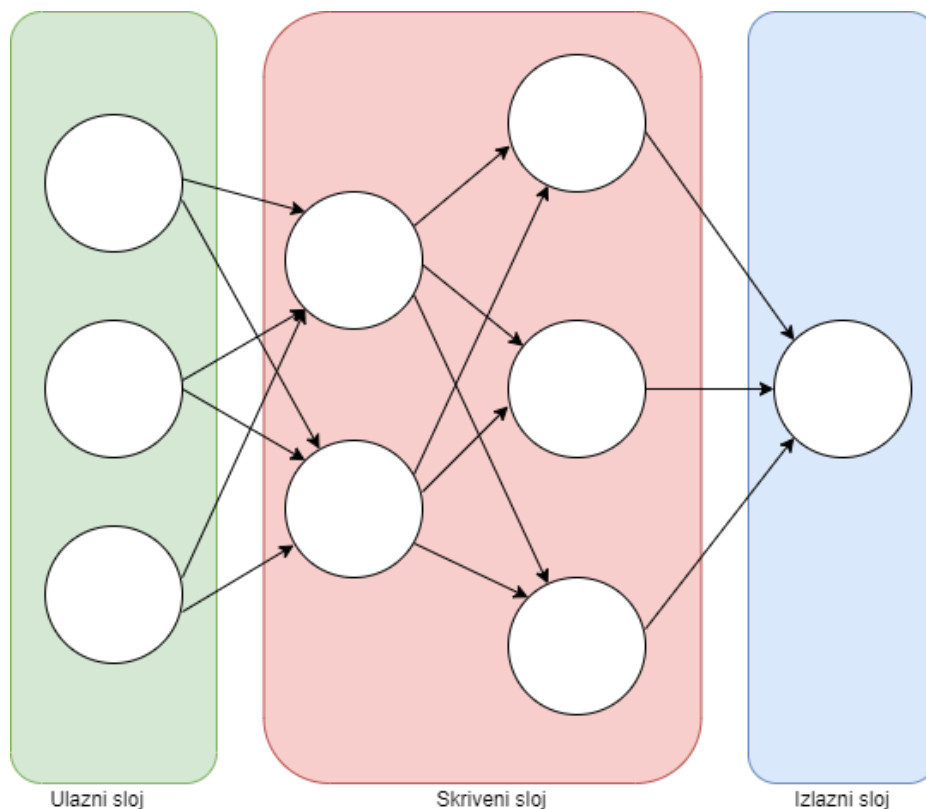


Slika 3. Aktivacijske funkcije u neuronima (gore: tipa prag, sredina: linearna po odsječcima, dolje: sigmoidna)

Kod razvoja neuronskih mreža bitna je i arhitektura mreže, odnosno način na koji su povezani neuroni u neuronskim mrežama. Postoje četiri glavne vrste neuronskih mreža i to su redom:

- Jednoslojne mreže bez povratnih veza (engl. *single-layer feedforward networks*)
- Višeslojne mreže bez povratnih veza (engl. *multi-layer feedforward networks*)
- Mreže s povratnim vezama (engl. *recurrent networks*)
- Ljestvičaste mreže (engl. *lattice structures*)

Najjednostavnije mreže su jednoslojne mreže koje se sastoje od jednog sloja neurona i to samo izlaznog sloja, ulazni sloj se ne broji jer u njemu ne postoji nikakvo procesiranje podataka. Kod ovih vrsta mreža ne postoje povratna veza s izlaza na ulaz. Složenija vrsta mreže su višeslojne mreže koje osim ulaznih i izlaznih slojeva imaju i jedan ili više skrivenih slojeva. Kada mreže posjeduju više skrivenih slojeva nazivaju se duboke neuronske mreže. Kod višeslojnih mreža javlja se i pojam povezanosti mreža, gdje se mreže u kojima je svaki neuron u svakom sloju povezan na sve neurone u sljedećem sloju nazivaju potpuno povezane mreže. Djelomično povezane mreže su mreže gdje nedostaje veza između nekih neurona. Na slici 4. prikazana je arhitektura višeslojne potpuno povezane neuronske mreže²³.



Slika 4. Arhitektura neuronske mreže

Treća vrsta mreža su mreže s povratnom vezom, koje ne moraju imati neurone u skrivenom sloju, a karakterizira ih prisutnost povratne veze koja daje dodatnu kvalitetu mreži. Međutim, ove mreže su složenije za rad od prethodno navedene dvije vrste mreža. Na kraju, zadnja vrsta neuronske mreže je ljestvičasta mreža koja se sastoji od neurona koji su izrađeni od jednodimenzionalnih, dvodimenzionalnih ili višedimenzionalnih polja s skupom izvornih čvorova koji daju ulazne signale. Svaki ulaz spojen je na sve neurone u polju i u ovoj vrsti mreže ne postoji povratna veza već su svi neuroni raspoređeni u polja²³.

2.4.3 Optimizacija modela

Kod razvoja modela bitno je odrediti optimalne hiperparametre i značajke modela koje će dati najbolju generalizaciju mreže. Odabiru se hiperparametri koji daju minimalnu pogrešku modela odnosno koji daju najbolje rezultate procjene izlaza modela. Ručno isprobavanje najboljih hiperparametara (po principu pokušaja i pogreške) trajalo bi predugo i zbog toga postoje algoritmi koji su razvijeni za tu primjenu. Jedan takav algoritam je *GridSearchCV* koji je dio Pythonove knjižnice *sklearn*. Optimizacija hiperparametara se provodi tako da se rječnik hiperparametara prosljeđuje u model i za svaku kombinaciju parametara radi se unakrsna provjera (engl. *Cross validation*).

Za svaku kombinaciju parametara dobiva se određena točnost modela i određuju se najbolji parametri²⁴. Hiperparametri mreže koju su se optimirali u ovom radu navedeni su u tablici 2.

Tablica 2. Hiperparametri modela neuronske mreže optimirani u ovom radu ^{25,26}:

Model	Hiperparametar	Značenje
Algoritam nasumičnih šuma	<i>n_estimators</i>	Broj stabala u šumi
	<i>max_depth</i>	Maksimalna dubina stabla
	<i>n_jobs</i>	Broj procesora koji se koristi
	<i>random_state</i>	Kontrolira slučajnost i izgradnju stabala
	<i>max_samples</i>	Broj uzoraka koji se uzima iz trening seta za procjenu u svakom stablu
	<i>min_samples_leaf</i>	Minimalan broj uzoraka koji je potreban u čvoru lista
Neuronske mreže	<i>hidden_layer_sizes</i>	Broj skrivenih slojeva i neurona u njima
	<i>solver</i>	Optimizacija težina ulaznih varijabli
	<i>learning_rate_init</i>	Početna stopa učenja
	<i>max_iter</i>	Maksimalan broj iteracija
	<i>early_stopping</i>	Zaustavljanje ako se validacijski rezultat ne mijenja
	<i>random_state</i>	Kontrolira slučajnost i izgradnju mreže
	<i>alpha</i>	Jačina L2 regulacijskog člana

Uz hiperparametre u radu su optimirane i značajke modela. Značajke modela se optimiraju pomoću *Permutation feature importance* algoritma koji je dio *sklearn* knjižnice. Algoritam radi tako da se modelu dodaju podaci i pokušava se prekinuti odnos između pojedinih značajki i tražene varijable. Svaka od ulaznih varijabli izdvaja se iz skupa podataka te se određuje kolika je njezina ovisnost o kvaliteti modela.

2.4.4 Inženjerstvo značajki

Inženjerstvo značajki još i nazvano otkrivanje značajki proces je korištenja prethodnog znanja za izdvajanje značajki iz originalnih podataka. Ovaj proces provodi se u svrhu dodavanja značajki i poboljšavanja kvalitete algoritama umjetne inteligencije. Tipični načini na koje se uvodi inženjerstvo značajki su:

- Numeričke transformacije (razlomci, množenja i skaliranje podataka)
- Kodirane kategorije kao što su *one-hot* ili ciljani koder
- Grupiranje značajki
- Grupiranje agregirane vrijednosti
- Analiza glavnih komponenti PCA (engl. *Principal component analysis*)

U ovom radu za izdvajanje značajki koristit će se numeričke transformacije i kodirane kategorije, a način na koji su provedene bit će prikazano u kasnijem poglavlju.

2.4.5 Mjere ocjene kvalitete modela

Kod procjene kvalitete modela korištene su dvije statističke veličine, a to su: koeficijent determinacije (engl., *Coefficient of determination*, R^2) i korijen srednje kvadratne pogreške (engl. *Root-Mean-Squared Error*, *RMSE*). Za svaki razvijeni model, modelom predviđene vrijednosti uspoređivale su se sa stvarnim vrijednostima iz testnog seta podataka.

Koeficijent determinacije, R^2 :

Za određivanje koeficijenta determinacije koristi se formula (1) ²⁷:

$$R^2 = \frac{a \sum_{i=1}^n y_i + b \sum_{i=1}^n x_i y_i - n \times \bar{y}^2}{\sum_{i=1}^n y_i^2 - n \times \bar{y}^2} \quad (1)$$

U jednadžbi (1) X predstavlja nezavisnu varijablu, Y je zavisna varijabla, a i b su parametri funkcije koja se aproksimira. Stvarne vrijednosti zavisne varijable prikazane su kao y , a modelom procijenjene vrijednosti prikazane su kao \bar{y} . Model je reprezentativniji što je koeficijent determinacije bliži jedinici. Tablica 3. prikazuje Chadockovu ljestvicu i značenja vrijednosti koeficijenta determinacije. U ovom radu korištena je funkcija *R2_score* koja je dio Sklearn pythonove knjižnice ²⁷.

Tablica 3. Chadockova ljestvica

Koeficijent determinacije, R^2	Značenje
0	odsutna veza
0.00 - 0.25	slaba veza
0.25 - 0.64	veza srednje jakosti
0.64 - 1.00	čvrsta veza
1	potpuna veza

Korijen srednje kvadratne pogreške, *RMSE*:

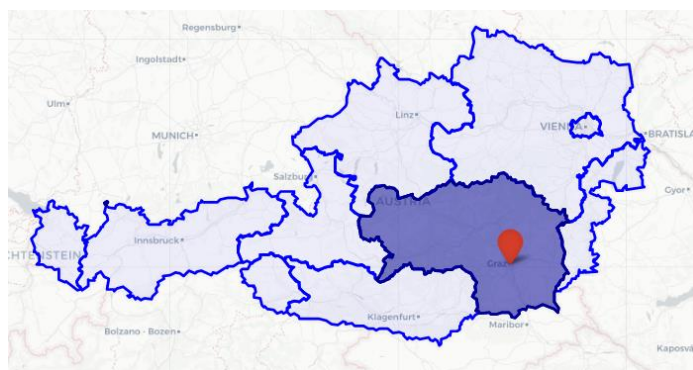
Uz koeficijent determinacije u ovom radu koristio se i korijen srednje kvadratne pogreške prikazan u formuli (2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x}_i)^2}{N}} \quad (2)$$

gdje N predstavlja broj podataka za koje se izračunava pogreška, x_i su stvarne vrijednosti koncentracije onečišćenja zraka, a \bar{x}_i su modelom procijenjene vrijednosti ²⁸.

2.5 GRAZ

Zbog javno dostupnih mnogobrojnih meteoroloških podataka za ovaj *case-study* uzet je grad Graz. Graz je glavni grad savezne države Štajerske (eng. Styria) te je drugi grad po veličini u Austriji. Na zadnjem popisu stanovništva koji je održan 01.01.2021., u Grazu je živjelo 291 134 stanovnika, a gustoća naseljenosti na istom tom popisu iznosila je 2282 stanovnika po km² ²⁹. Graz leži na rijeci Muri koja protječe kroz cijeli grad, a sa sjevera, istoka i zapada okružen je planinama i brdima. Grad se nalazi na 362m nadmorske visine. Klima je umjereno kontinentalna s toplijim ljetnim mjesecima i hladnijim razdobljima zimi. Zbog planinskih lanaca koji se protežu uz zapadnu stranu grada, broj padalina zimi je smanjen, dok su s druge strane u ljetnim mjesecima česta grmljavinska nevremena. Zimi u prosjeku padne 62cm snijega, a temperature padaju i do -15 °C, dok u ljetnim mjesecima temperature idu i iznad 33 °C. Prosječna temperatura u najhladnijem mjesecu (Siječanj) iznosi -0.8 °C, a prosječna temperatura u najtoplijem mjesecu (srpanj) je 20.7 °C. Godišnja količina padalina iznosi 845 milimetara po kvadratnom metru dok je broj sunčanih sati u godini 1890 ³⁰.



Slika 5. Prikaz Graza i savezne države u Austriji

Trenutno se kroz grad proteže 1100 km prometnica na kojima prometuje preko 140 000 vozila od čega je čak 117 000 osobnih automobila. Svakoga dana kroz grad putuje preko 136 000 ljudi od čega ih čak 35,7% putuje osobnim automobilima, pri čemu omjer ljudi po automobilu iznosi 1,22. Ostatak putnika svakodnevno putuje kao putnici u automobilima (9,5 %) i javnom prijevozu (19,9%) te biciklom (16,1 %) i pješice (18,8%). ³¹

2.6 KVALITETA ZRAKA U GRAZU

Početak 2021. godine Graz je imao period sa zadovoljavajućom kvalitetom zraka što je moguća posljedica COVID zabrane kretanja koja je bila na snazi u 2020. godini. Kvaliteta zraka je određena prema američkom indeksu kvalitete zraka (eng. *U. S. Air Quality Index*). Razine štetnih tvari u zraku bile su: PM2.5 9,5 $\mu\text{g}/\text{m}^3$, PM10 11,5 $\mu\text{g}/\text{m}^3$, dušikov dioksid 45 $\mu\text{g}/\text{m}^3$ i sumporov dioksid 15 $\mu\text{g}/\text{m}^3$ ³². Sa ovakvim koncentracijama tvari u zraku ne postoji opasnost od štetnih tvari. Grad prašine ili prijestolnica fine prašine neki su od naziva za Graz zbog loše kvalitete zraka u gradu. Buka koju uzrokuje promet, zagušljivi zrak i fina prašina samo su neki od problema ovoga grada. Onečišćenje grada je visoko u četvrtima Gries, Lend i Eggenberg. Postaje koje mjere onečišćenje u gradu su dugi niz godina na vrhu austrijske ljestvice onečišćenja zraka, a one koje pokazuju visoka godišnja onečišćenja su Don Bosco, Tiergartenweg i Petersgasse. Na mjernoj postaji Don Bosco granica fine prašine od 50 $\mu\text{g}/\text{m}^3$ u 2017. godini prijeđena je čak 54 puta, u 2018. taj broj bio je 39 dana, a u 2019. 15 puta. Ovakvo izrazito smanjenje broja dana kada koncentracije onečišćenja prelaze dozvoljenu granicu rezultat je vremenskih uvjeta. Postoji nekoliko razloga loše kvalitete zraka u Grazu. S jedne strane su to promet, grijanje i industrija, a s druge strane sam geografski položaj Graza. Graz se nalazi u kotlini što znači da je protok zraka loš, što je jako izraženo u zimskim mjesecima kada su gornji slojevi zraka topliji od onih blizu tla i nema prirodne izmjene zraka. Nedostatkom prirodnog kretanja zraka nema načina da se čestice prašine odvede s tla dalje od grada, a ovaj fenomen još se naziva i temperaturna inverzija zraka³². Načini na koje se pokušava smanjiti zagađenje u gradu su postavljanje mjernih postaja u Griesplatzu gdje se mjeri utjecaj prometa na kvalitetu zraka i uspoređuju se podaci s podacima dobivenim na postaji Don Bosco. Međutim, ti podaci trenutno nisu od neke koristi jer je uslijed COVID krize kvaliteta zraka porasla. Uz mjernu postaju u Grazu postavljen je i novi koncept mobilnosti kojim se potiče građane da koriste bicikle i javni prijevoz za svakodnevne migracije, te da više pješake³².

2.7 PYTHON I NJEGOVE KNJIŽNICE

Python je programski jezik kojeg je 1980. godine dizajnirao Guido van Rossum, a nastao je kao nadogradnja na ABC programski jezik. Python svoju primjenu nalazi u mnogim područjima, a neka od njih su: razvoj softvera, izrada *web* stranica i u novije vrijeme u području znanosti u podacima, a koristi se za zaštitu podataka u industriji. Python za svoj rad

koristi i veliku količinu knjižnica koje se koriste za: automatizaciju, analizu podataka, obradu slika, strojno učenje, razvijanje mobilnih aplikacija itd. U siječnju 2022. godine broj knjižnica koje se koriste u programskom jeziku Python premašio je brojku od 350000³³. U ovom radu od Pythonovih knjižnica koristile su se: *pandas*, *sklearn*, *geopandas*, *folium*, *matplotlib*. Knjižnice *folium* i *geopandas* koriste se za prikaz geografskih podataka dok se *matplotlib* koristi za grafički prikaz rezultata i statističku obradu podataka.

Pandas

Pandas je softver knjižnica napisana za programski jezik Python, a koristi se za manipulaciju i analizu podataka. Pandas nudi organizaciju struktura podataka i numeričke operacije na tablicama i vremenskim serijama.

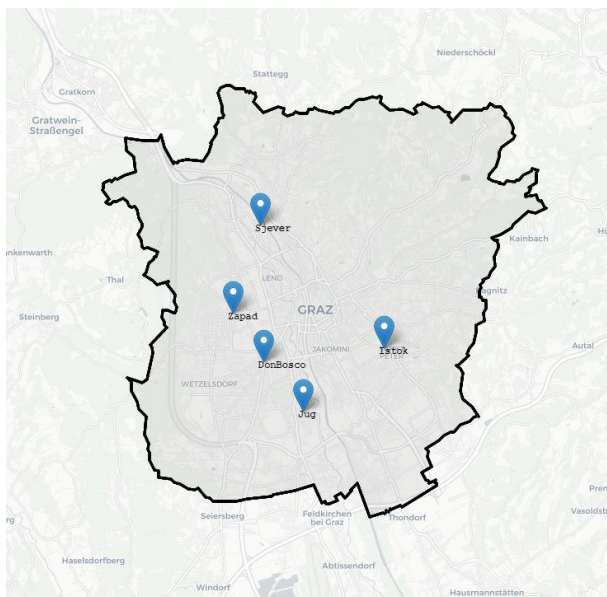
Scikit-learn

Scikit-learn ili skraćeno sklearn je softverska knjižnica razvijena za potrebe umjetne inteligencije, a napisana za programski jezik Python. U ovoj knjižnici moguće je pronaći različite algoritme za klasifikaciju i regresiju. Ova knjižnica povezana je s pandasom i moguće ih je koristiti simultano, a neke od funkcija koje se koriste u ovom radu, a dio su sklearn su: *MLPRegression* i *RandomForestRegression*.

