

Analiza i predviđanje razine kvalitete zraka u Republici Hrvatskoj primjenom metode linearne regresije u alatu Weka

Ilić, Julia

Graduate thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Economics and Business / Sveučilište u Zagrebu, Ekonomski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:148:959899>

Rights / Prava: [Attribution-NonCommercial-ShareAlike 3.0 Unported / Imenovanje-Nekomercijalno-Dijeli pod istim uvjetima 3.0](#)

Download date / Datum preuzimanja: **2024-08-08**



Repository / Repozitorij:

[REPEFZG - Digital Repository - Faculty of Economics & Business Zagreb](#)



Sveučilište u Zagrebu

Ekonomski fakultet

Specijalistički diplomski stručni studij

Elektroničko poslovanje u privatnom i javnom sektoru

**ANALIZA I PREDVIĐANJE RAZINE KVALITETE ZRAKA U
REPUBLICI HRVATSKOJ PRIMJENOM METODE
LINEARNE REGRESIJE U ALATU WEKA**

Diplomski rad

Student: Julia Ilić, JMBAG :0055411662

Zagreb, rujan 2023.

Sveučilište u Zagrebu

Ekonomski fakultet

Specijalistički diplomski stručni studij

Elektroničko poslovanje u privatnom i javnom sektoru

**ANALIZA I PREDVIĐANJE RAZINE KVALITETE ZRAKA U
REPUBLICI HRVATSKOJ PRIMJENOM METODE
LINEARNE REGRESIJE U ALATU WEKA
ANALYSIS AND PREDICTION OF AIR QUALITY IN
CROATIA USING LINER REGRESSION MODEL IN WEKA
TOOL**

Diplomski rad

Student: Julia Ilić, JMBAG :0055411662

Mentor: Prof.dr.sc. Mirjana Pejić Bach

Zagreb, rujan 2023.

Potpisana izjava o akademskoj čestitosti

IZJAVA O AKADEMSKOJ ČESTITOSTI

Izjavljujem i svojim potpisom potvrđujem da je diplomski rad / prijava teme diplomskog rada isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju korištene bilješke i bibliografija.

Izjavljujem da nijedan dio rada / prijave teme nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog izvora te da nijedan dio rada / prijave teme ne krši bilo čija autorska prava.

Izjavljujem, također, da nijedan dio rada / prijave teme nije iskorišten za bilo koji drugi rad u bilokoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.


(vlastoručni potpis studenta)

Zagreb, 10.05.2023.

(mjesto i datum)

STATEMENT ON THE ACADEMIC INTEGRITY

I hereby declare and confirm by my signature that the final thesis is the sole result of my own work based on my research and relies on the published literature, as shown in the listed notes and bibliography.

I declare that no part of the thesis has been written in an unauthorized manner, i.e., it is not transcribed from the non-cited work, and that no part of the thesis infringes any of the copyrights.

I also declare that no part of the thesis has been used for any other work in any other higher education, scientific or educational institution.


(personal signature of the student)

Zagreb, 10.05.2023.

(place & date)

SAŽETAK

Najveću prijetnju za ljudsko zdravlje i okoliš predstavljaju lebdeće čestice, dušični dioksid i prizemni ozon. Smatra se kako je oko 90% stanovnika u europskim gradovima i skoro 99% stanovnika na globalnoj razini izloženo svakodnevno određenim razinama onečišćivača u zraku koji imaju za posljedicu srčane bolesti, moždani udar, kronične plućne bolesti tj. bolesti dišnog sustava, rak. Unatoč smanjenju koncentracije štetnih čestica u zraku te promicanje "pametnih gradova", buđenje svijesti o ekologiji, koncentracije štetnih tvari u zraku su i dalje previsoke. U ovom radu objašnjena je specifična primjena podatkovne znanosti, te strojnog učenja za predikcije daljnjeg onečišćenja zraka. U svrhu analize podataka za izgradnju modela za predikciju kvalitete zraka na području Hrvatske, prikupljeni su podaci mjernih stanica za 2021g. koji pokazuju kvantitativne i mjerljive podatke štetnih tvari u zraku te ostali sekundarni podaci o broju stanovništva, broju registriranih vozila u prometu, te utvrđivanje postojanja raznih zagađivača kao što su industrije metala, plina, odlagališta smeća koje treba uzeti u obzir ponajviše zbog signifikantnosti za ishod predviđanja onečišćenja zraka. U procesu izgradnje prediktivnog modela strojnog učenja koristilo se nekoliko algoritama nadziranog učenja za zadatke rudarenja podacima kao su stablo odlučivanja i linearna regresija, te neuronske mreže za duboko učenje: višeslojni preceptor (MLP). Regresijskom analizom između više varijabli kao što je broj stanovništva, broja vozila u prometu, postojanje zagađivača, te izmjerni podaci o koncentraciji štetnih tvari u zraku, utvrditi će se njihova međuovisnost, što će na kraju rezultirati prediktivnim modelom emisije štetnih tvari u zraku kao što su lebdeće čestice $PM_{2.5}$ i PM_{10} , ugljični monoksid CO, sumporov dioksid SO_2 , dušikov dioksid NO_2 , O_3 prizemni ozon, sumporovodik (H_2S), amonijak (NH_3), te benzen (C_6H_6)

Ključne riječi: Strojno učenje, rudarenje podataka, linearna regresija, neuronske mreže, podatkovna znanost, kvaliteta zraka, stablo odlučivanja.

ABSTRACT

The dangerous threat to human health and the environment is represented by floating particles, nitrogen dioxide and ground-level ozone. It is believed that about 90% of the population in European cities and almost 99% of the population at the global level are exposed to certain levels of pollutants in the air every day, which result in heart disease, stroke, chronic lung diseases, i.e. diseases of the respiratory system, cancer. Despite the reduction of the concentration of harmful particles in the air and the promotion of "smart cities", the awakening

of awareness about ecology, the concentrations of harmful substances in the air are still too high. This paper explains the specific application of data science and machine learning in the prediction of further air pollution. For the purpose of data analysis for building a model for predicting air quality in Croatia, data from measuring stations for 2021 were collected, which show quantitative and measurable data on harmful substances in the air and other secondary data on the number of the population, the number of registered vehicles in traffic, and the determination of the existence of various pollutants such as metal industries, gas, garbage disposal sites that should be taken into account mainly due to their significance for the prediction outcome air pollution. In the process of building a predictive model of machine learning, several supervised learning algorithms were used for data mining tasks such as decision tree and linear regression, and neural networks for deep learning: multilayer preceptor (MLP). Regression analysis between several parameters such as the number of population, number of vehicles in traffic, the existence of pollutants, and measured data on the concentration of harmful substances in the air, will determine their mutual interdependence, which will eventually result in a predictive model of the emission of harmful substances in the air, especially PM_{2.5} and PM₁₀. floating particles PM_{2.5} and PM₁₀, sulfur dioxide SO₂, nitrogen dioxide NO₂, ground ozone O₃, hydrogen sulphide (H₂S), ammonia (NH₃), and benzene (C₆H₆).

Keywords: Machine learning, data mining, linear regression, neural networks, data science, air quality, decision tree.

Sadržaj

1. Uvod.....	7
1.1. Predmet i cilj rada	7
1.2. Izvori podataka i metode prikupljanja	7
1.3. Struktura rada	8
2. Kvaliteta zraka u Republici Hrvatskoj i njezin utjecaj na zdravlje ljudi i okoliša	9
2.1. Procesi koji utječu na kakvoću zraka.....	9
2.2. Državna mreža za trajno praćenje kakvoće zraka na teritoriju RH.....	17
2.3. Zakonska podloga praćenja i procjene onečišćenja zraka u RH	24
3. Uloga i primjena podatkovne znanosti u praćenju kvalitete zraka	26
3.1. Pojam i razvoj podatkovne znanosti	26
3.2. Područja primjene podatkovne znanosti u meteorologiji i hidrologiji	29
3.3. Specifičnosti primjene podatkovne znanosti u prikupljanju i obradi podataka o kvaliteti zraka	30
4. Rudarenje podataka i strojno učenje	37
4.1. Prediktivna analitika i rudarenje podataka	37
4.2. Metode rudarenja podataka	40
4.3. Strojno učenje.....	46
4.3.1. Neuronske mreže	47
4.3.2. Stablo odlučivanja	50
4.3.3. Linearna regresija	53
5. Korištenje alata Weka u analizi i predviđanju kvalitete zraka	57
5.1. Metodologija istraživanja	57
5.2. Rezultati istraživanja.....	58
5.2.1. Deskriptivna analiza varijabli.....	58
5.2.2. Izrada modela predviđanja razine onečišćenja zraka	60
5.3. Predviđanja budućih trendova rudarenja podataka i strojnog učenja o kvaliteti zraka .	67
6. Zaključak	69
Popis slika	74
Popis tablica.....	76
Životopis kandidata.....	77

1. Uvod

1.1. Predmet i cilj rada

Sposobnost za kvalitetnu analizu podataka i izradu pravila za gradnju prediktivnog modela od velikog je značaja i vrijednosti za poslovne organizacije, vlade, donositelje odluka općenito a posebno u današnje vrijeme globalizacije i ubrzanog načina života. Upravo zbog ubrzanog razvoja idustrijalizacije i urabanizacije tijekom proteklih desetljeća, pojavile su se posljedice koje direktno utječu na ljudsko zdravlje i ekosustav. Glavna prijetnja je zagađenje zraka kojima nisu obuhvaćene samo zemlje u razvoju, već i razvijene zemlje, izravno i neizravno. Klimatske promjene su dokaz kako je zagađenje i način života po mjeri današnjeg čovjeka učinilo njegov opstanak neodrživim. Razne organizacije prikupljaju na dnevnoj bazi podatke o kvaliteti zraka. Ostali podaci koji se prikupljaju u statističke svrhe u RH jesu podaci o gustoći prometa i broju stanovništva na području mjernih stanica, te će se isti podaci koristiti za analizu međuovisnosti. Također, za potrebe analize uzima se u obzir i postojanje industrijskih postrojbi, te odlagališta za otpad, pod zajedničkim imenom zagađivači, koje se nalaze u neposrednoj blizini mjernih stanica za praćenje kvalitete zraka. Podaci se trebaju analizirati u svrhu otkrivanja posljednično-uzročnih veza kako bi se izgradio prediktivni model. Na temelju intepretacije modela i saznanja donosile bi se odluke o smanjenju koncentracije štetnih tvari u zraku. Razvojem tehnologije tj. performansi računala moguće simulirati događaje na temelju prikupljenih podataka također i predvidjeti buduću razinu kvalitete zraka. U svrhu izgradnje prediktivnih modela za kvalitetu zraka u Republici Hrvatskoj na prikupljenim podacima koristiti će se algoritmi strojnog učenja, rudarenjem prikupljenih sekundarnih podataka u Weka okruženju za analizu znanja.

1.2. Izvori podataka i metode prikupljanja

Za potrebe ovog rada koristiti će se izvori sekundarnih podataka sa web stranica Državnog hidrometeorološkog zavoda, tj. Državne mreže za trajno praćenje kvalitete zraka i Nastavnog zavoda za javno zdrastvo Splitsko-dalmatinske županije, Odjel za ispitivanje zraka, tla i buke, koji sadrže podatke iz 2021. godine o izmjerenim podacima kvalitete zraka u Hrvatskoj koji su zabilježeni na mjernim postajama. Koristiti će se podaci o broju stanovnika sa popisa iz 2021, izvor je web stranice Državnog zavoda za statistiku RH. Također će se koristiti podaci o broju registriranih vozila, po vrsti vozila i vrsti goriva u 2021.godini sa web stranica Centar za vozila Hrvatske. Podaci o ukupnom broj vozila u prometu su prikupljeni na Stanica za tehnički

pregled. Tehnički pregled je obavezan po Zakonu RH za sva vozila u prometu. Kako bi se izgradio model sa većom preciznošću predikcije uzima se u obzir i postojanje zagađivača kao što su razne industrije i odlagališta u neposrednoj blizini mjernih stanica. Osim podataka koristiti će se znanstvena i akademska istraživanja, te razna dostupna literature kao izvor informacija i citiranja. U radu su korištene metode deskriptivne analize (korištene pri raščlanjivanju i opisu elemenata cjelina koje su obrađivane u ovom radu s ciljem utvrđivanja elemenata, sadržaja i sastavnica promatrane cjeline kao i odnosa u cjelini) i sinteze (pri sjedinjavanju jednostavnih misaonih tvorevina u složenije povezujući elemente, procese, pojave i odnose u cjelinu), induktivna i deduktivna metoda (u svrhu prezentiranja uopćenih zakonitosti i smanjenja apstrakcija) te metoda kompilacije (citiranjem i grafičkim prikazima koji su preuzeti iz korištene literature).