3 MATERIJALI I METODE

3.1 PRIKUPLJANJE PODATAKA

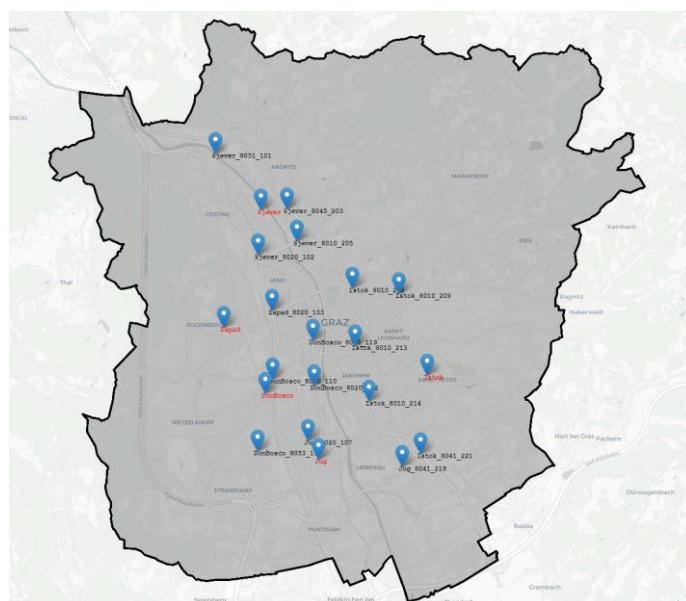
Meteorološki podaci koji su se koristili za razvoj modela mjereni su na pet lokacija u gradu.³⁴ U radu⁴ dokazano je da vrijeme ima utjecaj na koncentraciju lebdećih čestica u zraku. Mjerne postaje u gradu su prikazane na slici 6. Crni obrub predstavlja grad Graz.



Slika 6. Mjerne postaje u gradu

U radu Šimić *et al.*³⁵ dokazano je da promet utječe na koncentraciju lebdećih čestica, a samim time i na kvalitetu zraka. Zbog toga će se u ovom radu koristiti i podaci za promet dobiveni sa šesnaest lokacija u gradu. Na slici 7. prikazane su prometnice na kojima se mjeri promet i njihov odnos prema mjernim postajama u gradu. Crni obrub kao i u slici 6. predstavlja grad Graz. Slike 6. i 7. izrađene su pomoću pythonove knjižnice *Folium* za što je bilo potrebno poznavati geografsku duljinu i širinu prometnica i mjernih postaja. Kako bi se odredila udaljenost između prometnica i mjernih postaja korištene su pythonove knjižnice *folium* i *openrouteservice*. *Openrouteservice* se koristi za izračunavanje najkraćeg puta između prometnica i mjernih postaja dok je *Folium* korišten za određivanje zračne udaljenosti između točaka. U tablici 4. dane su vrijednosti zračnih i cestovnih udaljenosti između mjernih postaja i prometnica.

Svakoj od prometnica je dodano ime najbliže joj postaje radi lakšeg snalaženja u podacima.



Slika 7. Prometnice na kojima se mjeri promet (crno) i mjerne postaje u gradu (crveno)

Tablica 4. Prometnice i mjerne postaje te njihove udaljenosti

Prometnice	Najbliza mjerna postaja	Zračna udaljenost do postaje, km	Cestovna udaljenost do postaje, km
Istok_8010_208	Ost	2.69	3.01
DonBosco_8020_110	DonBosco	0.38	0.42
DonBosco_8020_119	DonBosco	1.68	1.96
DonBosco_8020_122	DonBosco	1.17	1.67
DonBosco_8053_111	DonBosco	1.36	1.43
Sjever_8010_205	Nord	1.12	2.91
Istok_8010_209	Ost	2.02	2.87
Istok_8010_213	Ost	1.83	2.68
Istok_8010_214	Ost	1.50	2.59
Sjever_8020_102	Nord	1.07	2.01
Zapad_8020_103	West	1.20	1.53
Jug_8020_107	Sud	0.51	0.84
Jug_8041_218	Sud	1.98	3.06
Istok_8041_221	Ost	1.86	3.07
Sjever_8045_203	Nord	0.61	2.81
Sjever_8051_101	Nord	1.70	2.32

Za razvoj modela pored podataka vezanih za promet korišteni su i meteorološki podaci koji su mjereni na pet postaja. Za usporedbu s meteorološkim podacima korišteni su i satelitski meteorološki podaci preuzeti s *Climate Data Store* uz pomoć pythonove knjižnice *cdsapi*.³⁴ U razvoj modela uključeni su i temporalni podaci kao što su dani u tjednu, godišnja doba, praznici, mjeseci, itd. U radu³⁵ dokazano je da dani u tjednu pridonose razvoju modela i procjeni lebdećih čestica u zraku.

3.2 METEOROLOŠKI PODACI

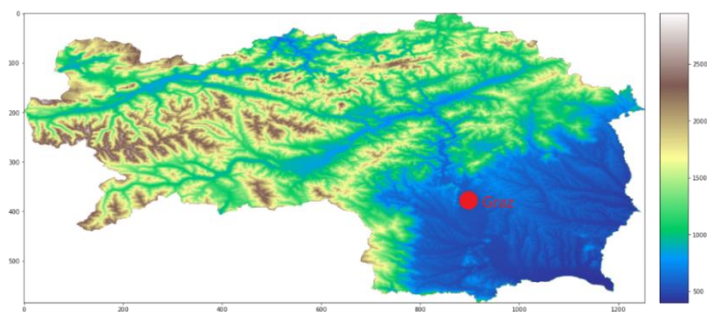
Meteorološki podaci dobiveni na pet lokacija u gradu prikazani su u tablici 5. Podaci su mjereni za razdoblje od 01.01.2018. do 15.03.2020.

Tablica 5. Meteorološki podaci mjereni na postajama

Postaje	Meteorološki podaci
Sjever	Smjer i brzina vjetra, temperatura, relativna vlažnost zraka, tlak, oborine i radijacija
Jug	Smjer i brzina vjetra, temperatura, relativna vlažnost zraka, tlak
Istok	Smjer i brzina vjetra, temperatura, relativna vlažnost zraka
Zapad	Smjer i brzina vjetra, temperatura, relativna vlažnost zraka
Don Bosco	Temperatura, relativna vlažnost zraka

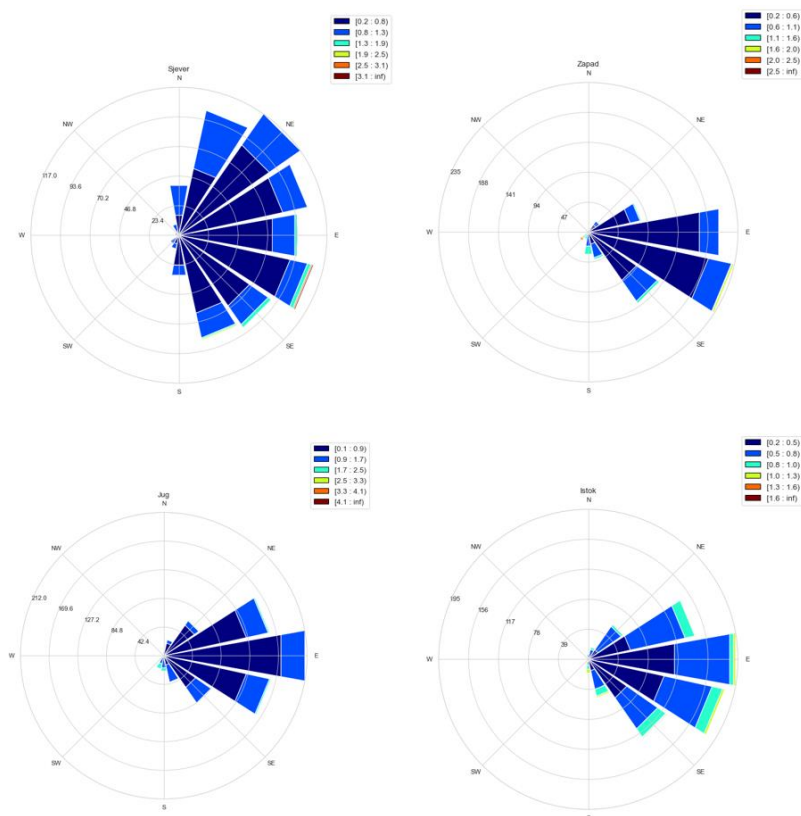
3.2.1 Vjetar

Vjetar predstavlja vodoravno strujanje zraka određeno smjerom odnosno stranom svijeta od koje puše vjetar i brzinom. Vjetar je zbog svoje velike prostorne i vremenske promjenjivosti vektorska veličina. Do nastanka vjetra dolazi iz više razloga, a neki od njih su: razlika tlaka između dva područja (vjetar puše iz područja višeg tlaka u područje manjeg tlaka i što je razlika veća to je i vjetar jači), Zemljina rotacija, Coriolisova sila (vjetar zbog djelovanja sile na sjevernoj polutki skreće desno, a na južnoj lijevo), centrifugalne sile kad je putanja čestica zraka zakrivljena i sila trenja s pologom³⁶. Kada se promatra vjetar u Grazu veliki utjecaj ima topografija, odnosno reljefni prostor oko Graza. Na slici 8. prikazana je topografija savezne države Štajerske i grad Graz.



Slika 8. Topografija Štajerske i lokacija Graza (crveno)

Iz topografije savezne države vidljivo je kako se Graz nalazi u kotlini, odnosno sa zapadne strane grada nalazi se planinski lanac koji štiti grad od vjetrova sa zapada i sjevera. Međutim, s istočne i južne strane nalazi se nizina te je grad izložen vjetrovima baš iz tih smjerova. Na postaju sjever dolazi i vjetar iz smjera sjeveroistoka i to zbog prolaza koji se nalazi između planinskih lanaca. Na slici 9. je prikazana ruža vjetrova za četiri mjerne postaje u gradu. Boje na slikama označuju jačinu vjetra, a smjer vjetra je prikazan pomoću stupaca.

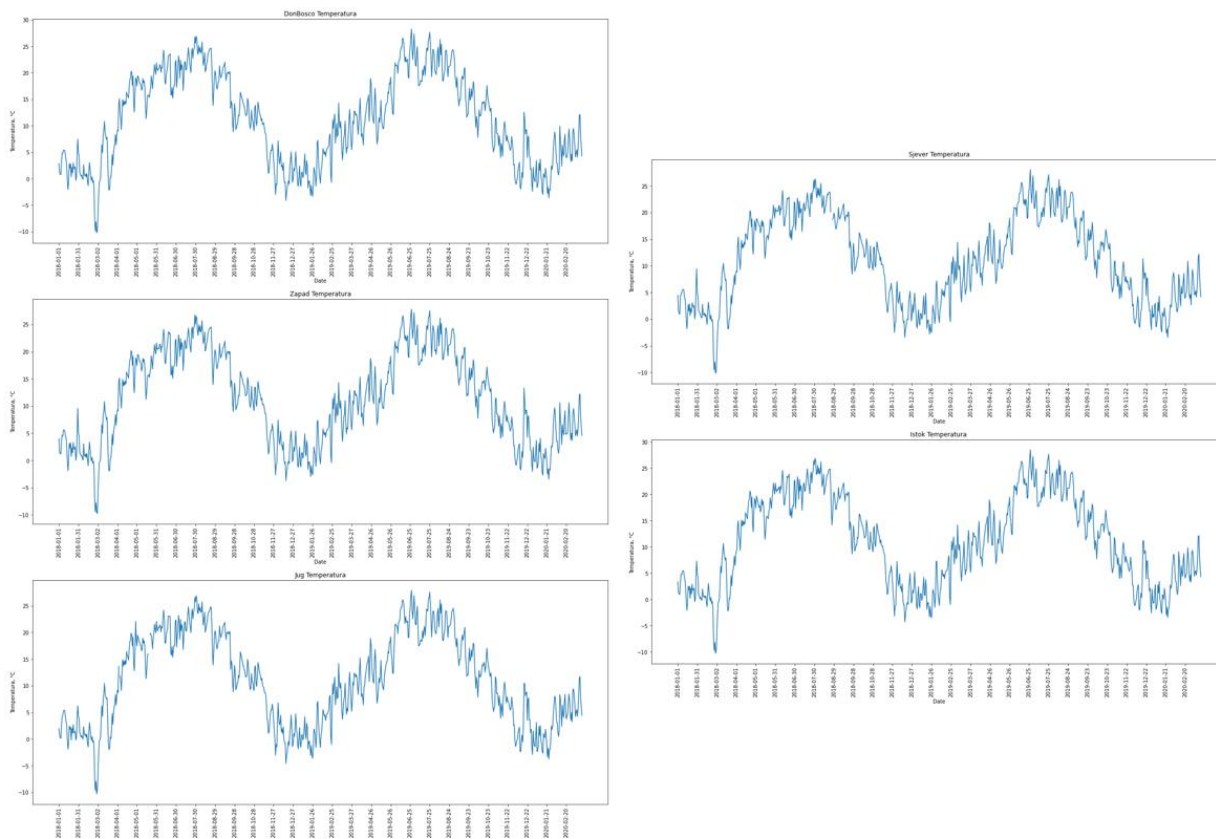


Slika 9. Smjer i brzina vjetra na mjernim postajama (Sjever – gore lijevo, Zapad – gore desno, Jug – dolje lijevo, Istok – dolje desno)

Na slici 9. prikazani su smjer i brzine puhanja vjetra na četiri mjerne postaje. Kada se promatra postaja Sjever može se vidjeti kako većina vjetra dolazi sa sjeveroistoka i jugoistoka što i odgovara topografiji i lokaciji mjerne postaje u gradu. Ova postaja ujedno bilježi i najviše promjena u smjeru puhanja vjetra. Najjači ili najbrži vjetar dolazi sa sjeveroistoka. Postaja zapad bilježi najviše vjetra sa smjera istoka i jugoistoka. Smjer vjetra koji se mjeri na postajama Jug i Istok je istog smjera što također odgovara topografiji i nizini koja se nalazi istočno i južno od Graza. Najveće brzine puhanja vjetra zabilježene su u proljetnim mjesecima na svim postajama. Što se tiče najveće brzine puhanja vjetra za promatrano razdoblje, ona je zabilježena na postaji Jug i iznosi 4.1 m/s.

3.2.2 Temperatura zraka

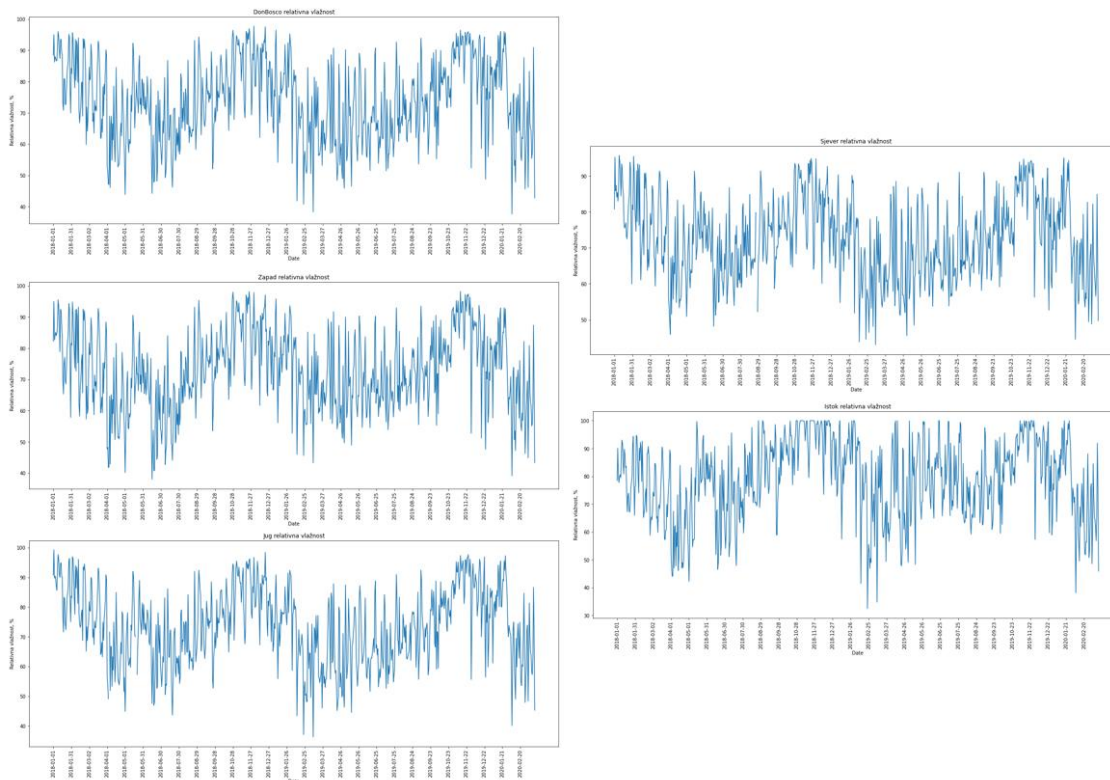
Temperatura zraka predstavlja temperaturu mjerenu na 2 m visine. Na temperaturu zraka ne utječe toplinsko zračenje tla i okoline ili izravno zračenje Sunca. Temperatura zraka ovisi o dobu dana i godini, te tako tijekom dana temperatura ovisi o količini naoblake i vjetru dok kroz godinu ona ovisi o položaju Zemlje u odnosu na Sunce. U Grazu je najhladniji mjesec siječanj, a najtopliji srpanj³⁷. Za razliku od vjetra, temperatura se mjeri na svim postajama u gradu te je na slici 10. prikazana temperatura za svaki dan u godini na svakoj od postaja. Slika 10. potvrđuje tezu da je siječanj najhladniji mjesec godine, a srpanj najtopliji. Sa slike 10. je vidljivo i blago povećanje u temperaturi u 2019. godini u odnosu na 2018. Temperature se ne mijenjaju znatno između postaja zbog toga što su relativno blizu jedna drugoj. U ljetnim mjesecima uz visoke temperature dolazi do smanjenja koncentracije lebdećih čestica u zraku što će biti potvrđeno u jednom od slijedećih poglavlja. Zimske mjesece i niske temperature prati povećana koncentracija lebdećih čestica.