1.3. Struktura rada

Ovaj rad se sastoji od šest cjelina (poglavlja). U prvoj cjelini je objašnjeno predmet i cilj rada u kojem se prezentiraju metode istraživanja i vrsta prikupljenih podataka. Ističe se svrha i način analize prikupljenih podataka koja ima za cilj izgradnju modela predikcije kvalitete zraka u budućnosti na području Hrvatske. Druga cijelina se opisuje procese i događaje koji utječu na kvalitetu zraka u Hrvatskoj, kako to utječe na zdravlje stanovništva, te svrhu i način prikupljanja podataka od strane državne mreže, te opisuje sadašnju zakonsku regulativu. U trećem poglavlju prezentira se razvoj podatkovne znanosti do danas, značajna uloga i primjena podatkovne znanosti u znanstvenim istraživanjima metrologije i hidrologije, te specifična primjena za ovaj rad, obradi podataka o kvaliteti zraka. Četvrta cijela (pogavlje) opisuje procese (korake) strojnog učenja, tehniku rudarenja podataka, objašnjava najčešće korištene zbirke algoritama te posebno linearnu regresiju. Peta cijelina opisuje metodologiju istraživanja od obrade podataka do izgradnje modela, međusobne usporedbe rezultata korištenja algoritama strojnog učenja te predviđanja budućih trendova rudarenja podataka u svrhu predikcije kvalitete zraka. Na kraju, u šestoj cijelini donosi se zaključak koji objedinjuje sve elemente ovog rada, analizirajući pritom sve nedostatke i koristi koje se utvrdile temeljem ovog rada.

2. Kvaliteta zraka u Republici Hrvatskoj i njezin utjecaj na zdravlje ljudi i okoliša

2.1. Procesi koji utječu na kakvoću zraka

Unatoč razvoju tehnologije i raznih tehnoloških dostignuća, većina država svoj gospodarski rast ostvaruje kroz granu industriju. Industrija je i dalje predstavlja jednu od najvažnijih poluga gospodarske održivosti te ekonomske stabilnosti. Još krajem 18. st. industrija se intenzivno razvijala sa izumom parnog stroja. U 19 st. otkrićem električne energije došlo je do druge industrijske revolucije koja je dovela do ekspanzije rasta manufakture i tvornica raznih djelatnosti. Treća industrijska revolucija započinje pojavom automatizacije i mehanizacije u proizvodnji, dok četvrta revolucija je donijela tehnološki napred u pogledu razvoja virtualnih uređaja sa fizičkom podrškom i IoT (engl. Internet of things). U Hrvatskoj tek valom druge industrije počela se razvijati industrija, tekstilna, drvna i prerađivačka industrija, nakon nje slijede, elektroprivreda, prerada metala, kemijska te ostalo. Također zastupljene su bile cementare i kamenolomi, dok je prva rafinerija nafte u Hrvatskoj, a ujedno jedna od prvih u Europi, počela sa radom 1883. u Rijeci, neposredno u blizini riječke luke. Industrija kako je već spomenuto je jedna od najvažnijih poluga gospodarstva i ekonomskog rasta jedne zemlje. Razvojem industrije došlo je do kako ubrzanog razvoja gospodarstva tako i cijelog društava globalno gdje se manualni rad pretvorio u tvornički rad. Ubrzao se način života i došlo je stvaranja velikih gradova. U svojim istraživačkim znanstvenim radovima Zdenka Šimončić-Bobetko opisuje razvoj tekstilne i cementne industrije u prošlom stoljeću u Hrvatskoj te navodi zajedničke osobine koje su uslijedile razvoju tih industrija kao prometnu povezanost, izvor lokacije za eksploataciju, iskorištenja prirodnih sredstva zbog društvene i ekonomske koristi. ” Voda je služila za natapanje konoplje. Kudjeljare su se radi toga najčešće podizale u blizini rijeka i kanala. Međutim voda se mogla dobiti i drukčije, npr. bušenjem bunara, stvaranjem akumulacija ili iz vodovodne mreže ali sve je to poskupljivalo proizvodni proces” (Mirković, et al 179-185; Mišić, et al, 147-148.) Iako su materijali od kojih se radila odjeća bili prirodni materijali, kemikalije koje se koriste u obradi odjeće otpuštaju se u zrak i okoliš. Međutim jedna stvar se stalno ponavlja kao uvjet ekonomske održivosti industrije je prometna povezanost zbog transportiranja proizvoda. “Daljni faktor u određivanju prostornog smještaja tekstilne industrije bio je i promet. Prilikom izbora dana je obično prednost mjestima koja su bila na prometnim arterijama ili onima koja su prometno povezana s glavnim prometnicama”

(Šimončić-Bobetko, 1983., 101-204). Tvornice cementa su se gradile duž jadranske obale uz obrazloženja ekonomičnosti transporta za izvoz pomorskim putem, dok s druge strane transport cementa unutar zemlje sa obale je bio nepovoljan zbog kako se navodi dugih relacija. “Na lokaciju cemente industrije utjecala je svakako mogućnost plasmana naročito u izvozu, ali su od jednakog utjecaja, ako ne i većeg, bila kvaliteta i bogata nalazišta sirovina. Međutim, u Hrvatskoj je došlo i do osnivanja nekih tvornica cementa kod kojih kao lokacioni faktor dobiva na značenju i mjesto potrošnje kao npr. tvornica Croatia ili tvornica Societa istriana el cemento – Pula” (Šimončić-Bobetko, 1983., 97-167.). Prometno provezana mjesta čini su se atraktivnim i ekonomski isplativim za plasman robe i proizvoda, međutim tada još nije nastala spoznaja o brojnim štetnostima zagađenjem zraka povećanim prometom i proizvodnje na fosilna goriva te njihov dugoročni utjecaj na okoliš. Važnije je bilo zadovoljavanje potreba društva i ostvarivanje profita, te zadžavanje trendova sa svijetom. Treba istaknuti kako kako su cementare najveći potrošači električne energije tik uz brodogradnju, te kako u procesu proizvodnje koriste značajne količine goriva. U vrijeme razvoja koristili su ugljen kao tehnološko gorivo koji dovozili iz rudnika u unutar zemlje kao i uvozom. Od svih fosilnih goriva ugljen ima najvišu upotrebu kroz povijest, a najvažnije su proizvodnja čelika i električne energije. Za proizvodnju čelika u svijetu potrebno je oko 70% ugljena, dok za proizvodnju primarne energije je potrebno 23% a električne energije u svijetu oko 38%.

Osim industrija, kako se već spomenulo, neki od glavnih izvora ugljičnog dioksida su i emisije ispušnih plinova automobila te ostalih vrsta vozila na dizel i benzin. Pri sagorijevanju benzina, koji je dobiven od nafte, dolazi do kemijskih spojeva, koji se uglavnom sastoje od ugljičnog monoksida, dušikovih oksida, ugljikovodika i spojeva olova. Ispušni sustav dizelskih motora sadrži čađu, koja dovodi do stvaranja smoga, neizgorenih ugljikovodika, dušikovih oksida i sumpornog anhidrida. Takvi spojevi dovode do porasta temperature i to na globalnoj razini te uzorkuju ozbiljne bolesti opasne po život stanovnika u velikim gradovima a uz meteorološke uvjete širenjem na ruralna područja te posebno na poljoprivredna zemljišta.

Sa ekološke strane, ugljen je najopasniji izvor energije jer je najvećim dijelom sačinjen od ugljika i vodika, te u sebi ima nečistoće koje se ispuštaju u atmosferu, te njegovim izgaranjem nastaje ugljični dioksid CO₂. Ugljični dioksid je plin bez boje i mirisa, otrovan, a u atmosferi je jedan od stakleničnih plinova. Znanstvena istraživanja pokazuju kako je upravo globalno povećanje temperature uzrokovano upravo otpuštanjem ugljičnog dioksida u atmosferu. Ipak ne treba se zanemariti njegov značajni doprinos u zagrijavanju površine Zemlje, inače bi sve bilo pokriveno ledom. Funkcionira na način da kao jedan stakleničkih plinova propušta

kratkovalno zračenje (ultraljubičasto svjetlo vidljivo) sunčeve svjetlosti koja prolazi kroz atmosferu. Zatim Zemlja apsorbira energiju i zrači natrag u obliku dugovalnog zračenja, dok staklenički plinovi upijaju tu energiju i zagrijavaju zemlju. Čovjek je eksploatacijom ugljena te neefikasnim procesom sagorjevanja izazvao globalno zatopljenje i onečišćenje zraka. Iako danas postoje tehnologije koje mogu pročitati 99% sitnih čestica i ukloniti 95% nečistoća što ih prouzrokuju kisele kiše. Također zbog problema s kiselim kišama Vlade SAD-a i Kanade su financirale razvoj tehnologija koje mogu smanjiti emisiju ugljičnog dioksida u atmosferu efikasnijim sagorjevanjem ugljena.

Danas u Hrvatskoj postoji industrijska proizvodnja koja zauzima značajni dio u ukupnoj proizvodnji a to su prerađivačka, petrokemijska i metalprerađivačka industrija. Postoje nekoliko grana industrija među kojima najbrojnija prehrambena i to prema ukupnom prihodu, vodeće su proizvodnja hrane, pića i duhana. a slijede kemijska i naftna industrija, te tekstilna, metalna, drvna, kožarska te neka velika brodogradilišta. Također značajna je proizvodnja u građevinskom sektoru i energetici. Prema Državnom zavodu za statistiku u 2021.g. sa najvećim udjelom u proizvodnji od oko 81% najzastupljenija je prerađivačka industrija, dok je izvoz strojeva i prijevoznih sredstava iznosio 22%, prehrambenih proizvoda 9,5%, električne opreme 8%, gotovih metalnih proizvoda 8,0%, , kemijskih proizvoda 6,9%, farmaceutskih proizvoda 6,6%, koksa i naftnih prerađevina 5,0%, drva i drvnih prerađevina 5,9%, metala 5,2%, odjeće 4,6%, proizvoda od gume i plastike 4,0%, mineralnih proizvoda 3,9%, računala i računalne opreme 3,7%, kože 2,8%, papira 2,3%, , namještaja 1,8%, tekstila 1,2%, pića 1,1%, te duhanskih proizvoda 1,0%.

U Izvješću pučke braniteljice RH za 2021.g. prijavljuju se i istražuju najveći zagađivači zraka i okoliša u koje se ubrajaju naftne rafinerije, ostaci tvornice ferolegura uz more, te opasnih nesaniranih tj. propisno nezbrinuti otpad kao npr. nakon obrada silikomanganske troske. Posebno zabrinjava broj prijavi nepropisnih odlagališta otpada (građevinski, biootpad, otpadi kućanstva i ostalo). “Prema podacima DIRH-a, 2021. godine inspekcija zaštite okoliša u sedam slučajeva je utvrdila neprovođenje mjera zaštite propisanih rješenjem o prihvatljivosti zahvata za okoliš. Proveden je jedan nadzor nad ovlaštenicima za obavljanje stručnih poslova zaštite okoliša i četiri kontrolna nadzora te je upućeno 16 zahtjeva za mjerenja posebne namjene iz Zakona o zaštiti zraka. Nismo zaprimili podatke o provedenim nadzorima na lokacijama crnih točaka i bivših tvornica koje su u postupku sanacije” (Izvješće pučke pravobraniteljice, 2021., 150.).

Kao rezultat zagađenja od opasnog nesanimiranog otpada na adresu pučke pravobraniteljice pristižu svakodnevno prijave zbog zagađenja mora. Tako se u izvješću se navodi kako se postupalo u prijavljenim slučajevima. “Bilježimo porast pritužbi zbog onečišćenja morskog okoliša i priobalja, dok je inspekcija zaštite okoliša u 2021. godini evidentirala jedno onečišćenje mora ugljikovodicima kojemu nije utvrđen uzrok. Prema podacima MMPI-a lučke kapetanije, ustrojene unutar Uprave sigurnosti prometa koja je zadužena za kontinuiranu provedbu mjera zaštite morskog okoliša od štetnih utjecaja pomorskog prometa, u 2021. godini su postupale u 55 slučajeva po dojavama o onečišćenju mora, od kojih je onečišćenje utvrđeno i sanirano u 10 slučajeva, pri čemu za njih sedam nije poznat izvor onečišćenja” (Izvješće pučke pravobraniteljice, 2021, 152.). Također zabrinjavajuća činjenica je kako jedno odlagalište čini bolesnima stanovnike tog mjesta. “Prema podacima MGOR-a, 94 odlagališta su aktivna. Ona koja nastavljaju s radom do otvaranja CGO-a su sanirana i opremljena, a sanacija je u tijeku ili je završena na više od 75% odlagališta. Nije do kraja riješeno pitanje goriva iz otpada u postojećim CGO-ima, koje se prema podacima MGOR-a izvozi uz 40%-tno sufinanciranje FZOEU-a. Od kraja 2018. godine vodimo ispitni postupak zbog utjecaja CGO-a Marišćina na okoliš i zdravlje, u kojemu je i dalje ljeti povećana količina otpada, brzo se pune odlagališne plohe te se otpad nedovoljno odvaja” (Izvješće pučke pravobraniteljice, 2021.,153-154.).

Osim štetnih procesa koji se događaju u raznim tvornicama, na ostacima tvornica i nepropisnih odlagališta, na kvalitetu zraka i okoliša najviše utječe štetni plinovi prijevoznih sredstva a posebno u velikim gradovima sa većim brojem stanovništava. Kako bi bolje upoznali područje primjene ovog rada, potrebno je objasniti koji su to štetne tvari koje nastaju u prometu čije se dnevne i satne razine prate mjerne stanice na određenim područjima u Hrvatskoj. Štetni procesi koji se događaju u prometu, a naročito u velikim gradovima i industrijskim središtima, su dim, čađa i smog. Metrološka mjerenja su pokazala u svojim izvješćima kao čak 60% sumpornog dioksida se prenosi zrakom i na druga manja razvijenija mjesta pa čak i na susjedne države.