Slika 10. Temperature zraka u danu za razdoblje od 01.01.2018. do 15.03.2020. (Don Bosco – prvi s lijeva, Zapad – drugi s lijeva, Jug – treći s lijeva, Sjever – prvi s desna, Istok – drugi s desna)

3.2.3 Relativna vlažnost zraka

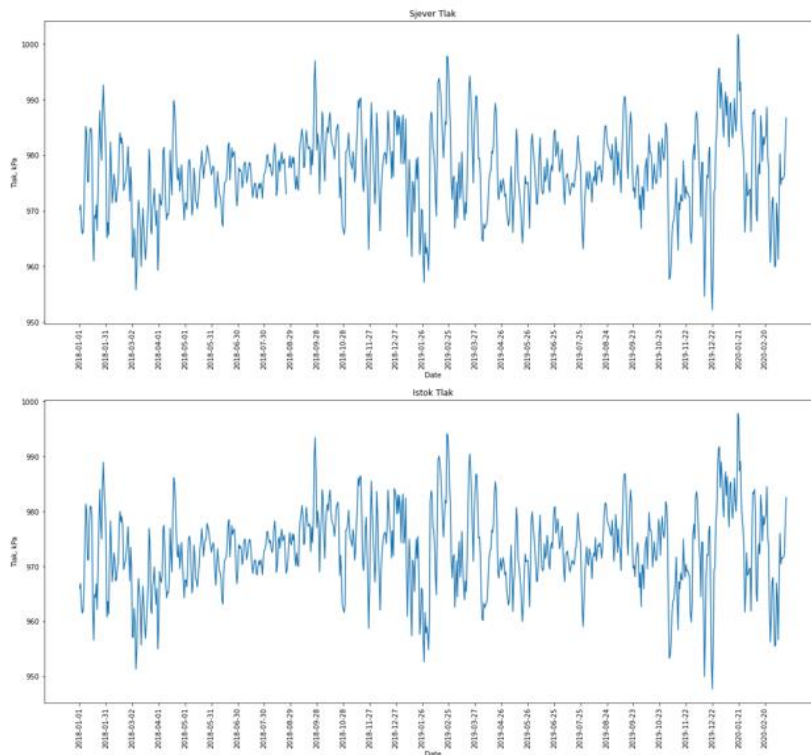
Relativna vlažnost zraka je udio vodene pare u zraku, a izražava se kao postotak. Dobiva se iz omjera parcijalnog tlaka vodene pare i parcijalnog tlaka zasićene vodene pare pri danom tlaku zraka i temperaturi. Kada je njezina vrijednost 0 % radi se o potpuno suhom zraku, a kada je vrijednost 100 % radi se o potpuno zasićenom zraku³⁸. Na slici 11. prikazana je relativna vlažnost zraka izmjerena na pet postaja u gradu. Kao i kod temperature, premala je udaljenost između postaja da bi se izmjerene veličine znatno razlikovale. Ipak postoji mala razlika gdje je na postaji Jug kroz cijelo razdoblje najviša relativna vlažnost zraka. Što je veća relativna vlažnost zraka to će biti veća i koncentracija lebdećih čestica. Ovo će biti potvrđeno u poglavlju Lebdeće čestice.



Slika 11. *Relativna vlažnost zraka u danu za razdoblje od 01.01.2018. do 15.03.2020. (Don Bosco – prvi s lijeva, Zapad – drugi s lijeva, Jug – treći s lijeva, Sjever – prvi s desna, Istok – drugi s desna)*

3.2.4 Tlak zraka

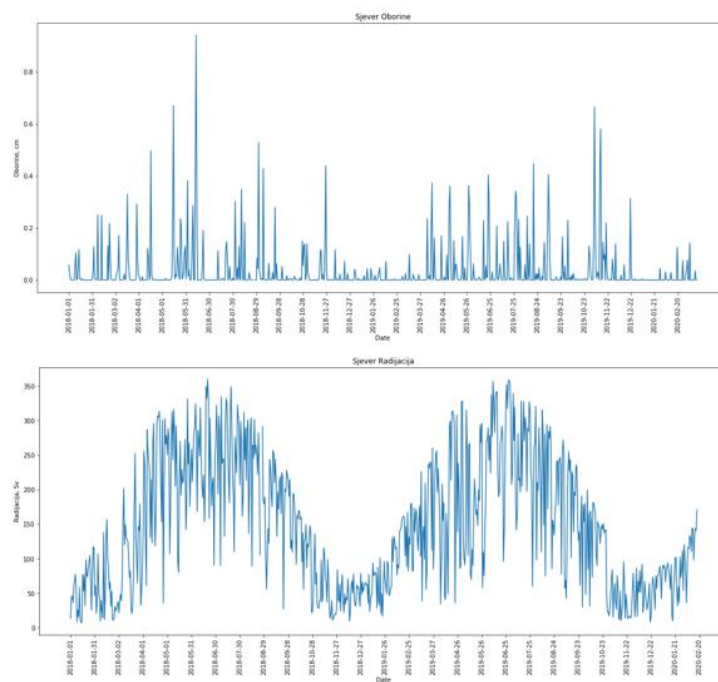
Tlak zraka skalarna je veličina koja opisuje djelovanje sile na površinu i dobiva se iz omjera sile i površine. Mjerna jedinica za tlak je paskal, ali se u meteorologiji češće koristi bar (10^5 Pa). Uz to još se može koristiti $1 \text{ atm} = 101325 \text{ Pa}$ ili mmHg, 1 mmHg iznosi $133,322 \text{ Pa}$ ³⁹. Atmosferski tlak koji se mjeri na dvije meteorološke postaje Sjever i Istok izražava se kao težina zraka nad površinom. Na slici 12. prikazan je tlak zraka za promatrano razdoblje. Vrijednosti na Sjevernoj mjernoj postaji veće su u odnosu na vrijednosti mjerene na Istoku. Kod velikih promjena tlaka koncentracija lebdećih čestica bit će veća i u odnosu na razdoblja s malim promjenama tlaka kada je koncentracija lebdećih čestica niska.



Slika 12. Tlak zraka izmjeren u Graz u razdoblju od 01.01.2018. do 15.03.2020. (Sjever – gore, Istok – dolje)

3.2.5 Oborine i radijacija

Pod pojam oborine ili padaline spada sva voda u kapljevitom ili čvrstom agregatnom stanju koja pada iz oblaka ili nastaje kondenzacijom iz zraka u obliku rose. Oborine se dijele na one nastale iz oblaka, dizanjem vlažnog zraka uz planine pod utjecajem vjetera ili polaganim podizanjem pod utjecajem ciklona. Mjeri se u mm po kvadratnom metru pomoću kišomjera⁴⁰. Radijacija ili zračenje je prijenos energije u obliku elektromagnetskih valova ili u snopovima subatomske čestice koje se gibaju kroz prostor. Zračenje se mjeri u Sievert (Sv)⁴¹. Na slici 13. prikazane su vrijednosti oborina i radijacije izmjerene na postaji Sjever. Velike količine padalina prateće mala koncentracija lebećih čestica u zraku jer oborinama dolazi do čišćenja atmosfere. Radijacija će biti veća u ljetnim mjesecima kada je koncentracija čestica u zraku malena.



Slika 13. Prikaz oborina i radijacije za promatrano razdoblje (Oborine – gore, radijacija – dolje)

3.2.6 Satelitski podaci

U ovom radu ispitat će se i usporediti utjecaj meteoroloških podataka dobivenih na lokalnim postajama i meteoroloških podataka dobivenih sa satelita, na učinkovitost modela. Satelitski podaci prikupljeni su sa stranice *Climate Data Store* pomoću pythonovih knjižnica *cdsapi*, *xarray*, *pathlib* i *os*. *Cdsapi* knjižnica je Europskog centra za srednjoročne vremenske prognoze (eng. *European Centre for Medium-Range Weather Forecasts* (ECMWF)) i koristi se za prikupljanje podataka o vremenu za bilo koje mjesto na Zemlji. Kako bi se dobili podaci za grad Graz koristile su se geografska dužina i širina koje iznose: 47.0707 i 15.4395⁴². Meteorološki podaci koji su preuzeti su: temperatura, naoblaka, tlak pare, vjetar i rosište. Svi podaci uzeti su kao srednje vrijednosti za razdoblje od 24 sata. Podaci za temperaturu i vjetar prate trend podataka koji su dobiveni na lokalnim postajama dok se tlak pare, rosište i naoblaka ne mjere na lokalnim postajama. Naoblaka predstavlja pokrivenost neba oblacima bez obzira na vrstu i visinu istih. Određuje se brojčano od 0 do 10, gdje 0 predstavlja potpuno vedro, a 10 nebo zastrto oblacima⁴³. U ovom radu podaci će biti prikazani postotkom gdje je 0 vedro, a 100 zastrto oblacima.

Naoblaka danju sprječava da izravno Sunčevo zračenje dopre do tla, a noću sprječava da tlo gubi toplinu. Rosište je temperatura pri kojoj će zrak postati zasićen vodenom parom, a da pri tom tlak i količina vodene pare ostaju nepromijenjeni. Rosište je dobar pokazatelj udjela vlage u zraku, naime promjene tlaka uz Zemljinu površinu su male pa tako visoke vrijednosti rosište ukazuju na visok udio vlage u zraku⁴⁴. Rosište se prema koncentraciji lebdećih čestica ponaša kao i temperatura zraka odnosno kada su vrijednosti rosišta niske koncentracija lebdećih čestica u zraku biti će visoka. Tlak pare ili parcijalni tlak je tlak koji bi plin imao da se sam nalazio u spremniku, a pri tome se plin nalazi u smjesi. Povišenje tlaka pare pratiti će toplija razdoblja pa će koncentracija lebdećih čestica biti obrnuto proporcionalna tj. kada je tlak pare visok koncentracija lebdećih čestica biti će niska i obratno.

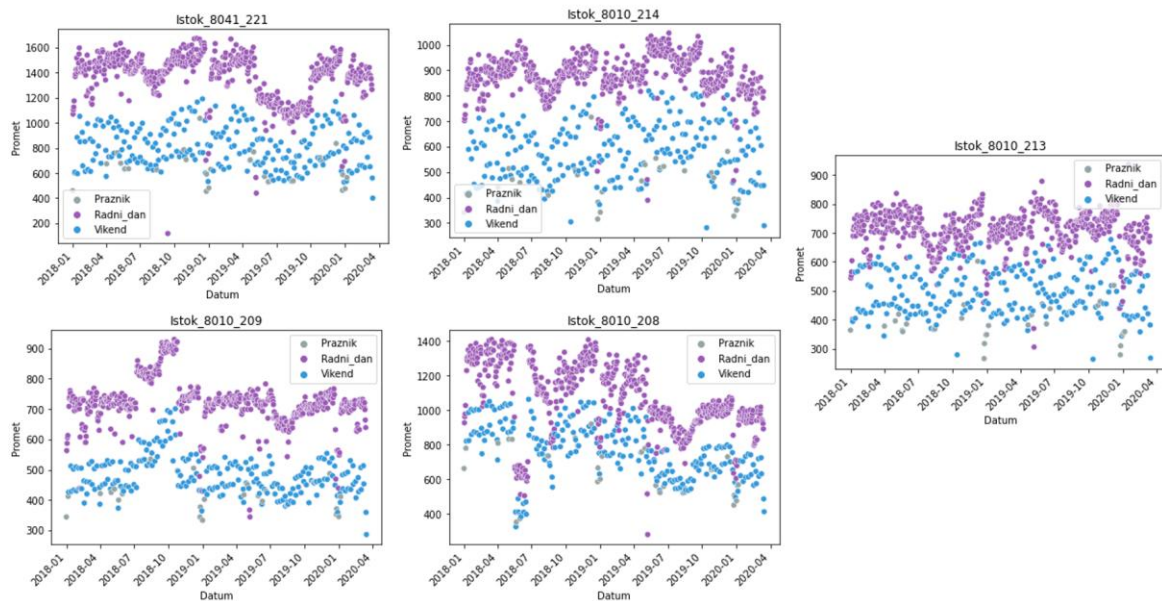
3.3 PROMET

Jedna od hipoteza koja se pokušava potvrditi ovim radom je da koncentracija lebdećih čestica u zraku ovisi o količini prometa. Iz tog razloga će u model za procjenu čestica u zraku biti uključeni i podaci za promet. Svaka od prometnica povezana je sa mjernom postajom na temelju zračne i cestovne udaljenosti što je i pokazano u prethodnom poglavlju.

3.3.1 Istok

U blizini postaje Istok nalazi se pet prometnica: 8041_221, 8010_214, 8010_209, 8010_208 i 8010_213. Najbliža prometnica postaji Istok je 8010_213 i očekuje se da će ona najviše pridonijeti onečišćenju zraka dok je prometnica 8010_208 najudaljenija i njezin utjecaj bi trebao biti najmanji (podaci o udaljenosti tablica 4.). Slika 14. prikazuje promet na svim prometnicama u blizini postaje Istok, a podaci su podijeljeni prema: radni dan, vikend i praznik. Kao što je i za očekivati, vikendom i praznicima će promet biti manji nego tijekom tjedna i samim time bi koncentracija lebdećih čestica trebala biti manja. Iz grafova se može vidjeti da je promet najgušći na prometnici Istok_8041_221., a najrjeđi na prometnicama Istok_8010_209 i Istok_8010_213. Na prometnici Istok_8010_208 došlo je do smanjenja prometa u 5. i 6. mjesecu 2018. godine zbog radova na cesti, no smanjenje u 5. mjesecu 2019. godine nastavilo je kroz cijelu 2019. i u razdoblju u 2020. godini. Prometnica Istok_8010_209 bilježi rast prometa u 2018. godini u razdoblju od 7. do 10. mjeseca, ali nakon toga u slijedećem razdoblju se promet vraća na stare vrijednosti.

Na prometnici Istok_8010_213 zabilježen je blagi pad u 2020. godini koji se također pripisuje radovima na cesti. Na svim prometnicama u 7. mjesecu svake godine može se vidjeti smanjenje u količini prometa zbog godišnjih odmora koji su česti u to doba. Na prometnici Istok_8041_221 u 2019. promet je smanjen u ljetnim mjesecima zbog radova i godišnjih odmora. Na kraju, prometnica Istok_8010_214 ne bilježi znatne promjene u količini prometa osim onih zbog godišnjih odmora.

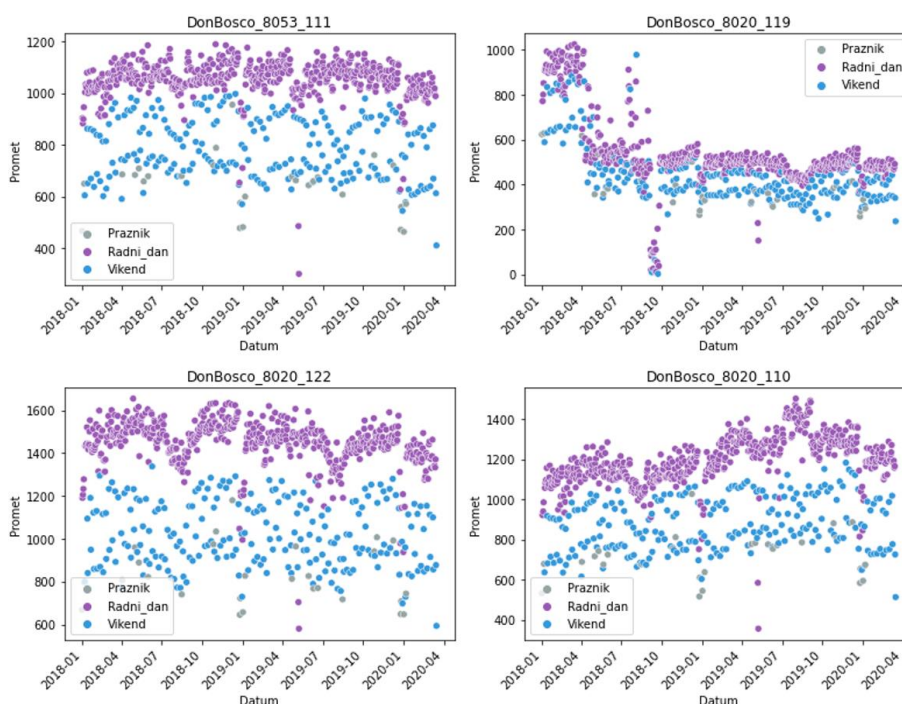


Slika 14. Promet na prometnicama u blizini postaje Istok (Istok_8041_221 – lijevo gore, Istok_8010_213 – sredina gore, Istok_8010_213 – desno sredina, Istok_8010_209 – lijevo dolje, Istok_8010_208 – sredina dolje)

3.3.2 Don Bosco

U blizini mjerne postaje Don Bosco nalaze se četiri prometnice i to su redom: 8053_111, 8020_119, 8020_122 i 8020_110. Prometnica 8020_110 najbliža je mjernej postaji, a 8020_122 najudaljenija (podaci o udaljenosti tablica 4.). Sa slike 15. vidljivo je kako je promet na tri od četiri prometnice iznimno gust, četvrta prometnica koja odstupa od prosjeka je DonBosco_8020_119 gdje je u 2018. godini zabilježen pad prometa koji se nastavio i u preostalom promatranom razdoblju. Prometnica DonBosco_8020_119 bilježi znatan pad prometa od oko 25 %, pad prometa javlja se u 5. mjesecu 2018. godine.

Kod ostalih prometnica nema znatnih promjena u količini prometa i samo se na prometnici DonBosco_8020_122 promet smanjuje u ljetnim mjesecima, dok se na ostale dvije prometnice promet znatnije ne mijenja kroz godinu.

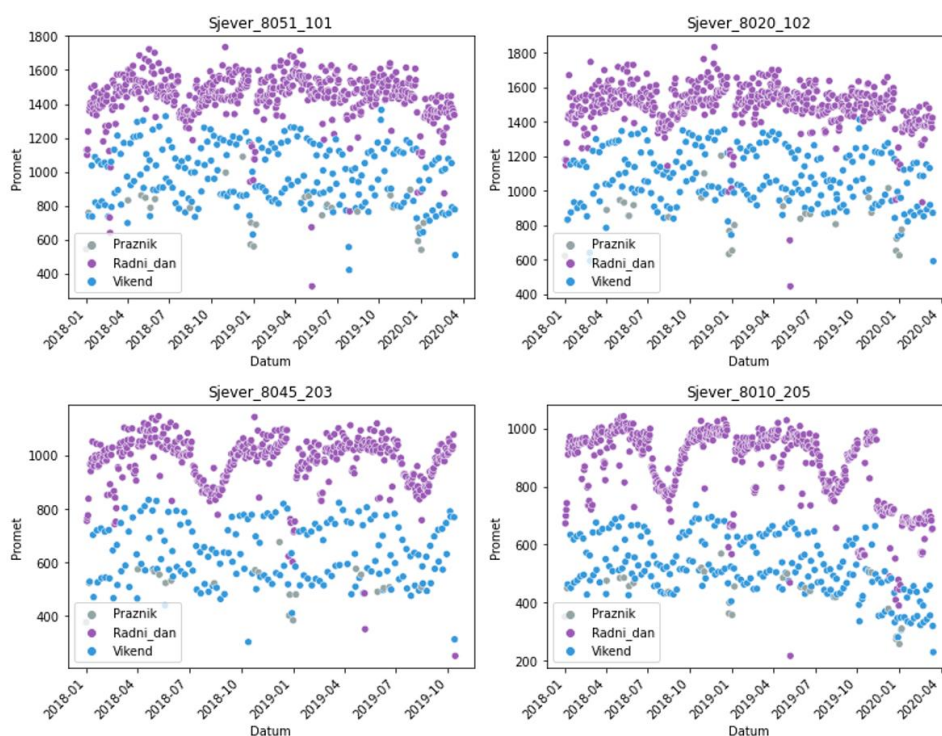


Slika 15. *Promet na prometnicama u blizini postaje Don Bosco (DonBosco_8053_111 – lijevo gore, DonBosco_8020_119 – desno gore, DonBosco_8020_122 – lijevo dolje i DonBosco_8020_110 – lijevo dolje)*

3.3.3 Sjever

U blizini postaje sjever nalaze se četiri prometnice i to su redom: 8051_101, 8020_102, 8045_203 i 8010_205 (podaci o udaljenosti tablica 4.). Najbliža prometnica u odnosu na mjernu postaju Sjever je 8020_102, a najudaljenija je 8045_203. Na slici 16. prikazan je promet u blizini postaje Sjever. Promet je izrazito gust na prometnicama Sjever_8051_101 i Sjever_8020_102 i ne mijenja se značajno kroz godinu dok je na ostale dvije prometnice gustoća prometa manja i u ljetnim mjesecima se promet razrijedi tijekom tjedna dok za vikend nema značajnih promjena. Na prometnici Sjever_8020_102 promet se kontinuirano smanjuje pa je tamo promet pao s 1800 automobila u danu na 1400 automobila u razdoblju od 2018. do 2020. godine. U 2018. godini je zabilježen pad vozila u ljetnim mjesecima dok u 2019. to nije slučaj.