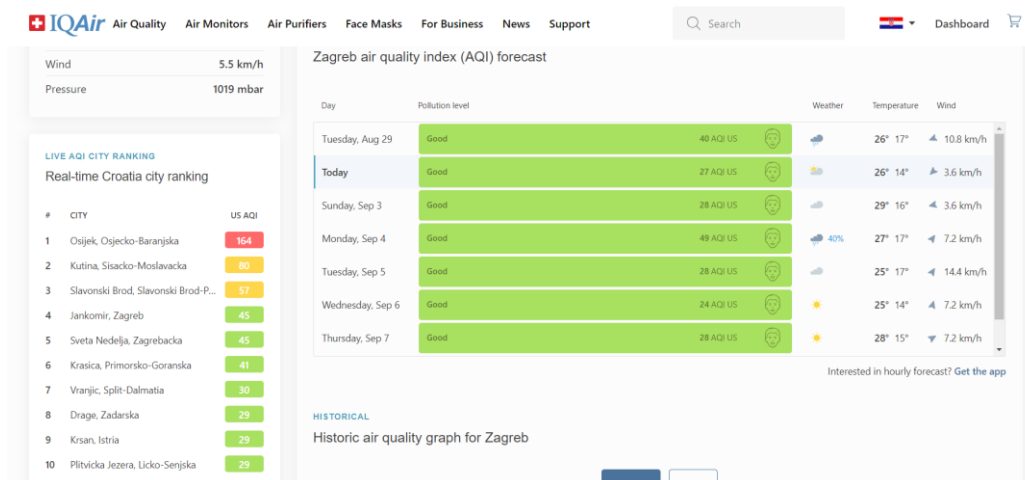
Na platformi za praćenje kvalitete zraku u svijetu u stvarnom vremenu AirVisual, švicarske tehnološke tvrtke IQAir specijalizirane za proizvodnju proizvoda i filtera za zaštitu zraka, mogu se pratiti na dnevne i satne koncentracije lebdećih čestica PM_{2.5} i PM₁₀, sumporov dioksid SO₂, NO₂ dušikov dioksid, te O₃ prizemni ozon. Podatke sa više od 80000 tisuća lokacija koje prikuplja od vladinih mjernih postaja, private tvrtke i praćenja sa jednostavnijih senzora koje bilježu zainteresirani građani, te ekološke udruge. Na mapi su Hrvatske prikazana je svaka mjerna postaja od svakog posebnog subjekta od koji se prikupljaju podaci. Osim onečišćujućih tvari u zraku, od državnih mjernih postaja kao što je Državni metrološki zavod se prikupljaju

meteorološki podaci kao smjer vjetra, temperature, vrijeme koje utječu na širenje štetnih tvari i lebdećih čestica u zraku. Platforma IQAir posebno izvješćuje državne postaje o razinama prizemnog ozona O₃ koji je posebno opasan za zdravlje ljude. Tako se na platformi mogu pratiti navedene koncentracije prema rangu velikih gradova sa najzagađenijim prema najčistijim zrakom na svijetu.

“Koristeći web stranicu ili aplikaciju za kvalitetu zraka (Android ili iOS), možete se pomaknuti prema dolje i pogledati cijeli popis odabranih gradova, poredanih prema AQI-u po satu i dnevno. Trenutačni popis se usredotočio na velike gradove s populacijom od >300 000 ljudi i koji predstavljaju širok raspon zemalja kako bi se omogućio globalni kontrast”. (IQ Air platforma dostupno na <https://www.iqair.com/world-air-quality-ranking> pristupano 03.09.2023).

Na platformi (Slika 1.) se može pratiti Zagreb i županije u Hrvatskoj, kao i pojedinačna mjesta, unosom na tražilici mjesta u državi, čime će odmah povezati na mjerne stanice koje lokacijom odgovoraju traženom mjestu. Kako se na platformi prate kvaliteta zraka u stvarnom vremenu moguće je dobiti odmah rezultate za odabrane gradove, mjesta i države.

Slika 1. Prikaz kvalitete zraka za Zagreb

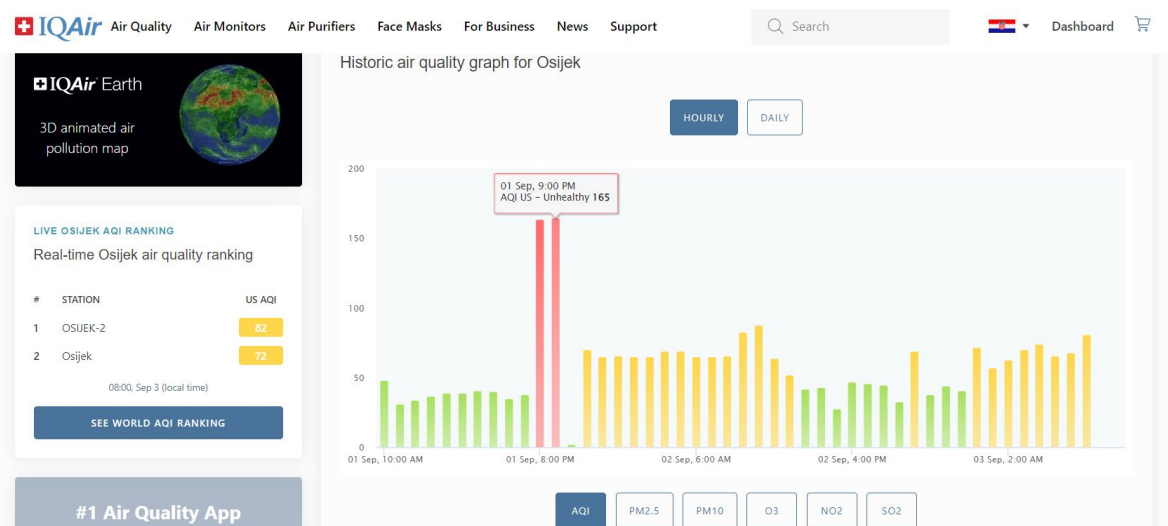


Izvor: IQAir 2023.

Također na istom prozoru može se usporediti Zagreb sa popisom županija gdje se nalaze mjerne postaje, pa se tako može vidjeti kako je najlošija kvaliteta zraka u Osječko-baranjskoj prema ukupnim štetnim tvarima u zraku na dan, 01.09.2023. Povijesni podaci mogu se pratiti

na stupičastom grafu i po nekoliko dana unazad po dnevnoj i satnoj razini, te za svakog zagađivača i ukupno. Na slici 2. su prikazana ukupna kvaliteta zraka za Osijek na satnoj razini.

Slika 2. Prikaz praćenje kvaliteta zraka za Osijek



Izvor: IQAir 2023.

Indeksi kvalitete zraka AQI se sastoji od 6 zagađivača zraka:

- PM2.5
- PM10
- Ugljikov monoksid CO
- Sumporov dioksid SO₂,
- Dušikov dioksid NO₂
- Prizemni ozon O₃

Najštetnije za kvalitetu zraka jesu PM2.5, PM10 engl. PM (Particulate Matter) što predstavlja lebdeće čestice. Osim na kvalitetu zraka utječu i na zrak u prostorima u kojem borave ljudi. Određena razina čestica prašine u atmosferi je prirodna, ali problematično postaje kad se njihova koncentracija u zraku povećana. Prirodne prisutne čestice u zraku su rezultat požara, vulkanske aktivnosti ili erozije, dok primarni izvor sitnih čestica prašine je posljedica ljudske aktivnosti kao što je spaljivanje opasnih otpada, tvornička i industrijska postrojenja prerada metala, rudarstvo te za ovaj rad nježnačajniji, izgaranje fosilnih goriva.

Ugljikov monoksid kemijski je spoj jednog atoma ugljika i jednog atoma kisika, formulom izraženo CO. To je plin bez boje i mirisa koji nastaje nepotpunim izgaranjem tvari u kojima ima ugljika, te je jako otrovan jer se veže na hemoglobin u krvi čime sprječava prijenos kisika (često zvan "tihi ubojica").

Sumporov (IV) oksid (sumporov dioksid, SO₂) pri sobnoj temperaturi je bezbojan, otrovan i nadražujući plin neugodna, oštra i bockava mirisa, koji podražuje na kašalj, a nastaje izgaranjem sumpora, sumporovodika, te prženjem sulfidnih ruda. Bitno je pratiti 24 satne koncentracije.

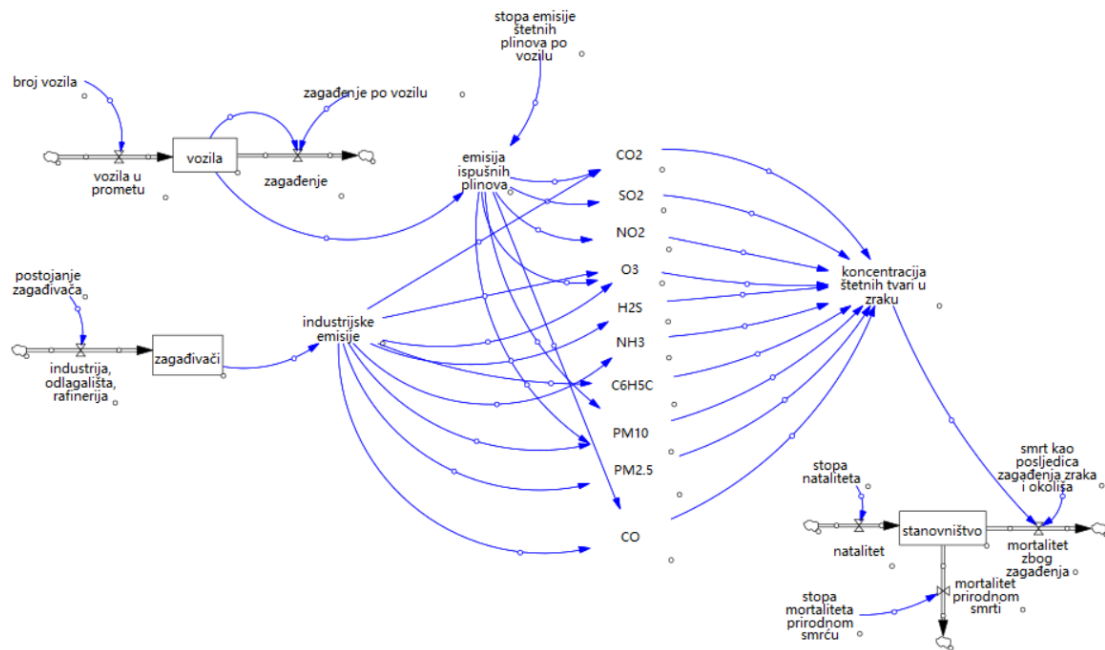
Dušik dioksid je kemijski spoj dušika i kisika formule NO₂. Ističe se jer je crveno-smeđi plin, otrovan i ima karakterističan oštri miris, sličan kloru. Kao glavni izvor dušikovih oksida je transport, zagađenje je povezano sa prometnicima i gusto naseljenim područijama. Osobito visoki koncentracije mogu se očekivati u hladnim zimskim danima sa malo vjetra. Općenito, dušikovi oksidi zagađuju zrak u gradovima, gdje nastaju u automobilskim motorima spajanjem kisika i dušika (oko 2 g po kilometru prijeđenog puta). Utječe na zdravlje tako što izaziva slabljenje imunološkog sistema, povećane alergijske reakcije, izaziva kašalj i bronhitis. Asmatičari mogu reagirati nakon kratkog izlaganja što dodatno utječe na oštećenje pluća.

Ozon (O₃) je snažan oksidirajući kemijski spoj. Simptomi izloženosti koji se pojavljuju uključuju: kašalj, iritaciju grla, bol prilikom dubokog udisaja zraka, bol u prsima, te glavobolju i mučninu. Ovo se posebno odnosi na djecu, osobe s bolestima pluća, starije osobe i ljudi koji obavljaju različite poslove na otvorenom, mogu biti posebno osjetljivi na ozon. Prosječna godišnja u seoskim sredinama se kreće oko 5 µg/m³, a u gradovima od 20 do 90 µg/m³. Prema preporukama SZO, prosječna godišnja koncentracija azot oksida ne bi trebala prelaziti 30 µg/m³.

Ekspanzivnim rastem industrije, koristeći prirodne resurse uz prerađivačke obrade, raste zagađenost zraka i okoliša. Potrebno je istaknuti kako ugljen kao nafta, prirodni plin i nuklearna energija su neobnovljivi izvori energije. Iako obnovljivi(neiscrpn) izvori energije su ekološki prihvatljivi i održivi, potrebno je promijeniti u potpunosti dosadašnji način proizvodnje te stvoriti preduvjete za korištenje energije sunce, vjetra, valova, plime i oseke, biomase, proizvodnja alternativnih "zelenih" goriva.

Postavlja se pitanje je li čovjek u nastojanju društvenog i ekonomskog razvoja a više u svrhu ostvarivanja ekstra profita upotrebom “loše energije” ubrzao uništenje prirodnih resursa koji pružaju potencijal za “dobru” ili zelenu energiju tj. održivu energiju?