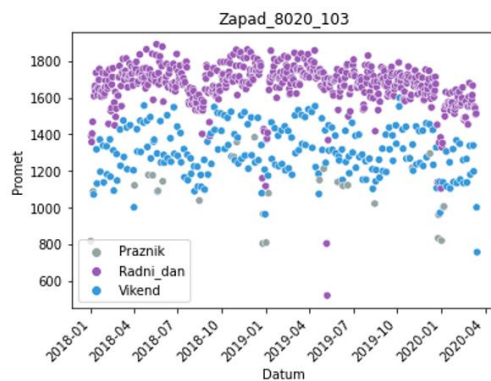
Ovo je ujedno i prometnica s najviše prometa u blizini postaje Sjever. Prometnica Sjever_8010_205 bilježi pad prometa u ljetnim mjesecima u 2018. i 2019. godini, ali je značajan i pad prometa u 11. i 12. mjesecu 2019. godine koji se nastavlja i u 2020. godini. Prometnici Sjever_8045_203 nedostaju podaci za trećinu 2019. godine i 2020. godinu zbog toga će se ti podaci procijeniti iz prijašnjih podataka. Na prometnici Sjever_8051_101 prisutna je velika količina vozila svakog dana u tjednu, u 2018. godini broj vozila se smanjuje u 7. i 8. mjesecu no u istim mjesecima u 2019. godine to smanjenje nije prisutno.



Slika 16. Promet na prometnicama u blizini postaje Sjever (Sjever_8051_101 – lijevo gore, Sjever_8020_102– desno gore, Sjever_8045_203– lijevo dolje i Sjever_8010_205– lijevo dolje)

3.3.4 Zapad

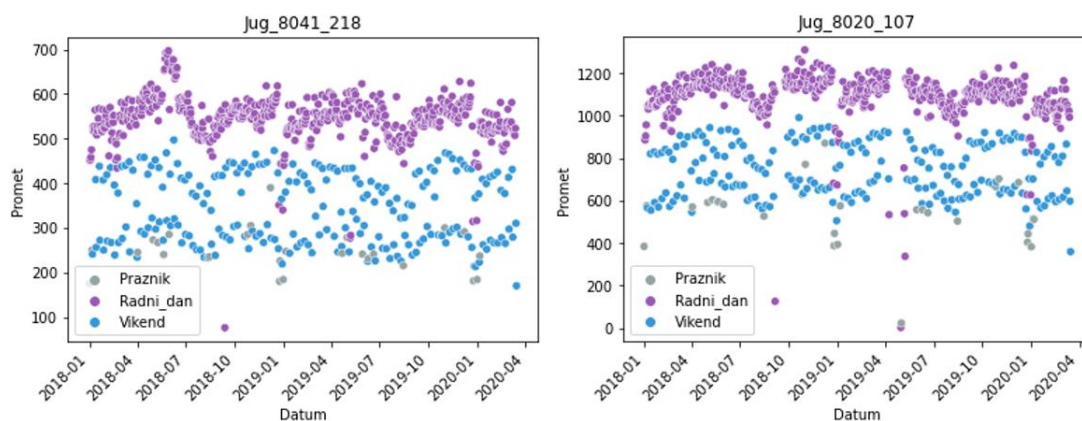
U blizini postaje zapad nalazi se samo jedna prometnica i to je 8020_103 (podaci udaljenosti u tablici 4.) . Međutim, ovo je jedna od prometnica s najgušćim prometom kroz cijelu godinu. Slika 17. prikazuje promet na prometnici Zapad_8020_103. Promet se smanjuje u ljetnim mjesecima u 2018. godini, ali u 2019. godini to smanjenje nije prisutno te je promet gust kroz cijelu godinu.



Slika 17. Promet na prometnici Zapad_8020_103

3.3.5 Jug

U blizini postaje Jug nalaze se dvije prometnice i to su: 8041_218 i 8020_107 (podaci o udaljenosti u tablici 4.). Prometnica Sud_8041_107 bliža je postaji Jug od druge te bi njezin utjecaj na postaju trebao biti izraženiji. Slika 18. prikazuje promet na prometnicama u blizini postaje Jug. Kod prometnice Jug_8020_107 za neke dane u 2018. i 2019. godini nema podataka pa će se oni procijeniti iz ostalih. Na prometnici Jug_8041_218 promet se mijenja značajno samo u ljetnim mjesecima.



Slika 18. Promet na prometnicama u blizini postaje Jug (Jug_8031_218 – lijevo i Jug_8020_107 – desno)

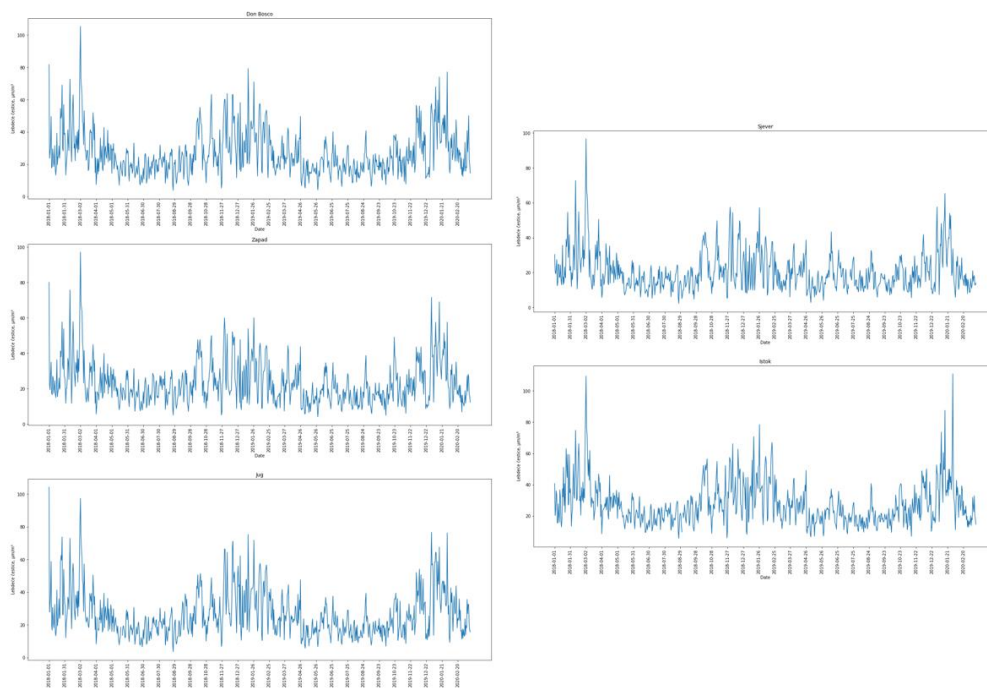
3.4 TEMPORALNI PODACI

U radu⁴ dokazano je kako temporalni podaci poput dana u tjednu, godišnjeg doba i praznika imaju utjecaj na koncentraciju lebdećih čestica u zraku te će se zbog toga oni uzeti u obzir

kod razvoja modela u ovom radu. Temporalni podaci uzeti za razvoj modela su: godina, dan u godini, mjeseci, dani u tjednu, godišnja doba, školski praznici i državni praznici.

3.5 KONCENTRACIJA LEBDEĆIH ČESTICA

Koncentracija lebdećih čestica mjeri se na svih pet mjernih postaja u gradu. Na slici 19. prikazana je koncentracija lebdećih čestica u zraku za razdoblje od 01.01.2018. do 15.03.2020. Na svim je postajama koncentracija čestica veća u zimskim mjesecima nego u ljetnim mjesecima. Na postajama Don Bosco, Istok i Jug koncentracija čestica je najveća dok je na preostale dvije ona manja. Ovakva razlika u koncentraciji je produkt prometa jer su navedene tri postaje izrazito prometno opterećene. Kada se promatra odnos između meteoroloških podataka može se uočiti da povećanjem vanjske temperature zraka dolazi do smanjenja lebdećih čestica u zraku, povećanjem relativne vlažnosti zraka doći će i do povećanja čestica u zrak. Na kraju treba spomenuti i oborine koje će pozitivno djelovati na koncentraciju lebdećih čestica u zraku odnosno smanjivati će koncentraciju čestica, a taj se proces naziva čišćenje atmosfere. Na slici 19. prikazane su koncentracije lebdećih čestica u zraku za promatrano razdoblje.



Slika 19. Koncentracija čestica u promatranom razdoblju za svih pet postaja (lijevo gore – Don Bosco, lijevo sredina – Zapad, lijevo dolje – Jug, desno gore – Sjever, desno dolje – Istok)

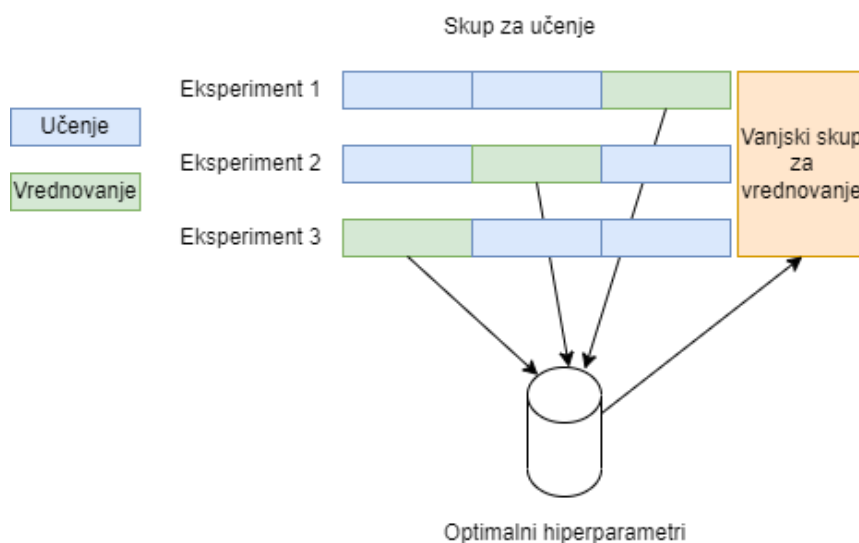
3.6 INŽENJERSTVO ZNAČAJKI

Kako bi se poboljšala kvaliteta razvijenih modela koristilo se inženjerstvo značajki i to na tri načina. Na početku su svi meteorološki podaci dobiveni s pet lokanih postaja prikazani kao podaci po satu, točnije za svaki dan i pojedinu veličinu bilo je dostupno 24 vrijednosti. Iz tih 24 vrijednosti za svaki meteorološki parametar bila je određena minimalna, maksimalna i srednja vrijednost. Drugi način poboljšanja modela koristeći inženjerstvo značajki bilo je povezivanje značajki numeričkim transformacijama (množenje i dijeljenje). Meteorološki podaci se razlikuju za svaku od postaja pa su se i podaci drugačije transformirali. Za svaku postaju množeni su temperatura i relativna vlažnost jer smanjenjem temperature dolazi do povećanja relativne vlažnosti zraka. Podaci za vjetar poznati su za sve postaje osim Don Bosca pa je povezivanje temperature i vjetra bilo moguće samo na preostale četiri postaje. Postaje Sjever i Istok mjere tlak što nije slučaj za preostale tri postaje. Tlak je povezan s temperaturom, relativnom vlažnošću i brzinom vjetra i to preko množenja i dijeljenja. Postaja Sjever je jedina postaja na kojoj se mjeri još i zračenje i oborine pa su i one uključene u razvoj modela. Zračenje je povezano s temperaturom i brzinom vjetra, a oborine s temperaturom. Treća metoda koja je korištena u ovom radu vezana je uz smjer puhanja vjetra, te je ova metoda primijenjena na sve postaje osim Don Bosco. Naime, ruža vjetrova prikazana na slici 9. podijeljena je na 8 dijelova, te svaki dio čini 45 stupnjeva. Za svaki dan određena je srednja vrijednost puhanja vjetra te se promatra u kojem se od 8 dijelova ruže vjetrova nalazi, dio u kojem vjetar puše dobiva 1 za taj dan, a ostali dijelovi dobivaju 0.

4 EKSPERIMENTALNI DIO

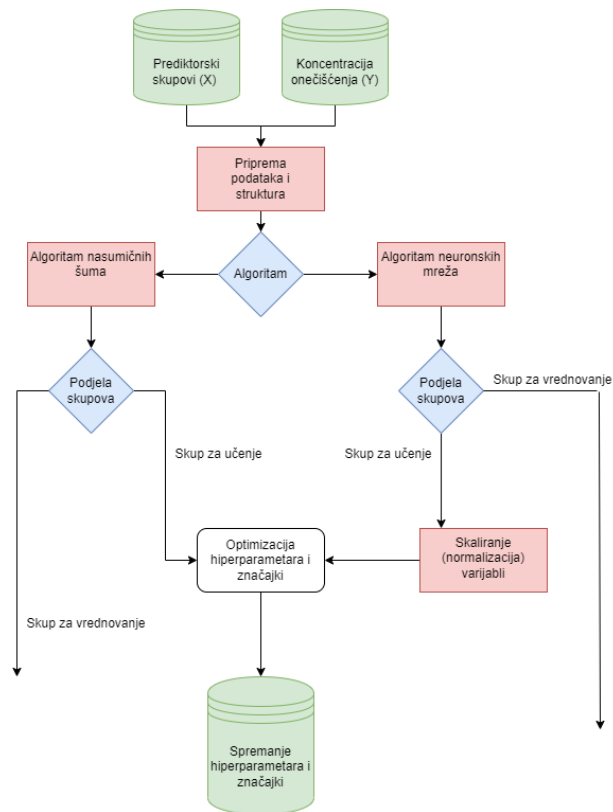
4.1 ODABIR HIPERPARAMETARA I ZNAČAJKI

Algoritmi nasumičnih šuma i algoritmi neuronskih mreža koji se primjenjuju za razvoj modela opisani su u poglavljima 2.5.1 i 2.5.2. Modeli su razvijeni za svaku od pet postaja, te su hiperparametri i značajke određene za svih pet postaja. Za optimizaciju neuronske mreže korišten je *GridSearchCV* algoritam opisan u poglavlju 2.5.3. Skup podataka podijeljen je na skup za učenje koji čini 75 % podataka, a sadrži 600 podataka, dok se skup za vrednovanje modela sastoji od 200 podataka koji pripadaju 25 % ukupnog skupa. Podaci nakon 15.03. nisu uzimani u obzir zato što od toga dana u Austriji nastupa zabrana kretanja (engl. *lockdown*). Optimizacija hiperparametara modela provedena je na skupu za učenje u postupku peterostruke križne validacije. Svaki skup za učenje podijeljen je na pet podskupova s jednakim brojem značajki, što je prikazano na slici 20.



Slika 20. *Primjer trostruke križne validacije*

Algoritam peterostruke križne validacije optimira hiperparametre modela na četiri podskupa, a kvaliteta se provjerava na preostalom petom skupu. Na slici 21. prikazan je cijeli algoritam optimizacije modela. Za svih 5 postaja korišten je ovaj algoritam optimizacije hiperparametara i značajki.



Slika 21. Algoritam odabira hiperparametara i značajki

U tablici 6. prikazani su hiperparametri koji su optimirani u spomenutom algoritmu, a svaki od parametara objašnjen je u poglavlju 2.5.3.

Tablica 6. Hiperparametri RF i NN modela i njihove moguće vrijednosti

Algoritam	Hiperparametar	Vrijednosti
Nasumičnih šuma	<i>n_estimators</i>	100, 200, 300
	<i>max_depth</i>	4, 6, 8, 10
	<i>n_jobs</i>	-1
	<i>random_state</i>	42
	<i>max_samples</i>	0.4, 0.45, 0.5
	<i>min_samples_leaf</i>	2, 5, 10
Neuronske mreže	<i>hidden_layer_sizes</i>	[100,500] [20,100] [5,10]
	<i>solver</i>	adam
	<i>learning_rate_init</i>	0.0001
	<i>max_iter</i>	300
	<i>early_stopping</i>	TRUE
	<i>random_state</i>	42
	<i>alpha</i>	0.0001, 0.05

Uz hiperparametre optimiraju se i ulazne značajke metodama opisanim u poglavlju 2.5.3. Za svaku od postaja različit je broj početnih značajki te je potrebno odrediti one najvažnije. Ulazne značajke za svaku od pet postaja dane su u tablici 7. te će se iz njih uzeti one najbolje za model. Izlazna značajka za svaku od postaja je koncentracija lebdećih čestica u zraku.