Slika 3. Prikaz povezanosti stanovništva, njihovih aktivnosti na stvaranje opasnih razina u zraku



izvor: autorski rad u alatu Vensim

Korištenjem alata Vensim, softvera za simulaciju, prikazala se dinamika i smjer utjecaja broja automobila u prometu na emisije ispušnih plinova, tj. izgaranjem goriva na visokim temperaturama, te utjecaj zagadivača (tvornička postrojenja, rafinerije, odlagališta) na stvaranje industrijske emisije štetnih tvari u zraku koji utječu na smrtnost stanovništva.

Međutim u ovom radu nećemo se baviti posljedicama po zdravlje stanovništva već gore navedenih varijabli za svako mjesto gdje postoji mjerna postaja kao broj stanovništva, broja automobila u prometu, postojanje zagadivača te količine opasnih štetnih tvari u zraku kako bi utvrdili njihovu povezanosti i međuovisnost.

što je mjerenje koncentracija SO₂, NO₂, CO, O₃, H₂S, NH₃, C₆H₆, PM₁₀, PM_{2.5} za svaku mjernu postaju.

Mjerenja se provode sukladno o Programu mjerenja razine onečišćenosti zraka u državnoj mreži za trajno praćenje kvalitete zraka (NN 73/16). Rezultati mjerenja koncentracije sumporovog dioksida (SO₂), dušikovog dioksida (NO₂), ugljikovog monoksida (CO), prizemnog ozona O₃, lebdećih čestica (PM₁₀ i PM_{2.5}), te

sumporovodik (H₂S),

amonijak (NH₃)

benzen (C₆H₆)

u zraku u zonama i aglomeracijama su prikazani u skladu sa Pravilnikom o praćenju kvalitete zraka (NN 72/2021).

Molekula sumporovodika (H₂S) je oblikom slična molekuli vode, ali je polarnost, odnosno razlika elektronegativnosti atoma u molekuli manja, s tim je povezivanje molekule manje, jer je sumpor manje elektronegativan od kisika. Sumporovodik (H₂S) je otrovan i zapaljiv plin bez boje pri sobnoj temperaturi to, neugodna mirisa po pokvarenim jajima upravo zbog vrlo neugodna mirisa, koji se osjeća u koncentracijama mnogo manjim od otrovnih, može ga se na vrijeme otkriti. Manje količine sumporovodika mogu se naći u sirovoj nafti, ali se u prirodnom plinu može naći i do 90%. Najveći industrijski zagađivač koji ga ispušta su naftne rafinerije gdje u postupku hidrodesulfurizacije¹ oslobađa sumpor iz nafte djelovanjem vodika. Nastali sumporovodik se pretvara u elementarni sumpor djelomičnim sagorijevanjem u Clausovom procesu².

Amonijak je spoj dušika i vodika formule NH₃. Pri normalnoj temperaturi i tlaku amonijak je bezbojan plin, oštra, karakteristična mirisa, lakši od zraka, lako topljiv u vodi. Otrovan je i korozivan prema pojedinim materijalima. Jedno od karakterističnih svojstava amonijaka je njegovo direktno reagiranje s kiselinama i formiranje soli. Zbog svojih raznih primjena, amonijak je jedna od anorganskih kemikalija s najvećom proizvodnjom. Postoje na desetine kemijskih postrojenja za proizvodnju amonijaka širom svijeta. Više od 80% proizvedenog

¹ Hidrodesulfurizacija - uklanjanje sumpornih spojeva iz naftnih derivata pomiču vodikom i katalizatorom.

² Clausov postupak (engl. Claus process; njem. Claus-Verfahren), najzastupljeniji postupak uklanjanja sumporovodika iz rafinerijskih i drugih plinskih smjesa pretvorbom u elementarni sumpor. U starijoj je inačici postupka plin koji sadržava sumporovodik izravno izgarao u sumpor prolazeći kroz sloj katalizatora (boksita) u tzv. Clausovoj peći.

amonijaka koristi se za proizvodnju gnojiva za poljoprivredne usjeve. Amonijak se i dalje koristi kao rashladno sredstvo u velikim industrijskim procesima, kao što su proizvodnja leda u velikim količinama i u prehrambenoj industriji. Amonijak je koristan kao određenim u rashladnim sistemima koji koriste temperaturnu razliku. Otkad je upotreba alkalnih halogenida doprinijela smanjenju ozonskog sloja, ponovo se povećava uporaba amonijaka kao rashladnog sredstva. Ponekad se dodaje vodi za piće zajedno s klorom da bi formirao kloramin, dezinfekcijsko sredstvo. Iako se amonijak se stvara putem normalnog metabolizma aminokiselina, toksičan je u velikim koncentracijama. Može dovesti do disfunkcija jetre, kao što se javlja kod ciroze, može prouzročiti povećane koncentracije amonijaka u krvi, hiperamonemiju koja doprinosi zbunjenosti i komi jetrene encefalopatije tj. poremećaj koje oštećuju funkciju mozga kao i ostalim neurološkim bolestima koje su uobičajene za ljude s poremećenim ureinim ciklusom i organskim kiselinama.

Benzen je najjednostavniji aromatski ugljikovodik. Molekulska formula mu je C_6H_6 . Upotrebljava se u kemijskoj industriji kao važno otapalo (otapa masti i ulja, smole, boje, itd.), kao polazna sirovina za dobivanje mnogih aromatskih spojeva. Od benzena se izvodi i velika skupina organskih spojeva (benzenoidni aromatski spojevi). Služi za proizvodnju lijekova, boja, lakova, plastičnih masa, eksploziva, sredstava za pranje, pesticida. Benzen se sve više izbjegava u industriji jer je otrovan i kancerogen je. Upravo zbog njegovog štetnog djelovanja u nekim industrijskim procesima, zamjenjuju ga njegovi derivati koji nisu toliko štetni za ljudsko zdravlje (npr. toluen ili ksilen).

Rezultati pojedinih štetnih tvari u zraku su iskazani kao udio valjanih podataka u ukupnom mogućem broju podataka, a koji je umanjen za vrijeme trajanja redovitog pojedinog uređaja. Rezultati moraju obuhvaćati minimalni udio podataka koji je jednak ili viši od 85% kako bi se mogli vjerodostojno koristiti.

Metoda mjerenja ili mjerna tehnika koja se koristi je u mjerenju kvalitete zraka određuje se prema pojedinačnom onečišćivaču te je propisana prema Pravilniku o praćenju kvalitete zraka (NN 72/2020) U tablici 1. prikazane su referentne metode mjerenja za određivanje koncentracija plinova i lebdećih čestica u zraku.