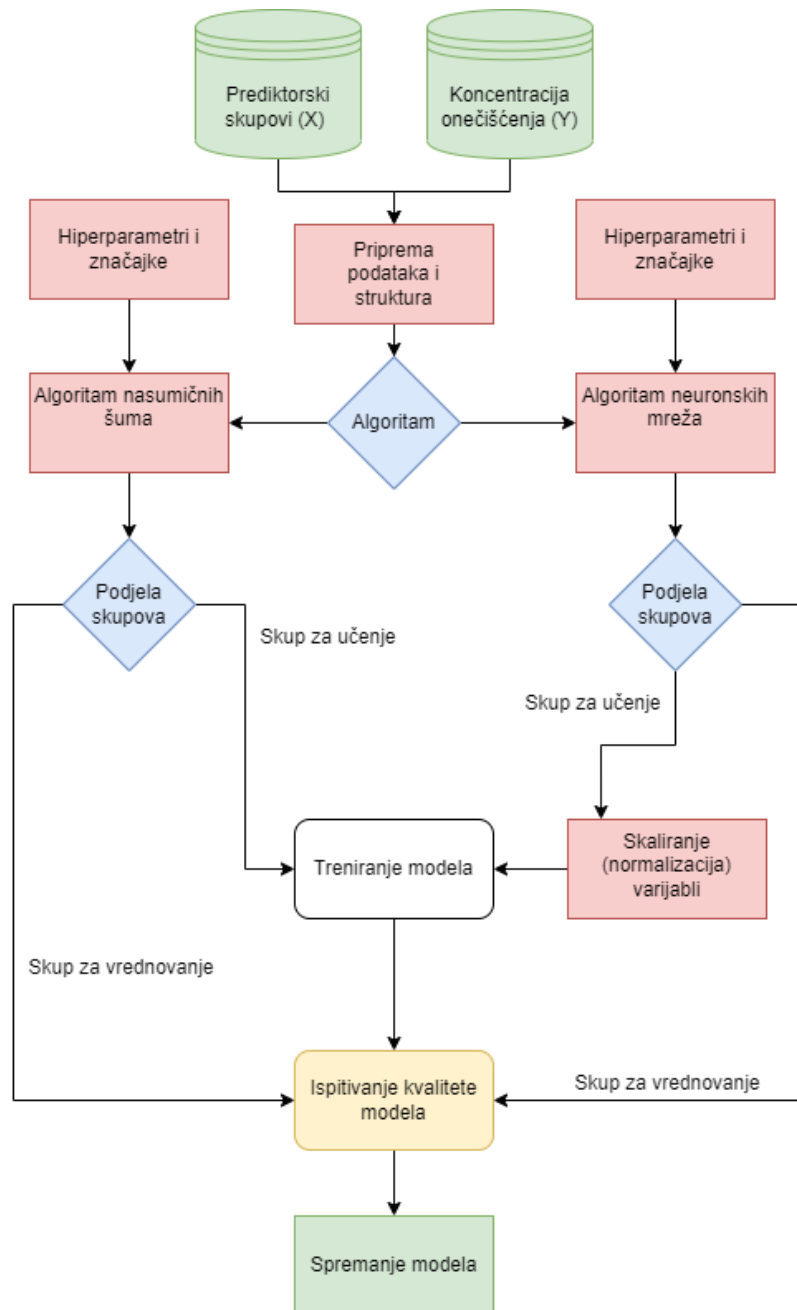
Tablica 7. Ulazne značajke modela za svaku od pet postaja

Lokacija	Temporalne varijable	Lokalne varijable	Satelitske varijable	Promet	Kombinacije lokalnih varijabli
Don Bosco	Dani u tjednu, mjeseci u godini, godine, državni praznici, školski praznici, godišnja doba	Temperatura, relativna vlažnost zraka	Temperatura zraka na 2 m, Naoblaka, Brzina vjetra na 10 m, Rosište, Tlak pare	DonBosco_8020_110, DonBosco_8020_119, DonBosco_8020_122, DonBosco_8053_111,	Temperatura i relativna vlažnost
Sjever	Dani u tjednu, mjeseci u godini, godine, državni praznici, školski praznici, godišnja doba	Temperatura, relativna vlažnost zraka, tlak, zračenje, oborine, brzina vjetra, maksimalna brzina vjetra u danu, smjer puhanja vjetra	Temperatura zraka na 2 m, Naoblaka, Brzina vjetra na 10 m, Rosište, Tlak pare	Sjever_8010_205, Sjever_8020_102, Sjever_8045_203, Sjever_8051_101	Temperatura i relativna vlažnost, temperatura i tlak, naoblaka i temperatura, relativna vlažnost i tlak, tlak i brzina vjetra, zračenje i brzina vjetra, zračenje i temperatura, temperatura i brzina vjetra
Istok	Dani u tjednu, mjeseci u godini, godine, državni praznici, školski praznici, godišnja doba	Temperatura, relativna vlažnost, tlak, brzina vjetra, maksimalna brzina vjetra u danu, smjer puhanja vjetra	Temperatura zraka na 2 m, Naoblaka, Brzina vjetra na 10 m, Rosište, Tlak pare	Istok_8010_208, Istok_8010_209, Istok_8010_213, Istok_8010_214, Istok_8041_221	Temperatura i relativna vlažnost, temperatura i tlak, naoblaka i temperatura, relativna vlažnost i tlak, tlak i brzina vjetra, temperatura i brzina vjetra
Zapad	Dani u tjednu, mjeseci u godini, godine, državni praznici, školski praznici, godišnja doba	Temperatura, relativna vlažnost, brzina vjetra, maksimalna brzina vjetra u danu, smjer puhanja vjetra	Temperatura zraka na 2 m, Naoblaka, Brzina vjetra na 10 m, Rosište, Tlak pare	Zapad_8020_103	Temperatura i relativna vlažnost, temperatura i brzina vjetra
Jug	Dani u tjednu, mjeseci u godini, godine, državni praznici, školski praznici, godišnja doba	Temperatura, relativna vlažnost, brzina vjetra, maksimalna brzina vjetra u danu, smjer puhanja vjetra	Temperatura zraka na 2 m, Naoblaka, Brzina vjetra na 10 m, Rosište, Tlak pare	Jug_8020_107, Jug_8041_218,	Temperatura i relativna vlažnost, temperatura i brzina vjetra

Izlazne vrijednosti modela također su predobrađene prije treniranja modela. Predobrada je provedena tako da su uklonjeni *outlieri*, točnije uklonjene su sve vrijednosti koje su veće od 99 percentila ili manje od 1 percentila. Vrijednosti koje su manje ili veće od zadanih parametara te odstupaju od trenda smatraju se izuzecima i rezultat su nekih anomalija (npr. povećani opseg prometa, hladnije vrijeme od prosjeka, jači vjetar itd.). Te vrijednosti unose pogrešku u model jer se ne mogu predvidjeti, npr. svi rezultati nakon 15.03 ne mogu se koristiti jer je tada nastupilo iznimno razdoblje *lockdown*-a te ih nije moguće sa sigurnošću predvidjeti modelom.

4.2 TRENIRANJE MODELA

Nakon što su se za svaku od pet postaja odredili optimalni hiperparametri modela i ulazne značajke, modeli su trenirani sa 5 različitih *random state*-ova kako bi se ocijenila kvaliteta modela. Za određivanje kvalitete modela korišteni su koeficijent determinacije i korijen srednje kvadratne pogreške na skupu za vrednovanje modela. Dijagram toka algoritma za treniranje modela prikazan je na slici 22. Prvo se za svaku od postaja odabiru podaci, kao npr. meteorološki podaci ili prometnica koja je najbliža postaji, a zatim se odabire algoritam razvoja modela. Nakon odabira algoritma skup podataka dijeli se na skup za učenje i skup za vrednovanje. Kod razvoja modela neuronskih mreža potrebno je skalirati podatke. Za svaki od algoritama uzimaju se odabrani hiperparametri i značajke modela. Preostaje još treniranje modela na skupu za učenje i ispitivanje kvalitete modela pomoću skupa za vrednovanje.



Slika 22. *Algoritam treniranja modela*

5 REZULTATI I RASPRAVA

5.1 REZULTATI OPTIMIZACIJE MODELA

Pomoću *GridSearchCV* i *Permutation Importance* algoritma provedena je optimizacija hiperparametara i ulaznih značajki modela za svaku od pet postaja. Odabrani hiperparametri i značajke ovisile su o primijenjenom algoritmu i korištenom skupu podataka. Ukupno je bilo 6 različitih skupova podataka i to su redom bili:

- Lokalni i temporalni podaci zajedno s kombinacijom lokalnih značajki
- Satelitski i temporalni podaci
- Lokalni, satelitski i temporalni podaci zajedno s kombinacijom lokalnih značajki
- Lokalni, temporalni podaci i promet te kombinacija lokalnih značajki
- Satelitski, temporalni podaci i promet
- Svi podaci zajedno

Zbog velikog broja podataka i jednostavnosti prikaza rezultata kao primjer optimizacije modela uzet je prvi skup podataka, tj. skup s lokalnim i temporalnim podacima te kombinacijom lokalnih značajki. U tablici 8. prikazani su optimirani hiperparametri.

Tablica 8. Hiperparametri dobiveni optimizacijom pomoću *GridSearchCV* algoritma

Lokacija	Algoritam	Hiperparametri modela
Don Bosco	Neuronske mreže	alpha': 0.0001, 'early_stopping': True, 'hidden_layer_sizes': (500, 100, 10), 'learning_rate_init': 0.0001, 'max_iter': 300, 'random_state': 42, 'solver': 'adam'
Don Bosco	Nasumične šume	max_depth': 10, 'max_samples': 0.5, 'min_samples_leaf': 2, 'n_estimators': 100, 'n_jobs': -1, 'random_state': 42},
Sjever	Neuronske mreže	alpha': 0.05, 'early_stopping': True, 'hidden_layer_sizes': (500, 100, 10), 'learning_rate_init': 0.0001, 'max_iter': 300, 'random_state': 42, 'solver': 'adam'
Sjever	Nasumične šume	max_depth': 10, 'max_samples': 0.5, 'min_samples_leaf': 2, 'n_estimators': 200, 'n_jobs': -1, 'random_state': 42},
Istok	Neuronske mreže	alpha': 0.05, 'early_stopping': True, 'hidden_layer_sizes': (500, 100, 10), 'learning_rate_init': 0.0001, 'max_iter': 300, 'random_state': 42, 'solver': 'adam'
Istok	Nasumične šume	max_depth': 10, 'max_samples': 0.5, 'min_samples_leaf': 2, 'n_estimators': 200, 'n_jobs': -1, 'random_state': 42},

Jug	Neuronske mreže	alpha': 0.05, 'early_stopping': True, 'hidden_layer_sizes': (500, 100, 10), 'learning_rate_init': 0.0001, 'max_iter': 300, 'random_state': 42, 'solver': 'adam'
Jug	Nasumične šume	max_depth': 10, 'max_samples': 0.5, 'min_samples_leaf': 2, 'n_estimators': 300, 'n_jobs': -1, 'random_state': 42},
Zapad	Neuronske mreže	alpha': 0.05, 'early_stopping': True, 'hidden_layer_sizes': (500, 100, 10), 'learning_rate_init': 0.0001, 'max_iter': 300, 'random_state': 42, 'solver': 'adam'
Zapad	Nasumične šume	max_depth': 10, 'max_samples': 0.5, 'min_samples_leaf': 2, 'n_estimators': 300, 'n_jobs': -1, 'random_state': 42},

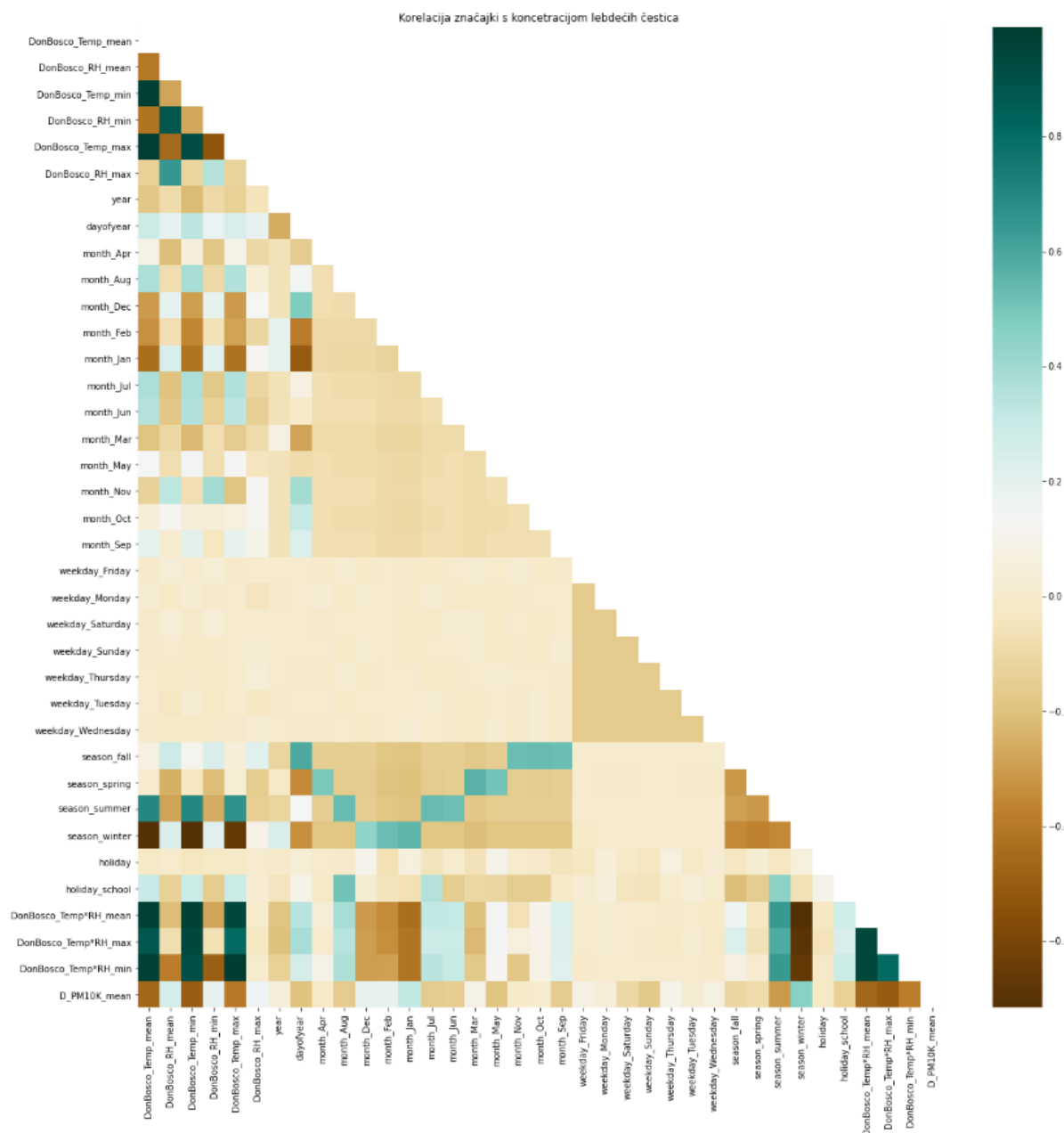
U tablici 9. prikazane su odabrane ulazne značajke modela za skup podataka s lokalnim i temporalnim podacima te kombinacijom lokalnih značajki. Na temelju ocjene *Permutation Importance* algoritma odabrana je 31 značajka za svaki model.

Tablica 9. Odabrane ulazne značajke modela za skup lokalnih meteorološki podataka

Lokacije	Algoritam	Značajke
Don Bosco	Neuronske mreže	Listopad, veljača, relativna vlažnost i temperatura max, prosinac, ožujak, travanj, zima, Temperatura min, petak, nedjelja, jesen, godina, temperatura i relativna vlažnost mean, temperatura mean, dan u godini, siječanj, ljeto, studeni, četvrtak, proljeće, školski praznici, ponedjeljak, temperatura max, srijeda, relativna vlažnost mean, utorak, temperatura i relativna vlažnost mean, subota, lipanj, srpanj, relativna vlažnost max
Don Bosco	Nasumične šume	Dan u godini, temperatura i relativna vlažnost max, temperatura min, relativna vlažnost mean, temperatura i relativna vlažnost min, temperatura max, temperatura mean, relativna vlažnost max, temperatura i relativna vlažnost mean, relativna vlažnost min, godina, nedjelja, školski praznici, petak, subota, veljača, ponedjeljak, listopad, četvrtak, ljeto, zima, proljeće, utorak, srijeda, prosinac, siječanj, ožujak, svibanj, travanj, rujana, studeni, praznici
Sjever	Neuronske mreže	Listopad, travanj, naoblaka mean, veljača, temperatura i relativna vlažnost max, ožujak, brzina vjetra mean, brzina vjetra max, naoblaka max, petak, školski praznici, temperatura i tlak, vjetar zapad-jugozapad, vjetar jug-jugoistok, temperatura i relativna vlažnost min, tlak min, temperatura min, zračenje i temperatura min, tlak mean, zračenje min, godina, temperatura i relativna vlažnost mean, relativna vlažnost i tlak mean, vjetar jugoistok-istok, temperatura max, brzina vjetara min, relativna vlažnost mean, siječanj, zima, proljeće, temperatura i brzina vjetra min, temperatura i tlak mean
Sjever	Nasumične šume	Brzina vjetra mean, brzina vjetra max, temperatura i relativna vlažnost max, dan u godini, temperatura i brzina vjetra mean, naoblaka i temperatura mean, tlak i brzina vjetra mean, zračenje min, tlak min, temperatura i brzina vjetra max, temperatura mean, relativna vlažnost i tlak min, tlak i brzina vjetra min, temperatura i relativna vlažnost mean, brzina vjetra mean, tlak i brzina vjetra max, relativna vlažnost i tlak mean, zračenje i temperatura mean, zračenje i brzina vjetra min, zračenje i temperatura min, relativna vlažnost mean, temperatura i brzina vjetra min, relativna vlažnost i tlak max, travanj, zračenje max, temperatura i relativna vlažnost min, temperatura i tlak min, tlak mean, naoblaka mean, zračenje i brzina vjetra mean
Istok	Neuronske mreže	Brzina vjetra mean, brzina vjetra max, nedjelja, godina, zima, travanj, temperatura i relativna vlažnost min, temperatura i tlak min, relativna vlažnost i tlakmin, vjetar zapad-jugozapad, Listopad, temperatura i relativna vlažnost max, temperatura i brzina vjetra min, temperatura min, školski praznici, temperatura i relativna vlažnost mean, petak, dan u godini, temperatura mean, ljeto, veljača, kolovoz, vjetar jugoistok-istok, brzina vjetra min, tlak min, ožujak, tlak mean, tlak max, proljeće, jesen, praznici, srijeda

Istok	Nasumične šume	Brzina vjetra mean, dan u godini, temperatura i brzina vjetra mean, temperatura i relativna vlažnost max, temperatura i tlak min, temperatura i brzina vjetra min, brzina vjetra max, temperatura i tlak max, temperatura i tlak max, tlak i brzina vjetra min, tlak min, relativna vlažnost i tlak max, tlak i brzina vjetra mean, temperatura max, temperatura i relativna vlažnost min, relativna vlažnost i tlak min, tlak i brzina vjetra max, temperatura i brzina vjetra max, brzina vjetra min, nedjelja, temperatura min, brzina vjetra max, tlak mean, relativna vlažnost min, tlak max, temperatura i relativna vlažnost mean, relativna vlažnost mean, relativna vlažnost i tlak mean, godina, školski praznici, temperatura mean, temperatura i tlak mean, brzina vjetra min
Jug	Neuronske mreže	Brzina vjetra mean, temperatura max, temperatura i relativna vlažnost max, relativna vlažnost min, brzina vjetra max, nedjelja, temperatura i relativna vlažnost min, veljača, godina, zima, ožujak, siječanj, travanj, prosinac, temperatura min, ljeto, temperatura i brzina vjetra max, brzina vjetra min, najveća brzina vjetra, temperatura mean, petak, rujan, dan u godini, temperatura i relativna vlažnost mean, jesen, subota, listopad, temperatura i brzina vjetra min, vjetar jug-jugoistok, temperatura i brzina vjetra mean, kolovoz, svibanj
Jug	Nasumične šume	Brzina vjetra mean, temperatura i relativna vlažnost max, dan u godini, temperatura i brzina vjetra mean, temperatura max, temperatura min, brzina vjetra max, brzina vjetra min, relativna vlažnost min, relativna vlažnost mean, temperatura i brzina vjetra min, temperatura i brzina vjetra max, temperatura i relativna vlažnost min, najveća brzina vjetra, temperatura mean, godina, relativna vlažnost max, temperatura i relativna vlažnost mean, ponedjeljak, veljača, nedjelja, travanj, zima, školski praznici, petak, jug-jugoistok, listopad, subota, proljeće, vjetar jugozapad-jug, vjetar istok-sjeveroistok, rujan
Zapad	Neuronske mreže	Listopad, travanj, brzina vjetra mean, brzina vjetra min, temperatura mean, brzina vjetra max, veljača, temperatura i relativna vlažnost mean, temperatura i relativna vlažnost max, zima, godina, temperatura min, nedjelja, najveća brzina vjetra mean, siječanj, relativna vlažnost min, prosinac, dan u godini, petak, vjetar jug-jugoistok, temperatura i relativna vlažnost min, vjetar jugozapad-jug, ponedjeljak, proljeće, jesen, temperatura max, vjetar istok-sjeveroistok, ožujak, ljeto, temperatura i brzina vjetra min, studeni, vjetar jugoistok-istok
Zapad	Nasumične šume	Brzina vjetra mean, temperatura i brzina vjetra mean, dan u godini, brzina vjetra max, temperatura max, temperatura i brzina vjetra min, brzina vjetra min, temperatura i relativna vlažnost max, najveća brzina vjetra, relativna vlažnost min, relativna vlažnost mean, temperatura i brzina vjetra max, temperatura min, temperatura i relativna vlažnost min, travanj, relativna vlažnost max, ponedjeljak, godina, listopad, školski praznici, veljača, vjetar jugozapad-jug, nedjelja, zima, vjetar istok-sjeveroistok, vjetar jug-jugoistok, petak, proljeće, srijeda, ožujak

Zbog velikog broja razvijenih modela na slici 23. prikazana je korelacija značajki i lebdećih čestica na primjeru postaje Don Bosco i skupu podataka s lokalnim meteorološkim podacima. Značajke čije vrijednosti su blizu 1 pokazuju pozitivnu korelaciju odnosno kako rastu njihove vrijednosti tako raste i vrijednost koncentracije lebdećih čestica, a vrijednosti blizu -1 pokazuju kako smanjenjem značajki dolazi do smanjenja lebdećih čestica.

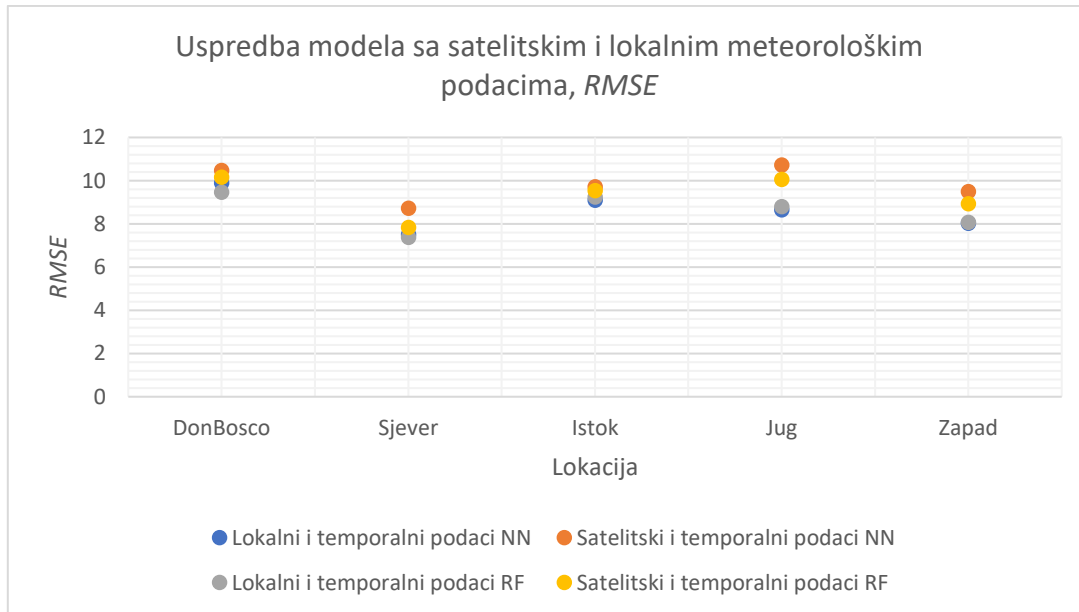


Slika 23. Korelacija ulaznih značajki i lebdećih čestica

5.2 SATELITSKI I LOKALNI METEOROLOŠKI PODACI

Nakon odabira hiperparametara i ulaznih značajki modela uslijedilo je treniranje modela. Modeli su trenirani 5 puta i dobiveno je 5 različitih procjena izlaza. Procijenjene vrijednosti izlaza uspoređene su sa stvarnim koncentracijama lebdećih čestica u zraku i izračunate su vrijednosti R^2 i $RMSE$. Dobivene vrijednosti R^2 i $RMSE$ korištene su za usporedbu utjecaja korištenog skupa podataka na procjenu modela pa su tako uspoređeni rezultati skupa s lokalnim meteorološkim podacima i skupa sa satelitskim meteorološkim podacima. $RMSE$ i

R^2 vrijednosti prikazane u ovom i slijedećim poglavljima računane su na skupu za vrednovanje modela. Na slici 24. prikazane su vrijednosti $RMSE$ za oba skupa podataka te za oba primijenjena algoritma (nasumične šume i neuronske mreža). Iz slike 24. se može vidjeti kako su modeli s lokalnim podacima bolji u procjeni koncentracije lebdećih čestica od modela sa satelitskim podacima. Kada se uspoređuju primijenjeni algoritmi, algoritam nasumičnih šuma postiže nešto bolje rezultate, te je i brži u računanju predikcije od algoritma neuronskih mreža.



Slika 24. Usporedba modela sa satelitskim i lokalnim meteorološkim podacima, $RMSE$

U tablici 10. prikazani su rezultati $RMSE$ za oba skupa podataka i primijenjene algoritme.

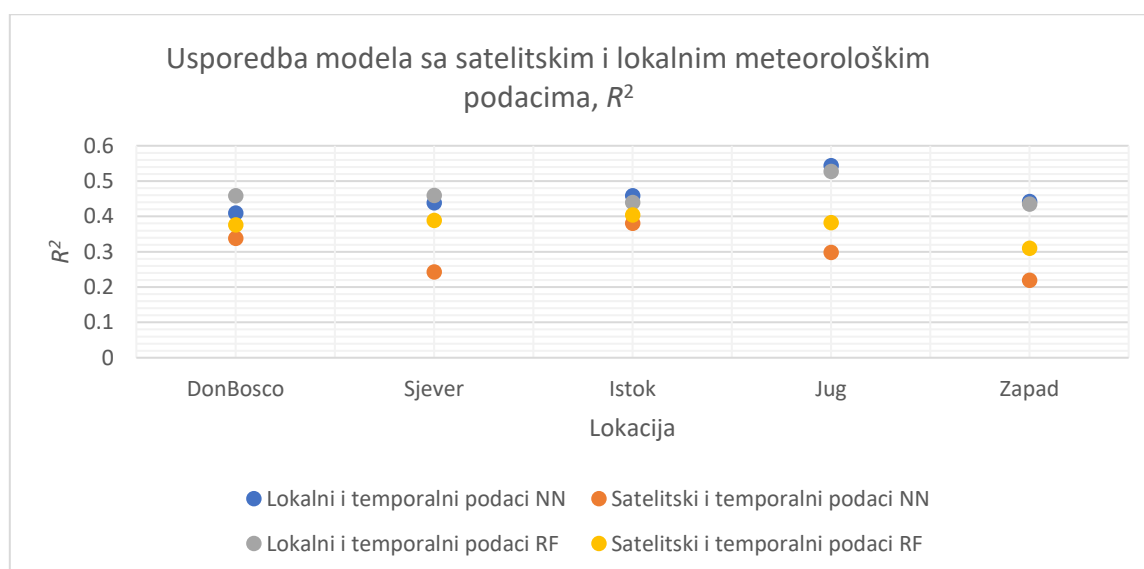
Tablica 10. Vrijednosti $RMSE$ za usporedbu satelitskih i lokalnih meteoroloških podataka

Lokacije	Lokalni i temporalni podaci NN	Satelitski i temporalni podaci NN	Lokalni i temporalni podaci RF	Satelitski i temporalni podaci RF
Don Bosco	9.89	10.48	9.47	10.17
Sjever	7.51	8.73	7.37	7.84
Istok	9.09	9.72	9.25	9.53
Jug	8.65	10.72	8.80	10.06
Zapad	8.02	9.50	8.08	8.93

Na slici 25. prikazani su koeficijenti determinacije modela R^2 i najbolje su se pokazali modeli s lokalnim meteorološkim podacima te se algoritam nasumičnih šuma pokazao malo boljim od neuronskih mreža. U tablici 11. prikazane su vrijednosti R^2 modela.

Tablica 11. Vrijednosti R^2 modela za usporedbu satelitskih i lokalnih meteoroloških podataka

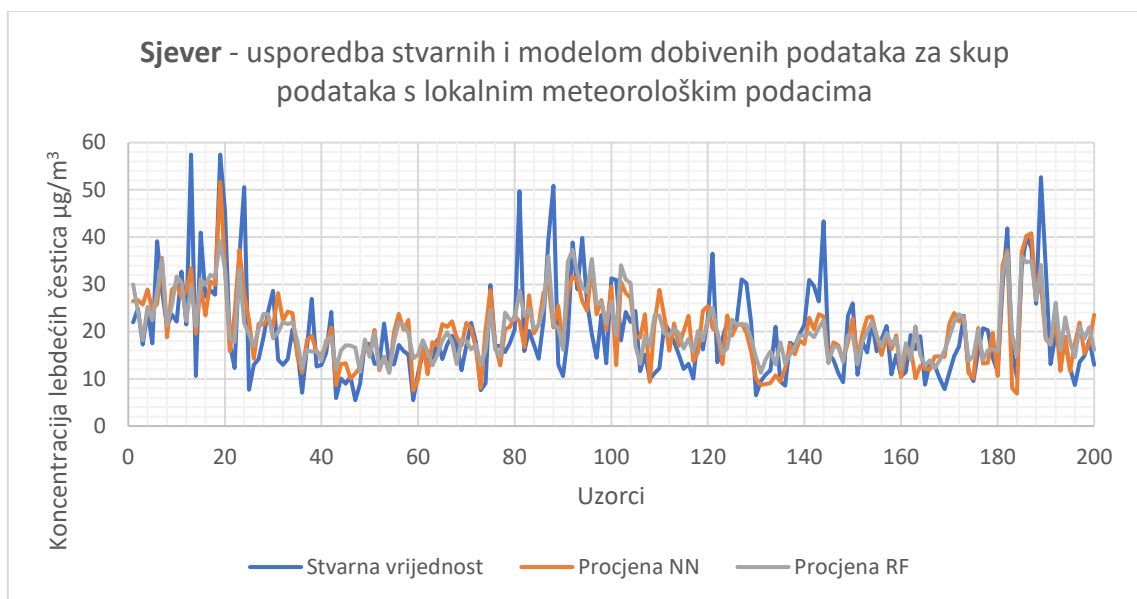
Lokacije	Lokalni i temporalni podaci NN	Satelitski i temporalni podaci NN	Lokalni i temporalni podaci RF	Satelitski i temporalni podaci RF
Don Bosco	0.40	0.33	0.45	0.37
Sjever	0.43	0.24	0.45	0.38
Istok	0.45	0.38	0.43	0.40
Jug	0.54	0.29	0.52	0.38
Zapad	0.44	0.21	0.43	0.30



Slika 25. Usporedba modela sa satelitskim i lokalnim meteorološkim podacima, R^2

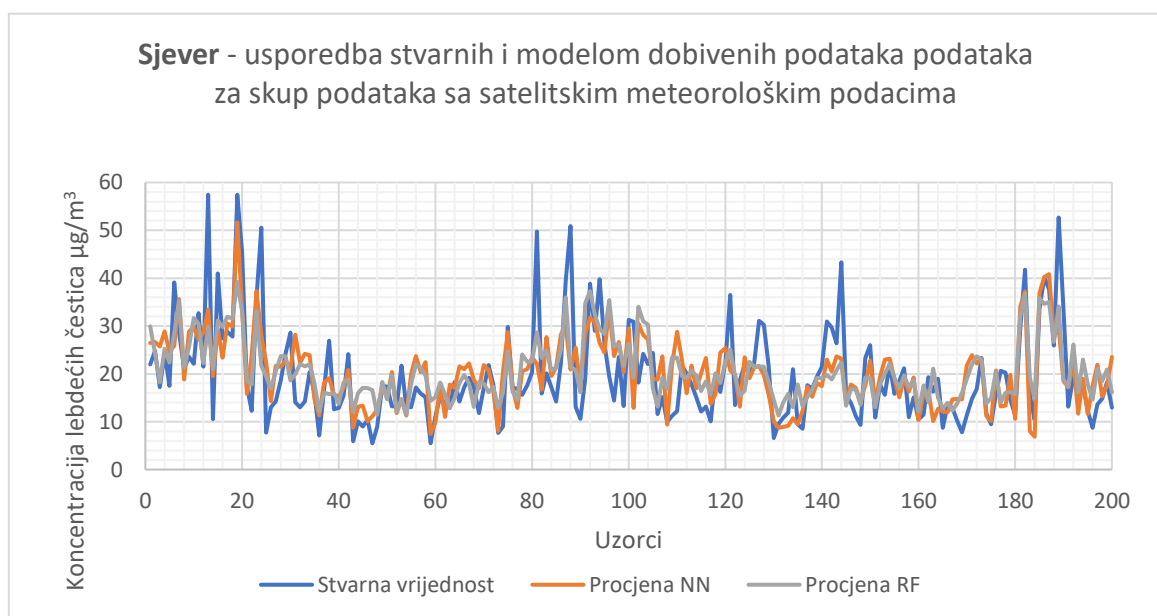
Rezultati usporedbe modela razvijenih sa meteorološkim i satelitskim podacima su i očekivani jer će lokalni utjecaj atmosfere imati više utjecaja na kvalitetu zraka od udaljenije atmosfere te će tako podaci koji su lokalno vezani za svaku od postaja nositi i više informacija.

Na slici 26. prikazana je usporedba vrijednosti stvarnih koncentracija lebdećih čestica i koncentracija koje su dobivene pomoću oba algoritma razvijenih na skupu s lokalnim meteorološkim podacima. Zbog jednostavnosti u ovom i ostalim poglavljima prikazat će se rezultati samo za postaju Sjever. Na slici 26. se može vidjeti kako model nešto lošije procjenjuje niske i visoke koncentracije čestica i to zato što su one rijetke pa model nema veliki broj podataka na kojima može učiti. Između dva algoritma nema velikih razlika te oba algoritma u cjelini dobro procjenjuju izlaznu varijablu.



Slika 26. *Usporedba stvarnih i modelom dobivenih podataka za skup podataka s lokalnim meteorološkim podacima*

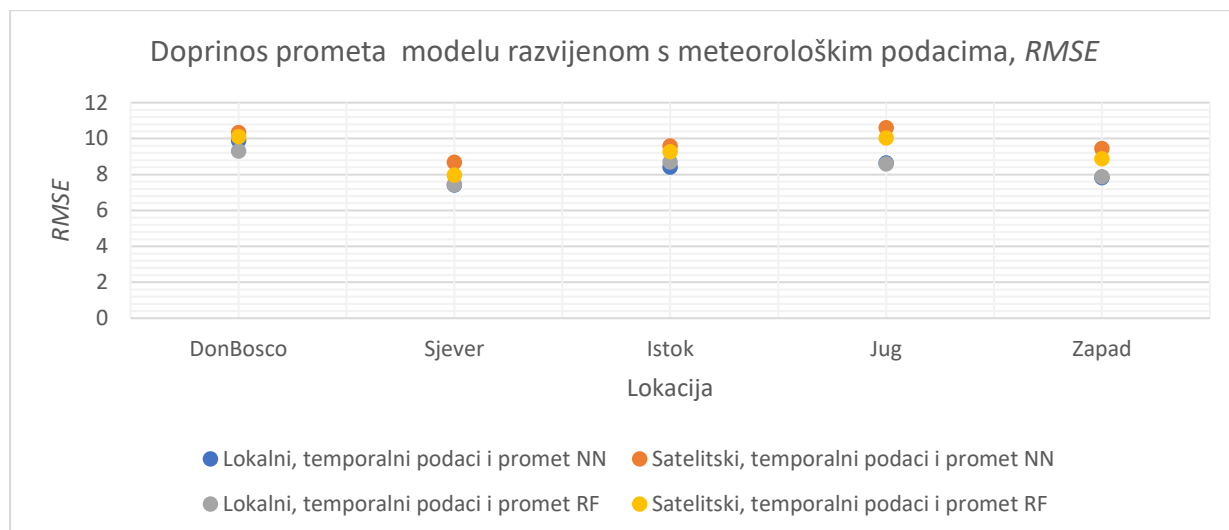
Na slici 27. prikazane su vrijednosti stvarnih koncentracija lebdećih čestica i koncentracija koje su dobivene pomoću dva algoritma razvijenih na skupu podataka sa satelitskim meteorološkim podacima. Kada se usporede slika 26. i 27. vidljivo je kako je model koji koristi skup s lokalnim meteorološkim podacima bolji u predikciji od modela sa satelitskim meteorološkim podacima. Razlog tome je što lokalni meteorološki podaci utječu više na koncentraciju čestica u zraku od satelitskih podataka iznad grada.



Slika 27. *Usporedba stvarnih i modelom dobivenih podataka za skup podataka sa satelitskim meteorološkim podacima*

5.3 DOPRINOS PROMETA MODELU ZA PROCJENU LEBDEĆIH ČESTICA

Uz lokalne i satelitske podatke u ovom radu korišteni su i podaci za promet dobiveni sa 16 lokacija u gradu i promatran je njihov utjecaj na kvalitetu modela. Na slici 28. prikazani su rezultati *RMSE* vrijednosti za skup (lokalni, temporalni podaci i promet te satelitski, temporalni podaci i promet). I ovdje su se skupovi s lokalnim meteorološkim podacima pokazali su se bolji dok u primijenjenim algoritmima nema značajnih razlika.



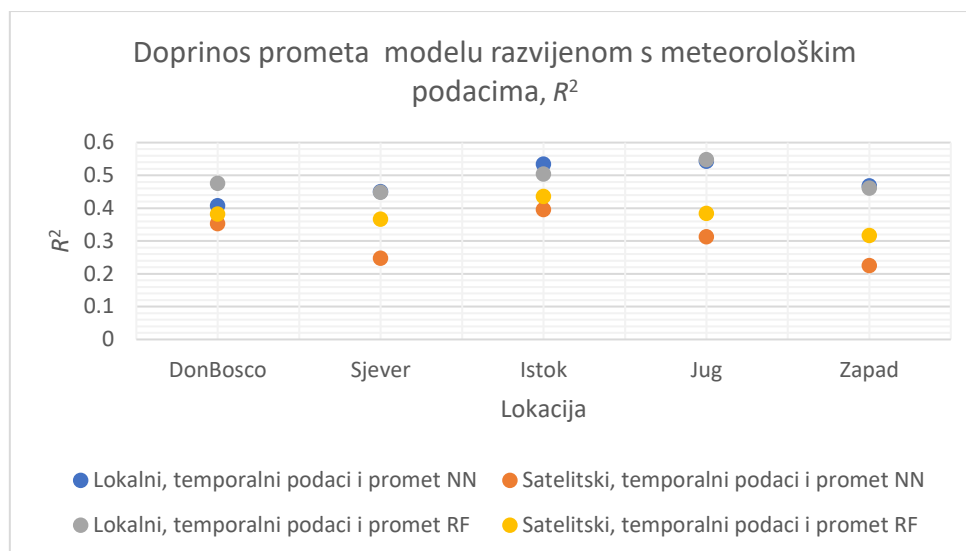
Slika 28. Doprinos prometa modelu razvijenog s meteorološkim podacima, *RMSE*

U tablici 12. prikazane su vrijednosti *RMSE* za doprinos prometa modelu za predviđanje lebdećih čestica u zraku. Kada se usporede rezultati s rezultatima iz prošlog poglavlja može se uočiti da je uvođenjem prometa u skup podataka došlo do poboljšanja modela.

Tablica 12. Vrijednosti *RMSE* za doprinos prometa meteorološkim podacima

Lokacije	Lokalni, temporalni podaci i promet NN	Satelitski, temporalni podaci i promet NN	Lokalni, temporalni podaci i promet RF	Satelitski, temporalni podaci i promet RF
Don Bosco	9.91	10.35	9.32	10.12
Sjever	7.43	8.70	7.45	7.98
Istok	8.43	9.60	8.70	9.28
Jug	8.65	10.61	8.61	10.04
Zapad	7.84	9.46	7.89	8.88

Na slici 29. prikazani su rezultati R^2 za doprinos prometa modelu s meteorološkim podacima te je i ovdje vidljivo kako su najbolje rezultate postigli modeli s lokalnim meteorološkim podacima i prometom. Kod postaja Jug i Zapad to povećanje iznosi i preko 10 posto u odnosu na satelitske podatke.



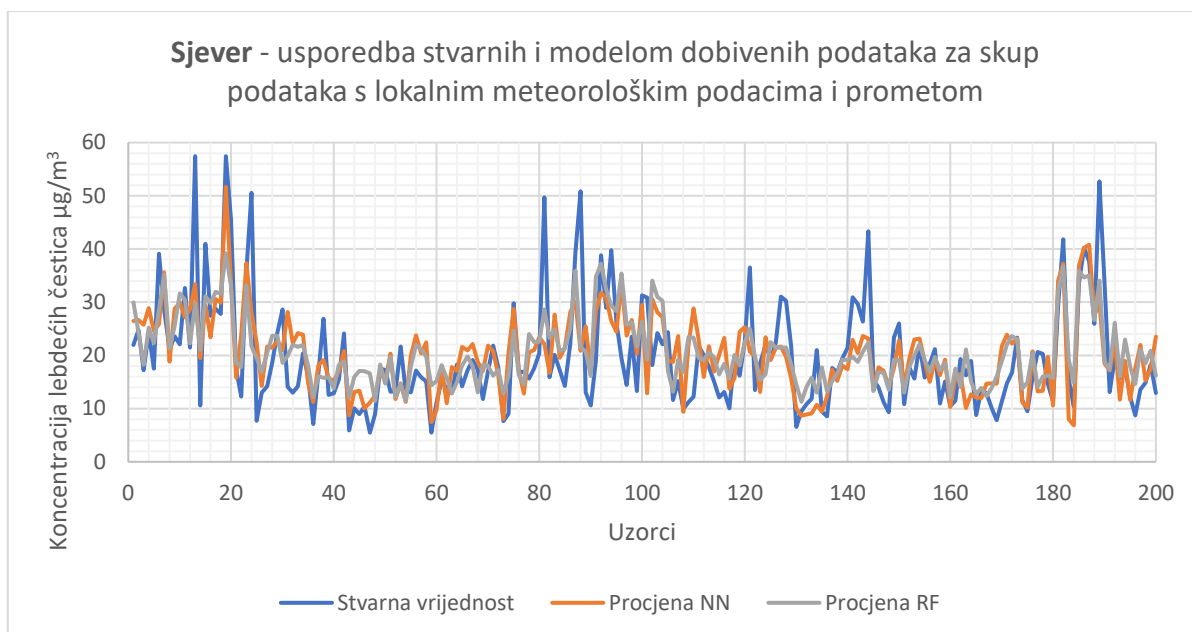
Slika 29. Doprinos prometa modelu s meteorološkim podacima, R^2

U tablici 13. prikazani su rezultati R^2 za doprinos prometa modelu s meteorološkim podacima i oni također potvrđuju tezu da je uvođenje podataka za promet poboljšalo kvalitetu modela.

Tablica 13. Vrijednosti R^2 za doprinos prometa modelu s meteorološkim podacima

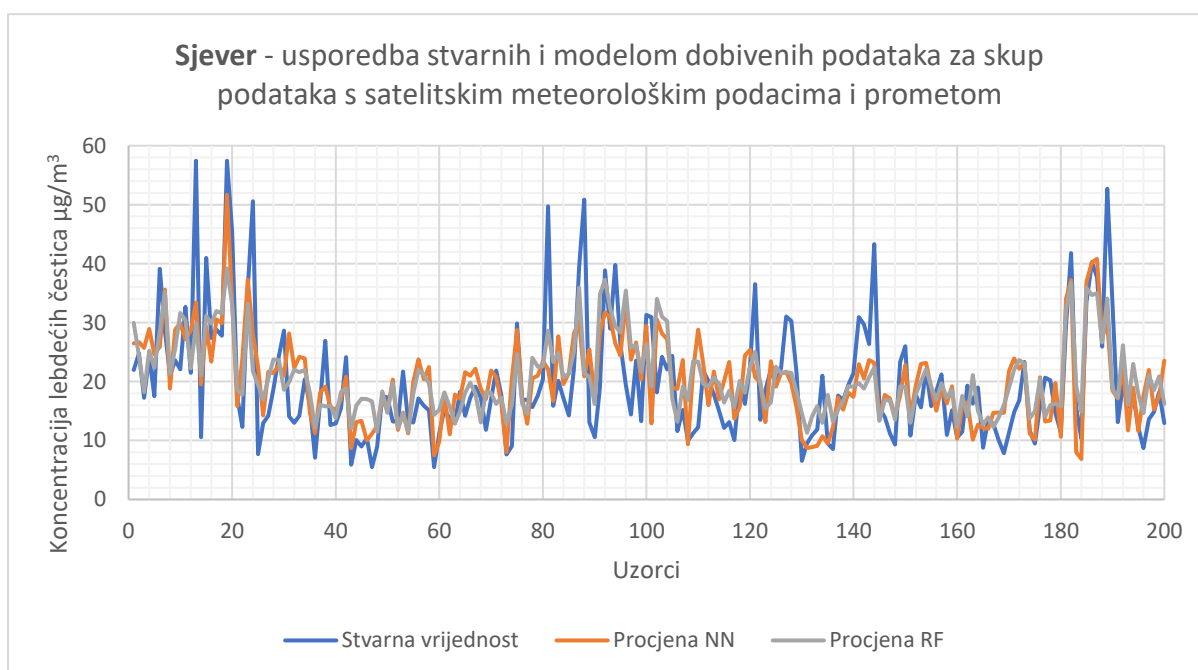
Lokacije	Lokalni, temporalni podaci i promet NN	Satelitski, temporalni podaci i promet NN	Lokalni, temporalni podaci i promet RF	Satelitski, temporalni podaci i promet RF
Don Bosco	0.40	0.35	0.47	0.38
Sjever	0.45	0.24	0.44	0.36
Istok	0.53	0.39	0.50	0.43
Jug	0.54	0.31	0.54	0.38
Zapad	0.46	0.22	0.46	0.31

Uvođenjem prometa kao ulazne varijable u model rezultiralo je manjim poboljšanjem u kvaliteti modela za sve postaje, pa se može zaključiti da prometnice koje okružuju postaje imaju direktan utjecaj na kvalitetu zraka mjenenog na postaji. Iz dobivenih rezultata može se vidjeti da su prometnice mnogo važnije značajke kod algoritma nasumičnih šuma, tj. njihova težina je bila znatno veća od težina u algoritmu neuronskih mreža. Na slici 30. prikazana je usporedba vrijednosti stvarnih koncentracija lebdećih čestica i koncentracija koje su dobivene pomoću dva algoritma na skupu s lokalnim meteorološkim podacima i podacima o prometu. Model treniran na skupu podataka s lokalnim i meteorološkim podacima bolje je procjenjivao izlaz od modela koji je treniran samo na lokalnim meteorološkim podacima. Međutim, model se i dalje nešto lošije vladala kod procjene visokih koncentracija lebdećih čestica.



Slika 30. *Usporedba stvarnih i modelom dobivenih podataka za skup podataka s lokalnim meteorološkim podacima i prometom*

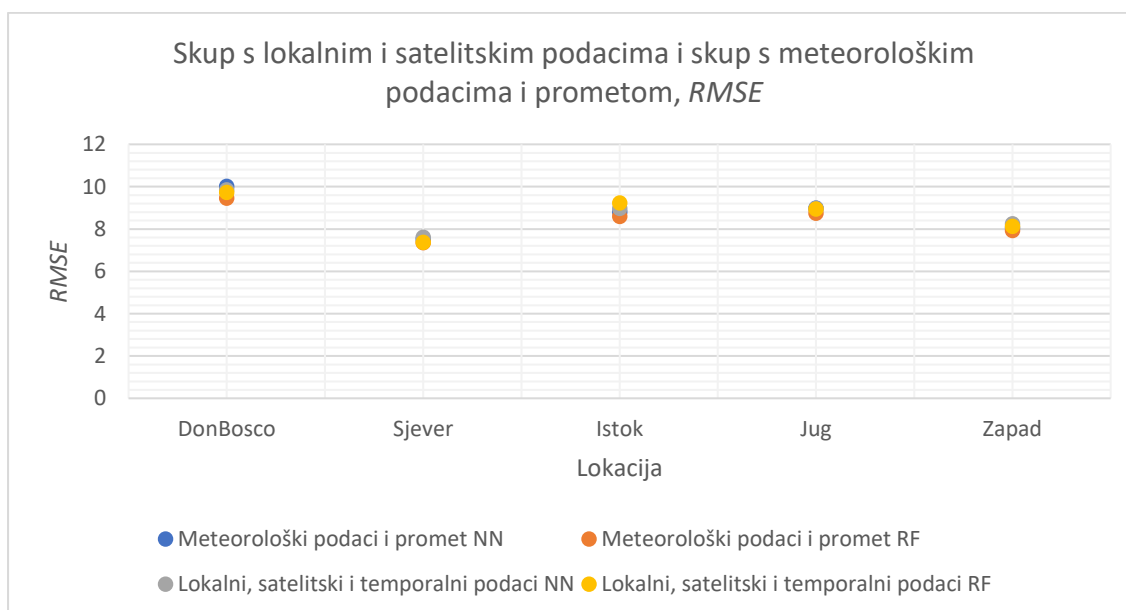
Na slici 31. prikazane su vrijednosti stvarnih koncentracija lebdećih čestica i koncentracija koje su dobivene pomoću dva algoritma za skup podataka sa satelitskim meteorološkim podacima i podacima o prometu. Model sa satelitskim meteorološkim podacima i prometom je ponovno nešto lošiji od modela s lokalnim meteorološkim podacima i prometom.



Slika 31. *Usporedba stvarnih i modelom dobivenih podataka za skup podataka sa satelitskim meteorološkim podacima i prometom*

5.4 USPOREDBA MODELA RAZVIJENOG SA SVIM DOSTUPNIM PODACIMA

U ovom poglavlju prikazani su rezultati za modele razvijene sa preostala dva skupa podataka, tj. skupa podataka gdje su spojeni lokalni i satelitski meteorološki podaci i skup u kojem su meteorološkim podacima dodani i podaci za promet. Na slici 32. prikazani su rezultati procjene modela sa ta dva skupa podataka sa koje se može vidjeti da su rezultati približno jednaki.



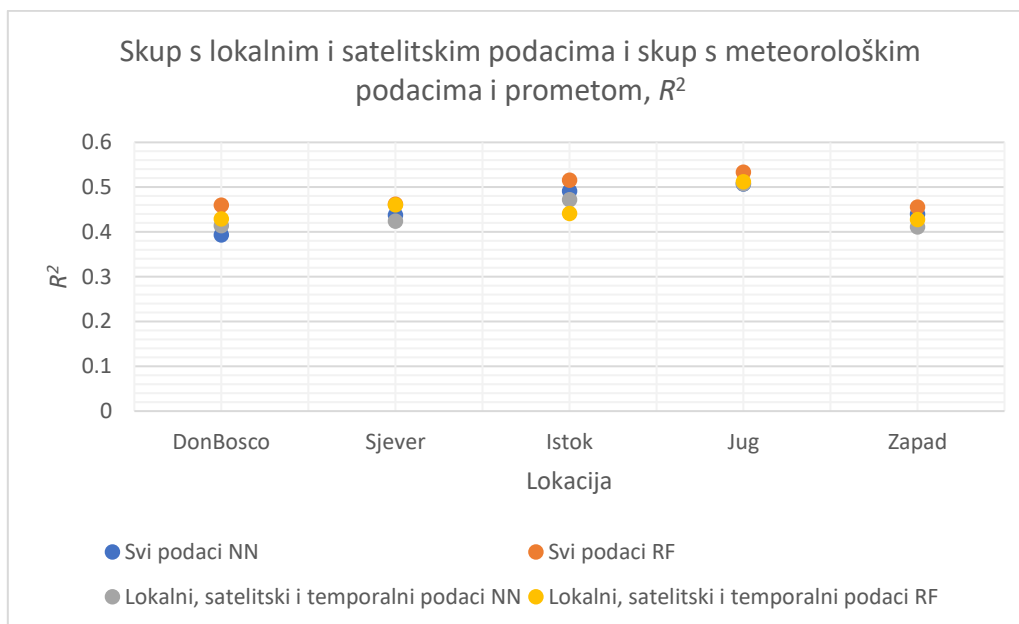
Slika 32. Rezultati modela s meteorološkim podacima (lokalni i satelitski) i rezultati modela s meteorološkim podacima (lokalni i satelitski) i prometom, RMSE

U tablici 14. prikazani su rezultati *RMSE* vrijednosti za dva skupa s meteorološkim podacima (lokalni i satelitski) i skup s meteorološkim podacima i prometom.

Tablica 14. Vrijednosti *RMSE* vrijednosti za dva skupa s meteorološkim podacima (lokalni i satelitski) i skup s meteorološkim podacima i prometom.

Lokacije	Lokalni, satelitski i temporalni podaci NN	Meteorološki podaci i promet NN	Lokalni, satelitski i temporalni podaci RF	Meteorološki podaci i promet RF
Don Bosco	9.85	10.02	9.73	9.47
Sjever	7.61	7.52	7.37	7.36
Istok	8.98	8.81	9.24	8.60
Jug	8.98	8.98	8.94	8.75
Zapad	8.25	8.05	8.13	7.93

Na slici 33. prikazani su koeficijenti R^2 te je vidljivo kako nije došlo do značajnih razlika u kvaliteti modela s obzirom na korištene podatke i algoritme.



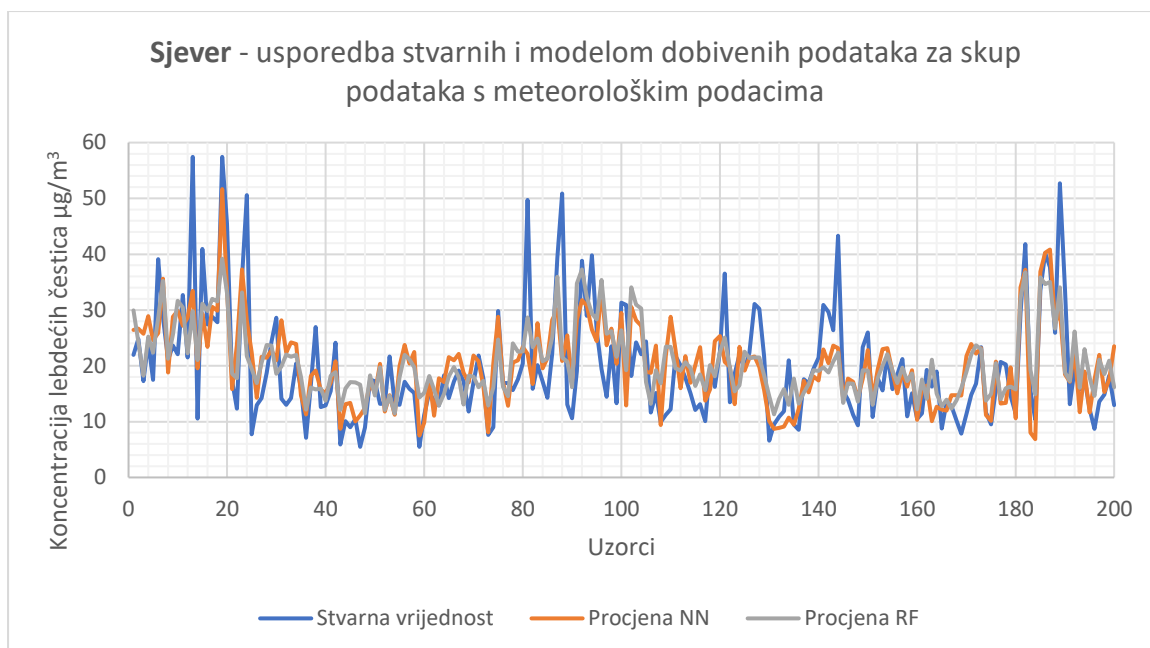
Slika 33. Rezultati modela s meteorološkim podacima (lokalni i satelitski) i rezultati modela s meteorološkim podacima (lokalni i satelitski) i prometom, R^2

U tablici 15. prikazani su rezultati R^2 vrijednosti za dva skupa s meteorološkim podacima (lokalni i satelitski) i skup s meteorološkim podacima i prometom.

Tablica 15. Vrijednosti R^2 za vrijednosti za dva skupa s meteorološkim podacima (lokalni i satelitski) i skup s meteorološkim podacima i prometom.

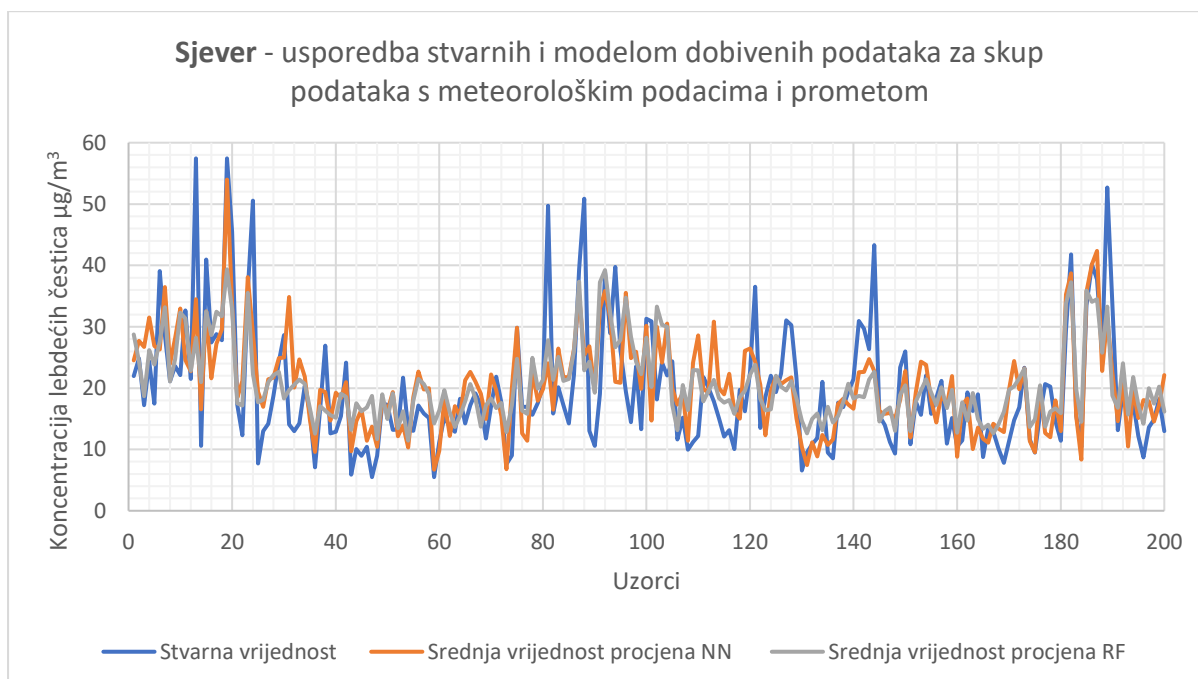
Lokacije	Lokalni, satelitski i temporalni podaci NN	Meteorološki podaci i promet NN	Lokalni, satelitski i temporalni podaci RF	Meteorološki podaci i promet RF
Don Bosco	0.41	0.39	0.42	0.45
Sjever	0.42	0.43	0.46	0.46
Istok	0.47	0.49	0.44	0.51
Jug	0.50	0.50	0.51	0.53
Zapad	0.41	0.43	0.42	0.45

Na slici 34. prikazane su vrijednosti stvarnih koncentracija lebdećih čestica i koncentracija koje su dobivene pomoću dva algoritma za skup podataka s lokalnim i satelitskim meteorološkim podacima. Spajanjem meteoroloških podataka došlo je do poboljšanja modela te je model bolje procjenjivao koncentraciju čestica, no i u ovom modelu problem je bio s visokim vrijednostima koncentracije lebdećih čestica.



Slika 34. *Usporedba stvarnih i modelom dobivenih podataka za skup podataka s meteorološkim podacima*

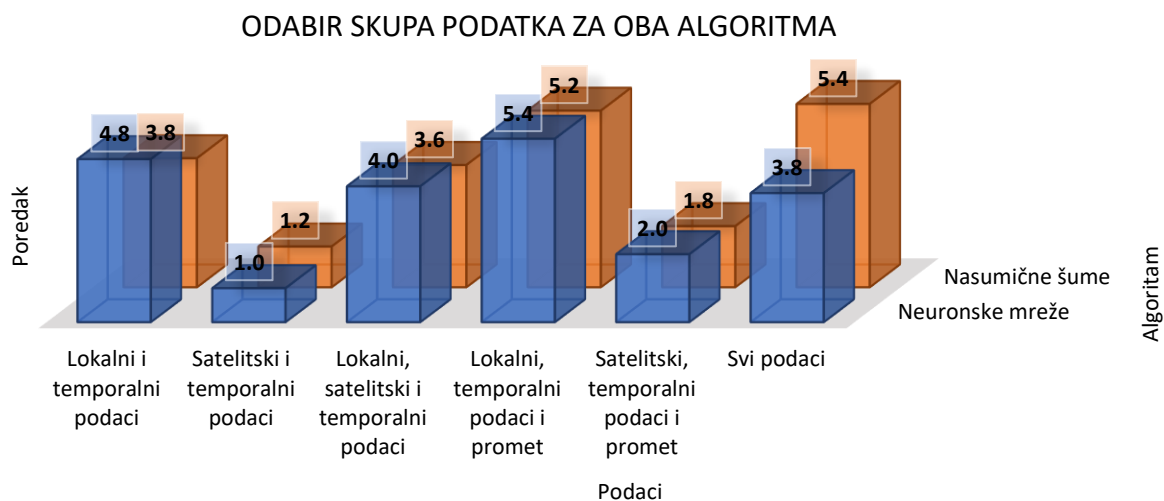
Na slici 35. prikazane su usporedbe vrijednosti stvarnih koncentracija lebdećih čestica i koncentracija koje su dobivene pomoću dva algoritma za skup podataka s meteorološkim podacima i prometom.



Slika 35. *Usporedba stvarnih i modelom dobivenih podataka za skup podataka s meteorološkim podacima i prometom*

5.5 ODABIR NAJBOLJEG ALGORITMA I SKUPA PODATAKA

U prethodnim poglavljima usporedili su se rezultati modela s različitim skupovima podataka i određen je utjecaj prometa na kvalitetu modela, a u ovom poglavlju prikazat će se najbolji skup podataka za svaki primijenjeni algoritam. Za procjenu modela s najboljim skupom podataka korišteni su R^2 i $RMSE$ vrijednosti te je najboljoj vrijednosti pridijeljen broj 6, a najlošijoj 1. Nakon toga je određena srednja vrijednost za svaki od korištenog skupa podataka i dobiveni rezultati prikazani su na slici 36. Nakon sistematizacije podataka vidljivo je kako su za oba primijenjena algoritma razvijeni modeli sa satelitskim skupom podataka dali najlošije rezultate i to sa i bez prometa. Za algoritam neuronskih mreža najbolji rezultati dobiveni su na skupu s lokalnim podacima gdje je skup koji uzima promet kao značajku nešto bolji u odnosu na onaj bez prometa. S druge strane, algoritam nasumičnih šuma najbolje rezultate pokazao je koristeći skup sa svim podacima, makar je i skup sa lokalnim podacima i prometom jako blizu najboljeg. Zanimljivo je da je kod modela neuronske mreže skup sa svim podacima treći po kvaliteti, a razlog tomu je što je u skupu sa svim podacima odabrano više kategoričnih varijabli od onih numeričkih što kod slučaja neuronskih mreža ne doprinosi boljoj kvaliteti modela.



Slika 36. Odabir skupa podataka za oba algoritma

6 ZAKLJUČAK

U ovom radu prikazana je primjena metoda umjetne inteligencije u svrhu razvoja modela za procjenu koncentracije lebdećih čestica u zraku. Modeli neuronskih mreža i nasumičnih šuma razvijeni su za procjenu koncentracije lebdećih čestica u zraku u austrijskom gradu Graz. Modeli su razvijeni za svaku od pet dostupnih mjernih postaja koristeći šest različitih skupova podataka što ukupno čini trideset modela. Za razvoj modela i primjenu algoritama neuronskih mreža i nasumičnih šuma korišten je programski jezik Python i njegove brojne knjižnice.

Prvo se ispitivao utjecaj meteoroloških podataka na kvalitetu modela i za to su korišteni lokalni podaci dobiveni sa pet mjernih postaja kao i satelitski podaci preuzeti sa stranice *Climate Data Store*. Modeli koji su razvijeni koristeći podatke s lokalnih mjernih postaja dali su bolje rezultate u oba primijenjena algoritma. Razlog boljih rezultata koristeći podatke s lokalnih postaja se može naći u tome što su lokalno mjereni meteorološki uvjeti precizniji te imaju veći utjecaj na kvalitetu lokalnog zraka od atmosfere iznad grada. Nakon toga se određivao doprinos prometa kao ulazne značajke na kvalitetu modela. Podaci za promet dobiveni su sa šesnaest mjernih postaja u gradu gdje je svaka prometnica pridružena najbližoj mjernoj postaji. Modeli razvijani na skupu sa meteorološkim podacima kojima su nadodani i podaci za promet pokazali su manja poboljšanja u kvaliteti modela. Iz dobivenih rezultata se stoga može zaključiti da promet i s udaljenih prometnica utječe na kvalitetu zraka na lokalnim mjernim postajama. Važno je napomenuti da su i ovdje modeli s lokalnim meteorološkim podacima i prometom bili bolji od modela razvijenih sa satelitskim meteorološkim podacima.

Na kraju je promatran i sinergijski utjecaj svih dostupnih ulaznih varijabli na kvalitetu modela pa su tako lokalni i satelitski podaci spojeni u jedan skup, a drugi skup je sadržavao i promet. Skup s lokalnim i satelitskim meteorološkim podacima i prometom je dao najbolje rezultate modela za procjenu lebdećih čestica u zraku pogotovo kod algoritma nasumičnih šuma. Kod algoritma neuronske mreže optimalan skup za razvoj modela je bio skup s lokalnim podacima i prometom. Zbog toga je to kontinuirani skup s podataka koji sadrži veću kvalitetu tj. bolju rezoluciju.

Modeli neuronskih mreža i nasumičnih šuma razvijeni u ovom radu mogu se primijeniti za predviđanje koncentracije lebdećih čestica na novim skupovima podataka.

7 POPIS OZNAKA I KRATICA

RF – Algoritam nasumičnih šuma

NN – Algoritam neuronskih mreža

RMSE – Korijen srednje kvadratne pogreške

R^2 – Koeficijent determinacije

PM – Lebdeće čestice

8 LITERATURA

1. Jeong CH, Yousif M, Evans GJ. Impact of the COVID-19 lockdown on the chemical composition and sources of urban PM_{2.5}. *Environ Pollut.* 2022;292(PB):118417. doi:10.1016/j.envpol.2021.118417
2. Ambient (outdoor) air pollution. Accessed May 9, 2022. [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
3. Segersson D, Eneroth K, Gidhagen L, et al. Health impact of PM₁₀, PM_{2.5} and black carbon exposure due to different source sectors in Stockholm, Gothenburg and Umea, Sweden. *Int J Environ Res Public Health.* 2017;14(7):11-14. doi:10.3390/ijerph14070742
4. Lovrić M, Pavlović K, Vuković M, Grange SK, Haberl M, Kern R. Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. *Environ Pollut.* 2021;274(xxxx). doi:10.1016/j.envpol.2020.115900
5. Brunekreef B, Holgate ST. Air pollution and health. *Lancet (London, England).* 2002;360(9341):1233-1242. doi:10.1016/S0140-6736(02)11274-8
6. New WHO Global Air Quality Guidelines aim to save millions of lives from air pollution. Accessed May 9, 2022. <https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution>
7. Allen JL, Klocke C, Morris-Schaffer K, Conrad K, Sobolewski M, Cory-Slechta DA. Cognitive Effects of Air Pollution Exposures and Potential Mechanistic Underpinnings. *Curr Environ Heal reports.* 2017;4(2):180. doi:10.1007/S40572-017-0134-3
8. Silva RA, West JJ, Zhang Y, et al. Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change. *Environ Res Lett.* 2013;8(3):034005. doi:10.1088/1748-9326/8/3/034005
9. 9 out of 10 people worldwide breathe polluted air, but more countries are taking action. Accessed May 9, 2022. <https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action>
10. Pennise D, Smith KR. Biomass Pollution Basics.
11. Goldstein AH, Koven CD, Heald CL, Fung IY. Biogenic carbon and anthropogenic pollutants combine to form a cooling haze over the southeastern United States. *Proc Natl Acad Sci U S A.* 2009;106(22):8835-8840. doi:10.1073/PNAS.0904128106/ASSET/4780F062-A21F-43C8-96F7-17F12F4FB175/ASSETS/GRAPHIC/ZPQ9990981510005.JPEG
12. Michelozzi P, Forastiere F, Fusco D, et al. Air pollution and daily mortality in Rome, Italy. *Occup Environ Med.* 1998;55(9):605. doi:10.1136/OEM.55.9.605
13. World's Most Polluted Cities in 2021 - PM_{2.5} Ranking | IQAir. Accessed May 9, 2022. <https://www.iqair.com/world-most-polluted-cities>
14. Karagulian F, Belis CA, Dora CFC, et al. Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmos*

- Environ.* 2015;120:475-483. doi:10.1016/J.ATMOSENV.2015.08.087
15. Amato F, Pandolfi M, Escrig A, et al. Quantifying road dust resuspension in urban environment by Multilinear Engine: A comparison with PMF2. *Atmos Environ.* 2009;17(43):2770-2780. doi:10.1016/J.ATMOSENV.2009.02.039
 16. Belis CA, Larsen BR, Amato F, et al. European Guide on with Receptor Models Air Pollution Source Apportionment Report EUR 26080 EN. doi:10.2788/9307
 17. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 3rd Edition | Wiley. Accessed May 9, 2022. <https://www.wiley.com/en-us/Atmospheric+Chemistry+and+Physics:+From+Air+Pollution+to+Climate+Change,+3rd+Edition-p-9781118947401>
 18. Hamra GB, Guha N, Cohen A, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environ Health Perspect.* 2014;122(9):906-911. doi:10.1289/EHP/1408092
 19. Raaschou-Nielsen O, Andersen ZJ, Beelen R, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol.* 2013;14(9):813-822. doi:10.1016/S1470-2045(13)70279-1
 20. What is Machine Learning? | IBM. Accessed June 6, 2022. <https://www.ibm.com/cloud/learn/machine-learning>
 21. 8 Steps to Build a Machine Learning Model - copyassignment.com. Accessed June 6, 2022. <https://copyassignment.com/8-steps-to-build-a-machine-learning-model/>
 22. Lovrić M. Razvoj i primjena modela za procjenu ekotoksikološkog rizika bioaktivnih kemijskih spojeva. Published online 2021.
 23. Dongare AD, Kharde RR, Kachare AD. Introduction to Artificial Neural Network. *Certif Int J Eng Innov Technol.* 2008;9001(1):2277-3754.
 24. 3.2. Tuning the hyper-parameters of an estimator — scikit-learn 1.1.1 documentation. Accessed June 6, 2022. https://scikit-learn.org/stable/modules/grid_search.html
 25. sklearn.neural_network.MLPRegressor — scikit-learn 1.1.1 documentation. Accessed June 6, 2022. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html
 26. sklearn.ensemble.RandomForestRegressor — scikit-learn 1.1.1 documentation. Accessed June 6, 2022. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>
 27. Bucchianico A Di. Coefficient of Determination (R²). *Encycl Stat Qual Reliab.* Published online March 15, 2008. doi:10.1002/9780470061572.EQR173
 28. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? *Geosci Model Dev Discuss.* 2014;7:1525-1534. doi:10.5194/gmdd-7-1525-2014
 29. Facts DF. *Austria. Figures. Data. Facts.*
 30. Graz climate: weather by month, temperature, precipitation, when to go. Accessed May 9, 2022. <https://www.climatestotravel.com/climate/austria/graz>

31. Thomas Stanzer, Irene Brischnik B phil. NB. Gentle Mobility The Graz Model of Success.
32. Graz Air Quality Index (AQI) and Austria Air Pollution | IQAir. Accessed May 9, 2022. <https://www.iqair.com/austria/styria/graz>
33. Welcome to Python.org. Accessed June 6, 2022. <https://www.python.org/>
34. Mario Lovrić, Michael Haberl, Mark Kröll, Valentino Petrić, Milana Vučković KP. *Air Pollution, Atmospheric and Local Meteorological Data for Graz, Austria from 2014 to End of 2021*,.; 2022. doi:<https://doi.org/10.5281/zenodo.6812067>
35. Šimić I, Lovrić M, Godec R, Kröll M, Bešlić I. Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon. *Environ Pollut.* 2020;263(2). doi:10.1016/j.envpol.2020.114587
36. vjetar | Hrvatska enciklopedija. Accessed May 10, 2022. <https://www.enciklopedija.hr/natuknica.aspx?ID=64995>
37. temperatura zraka | Hrvatska enciklopedija. Accessed May 10, 2022. <https://www.enciklopedija.hr/natuknica.aspx?ID=60787>
38. relativna vlažnost | Hrvatska enciklopedija. Accessed May 10, 2022. <https://www.enciklopedija.hr/natuknica.aspx?ID=52377>
39. tlak | Hrvatska enciklopedija. Accessed May 10, 2022. <https://enciklopedija.hr/natuknica.aspx?ID=61540>
40. oborina | Hrvatska enciklopedija. Accessed May 10, 2022. <https://www.enciklopedija.hr/natuknica.aspx?ID=44599>
41. radijacija | Hrvatska enciklopedija. Accessed May 10, 2022. <https://www.enciklopedija.hr/natuknica.aspx?id=51449>
42. Where is Graz, Austria on Map Lat Long Coordinates. Accessed May 10, 2022. <https://www.latlong.net/place/graz-austria-10475.html>
43. naoblaka | Hrvatska enciklopedija. Accessed May 10, 2022. <https://www.enciklopedija.hr/natuknica.aspx?id=42909>
44. rosište | Hrvatska enciklopedija. Accessed May 10, 2022. <https://www.enciklopedija.hr/natuknica.aspx?ID=53391>

9 ŽIVOTOPIS

Valentino Petrić [REDACTED] Pohađao je Gimnaziju Antun Gustav Matoš u Samoboru gdje je maturirao 2017. godine. 2017. godine upisao je Fakultet kemijskog inženjerstva i tehnologije gdje je 2020. godine pod vodstvom mentora prof. dr. sc. Domagoj Vrsaljka stekao akademsku titulu Sveučilišni prvostupnik inženjer kemijskog inženjerstva s radom Proizvodnja milireaktora stereolitografijom. Diplomski rad izradio je pod mentorstvom doc. dr. sc. Željke Ujević Andrijić na Zavodu za mjerenja i automatsko vođenje procesa u 2022. godini. U sklopu *Erasmus* stručne prakse proveo je tri mjeseca u Know – Centeru u Grazu na institutu za umjetnu inteligenciju.