Tablica 1. Prikaz referentnih metoda mjerenja

Onečišćujuća tvar	Princip mjerne/analitičke metode	Metoda mjerenja
SO₂	UV fluorescencija	HRN EN 14212 – Mjerenje koncentracije sumporova dioksida u zraku ultraljubičastom fluorescencijom (EN 14212)
NO₂	Kemiluminiscencija	HRN EN 14211 – Metoda za mjerenje koncentracije dušikova dioksida i dušikova monoksida u zraku kemiluminiscencijom (EN 14211)
CO	IR spektroskopija	HRN EN 14626 – Mjerenje koncentracije ugljikova monoksida nedisperzivnom infracrvenom spektroskopijom (EN 14626)
O₃	UV apsorpcija	HRN EN 14625 – Mjerenje koncentracije ozona ultraljubičastom fotometrijom (EN 14625)
Benzen	GC-PID ili GC-FID	HRN EN 14662-1 – Mjerenje koncentracije benzena – 1.dio: Uzorkovanje prosisavanjem uz termičku desorpciju i analizu plinskom kromatografijom (EN 14662-1)
		HRN EN 14662-2 – 2.dio: Uzorkovanje prosisavanjem uz desorpciju otapalom i analizu plinskom kromatografijom (EN 14662-2)
		HRN EN 14662-3 – 3.dio: Automatsko uzorkovanje prosisavanjem uz istovremenu analizu plinskom kromatografijom (EN 14662-3)
PM₁₀ i PM_{2,5}	Gravimetrija	HRN EN 12341 – Standardna gravimetrijska mjerna metoda za određivanje masenih koncentracija PM ₁₀ i PM _{2.5} frakcija lebdećih čestica (EN 12341)
H₂S	UV fluorescencija uz prethodno uklanjanje SO _x i konverziju H ₂ S u SO ₂	Kontinuirano mjerenje analizatorom
NH₃	Mjerenje koncentracija amonijaka – automatska mjerna metoda – kemiluminiscencija uz prethodnu konverziju NH ₃ u NO _x	Kontinuirano mjerenje analizatorom

Izvor: Izvješće za 2021. Država mreža za trajno praćenje kvalitete zraka

Primjena modela može slijediti raspoložive smjernice FAIRMODE (Forum for Air Quality Modelling) o referentnim tehnikama modeliranja koje su dostupne na mrežnim stranicama FAIRMODE.

Prema Pravilniku o praćenju kvalitete zraka kod procjene razine kvalitete zraka, može se kod mjerenja koncentracije onečišćujućih tvari u zraku nadomjestiti rezultatima modeliranja kvalitete zraka ili drugih objektivnih metoda ocjenjivanja razine onečišćenosti zraka samo za zone i aglomeracije u kojima razina onečišćenosti onečišćujućih tvari ne prekoračuje donji prag.

Primjenjuju se gornji i donji pragovi za određivanje uvjeta za procjenu koncentracija onečišćujućih tvari u zraku promatrane zone ili aglomeracije s obzirom na zaštitu ljudi,

vegetacije i ekosustava prema Uredbi o razinama onečišćujućih tvari u zraku (NN 77/2020). U tablici 2. prikazani su iznosi gornjih i donjih pragova procjene te njihove dozvoljene učestalosti prema pojedinom onečišćivaču.

Tablica 2. Gornji i donji pragovi procjene koncentracije po onečišćujućoj tvari

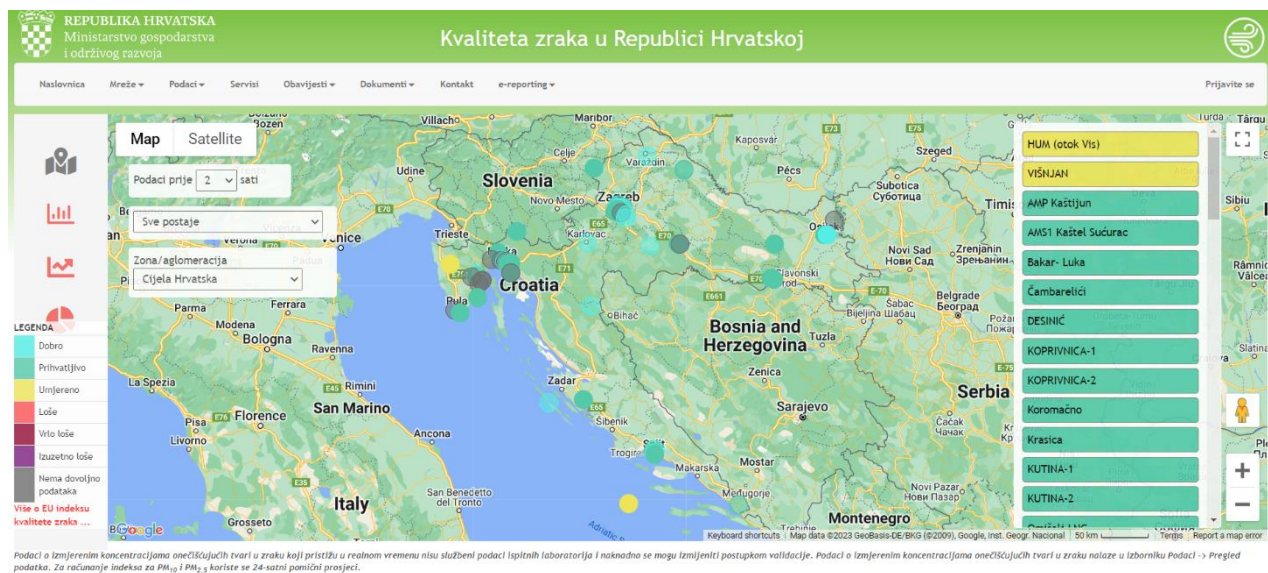
Onečišćujuća tvar	Prag procjene	Razdoblje praćenja	Vrijeme usrednjavanja	Iznos praga procjene	Učestalost dozvoljenih prekoračenja
SO ₂	gornji	kalendarska godina	24 sata	75 µg/m ³ (60% GV)	prag procjene ne smije biti prekoračen više od 3 puta u bilo kojoj kalendarskoj godini
	donji	kalendarska godina	24 sata	50 µg/m ³ (40% GV)	prag procjene ne smije biti prekoračen više od 3 puta u bilo kojoj kalendarskoj godini
NO ₂	gornji	kalendarska godina	1 sat	140 µg/m ³ (70% GV)	prag procjene ne smije biti prekoračen više od 18 puta u bilo kojoj kalendarskoj godini
			1 godina	32 µg/m ³ (80% GV)	
	donji	kalendarska godina	1 sat	100 µg/m ³ (50% GV)	prag procjene ne smije biti prekoračen više od 18 puta u bilo kojoj kalendarskoj godini
			1 godina	26 µg/m ³ (65% GV)	
PM ₁₀	gornji	kalendarska godina	24 sata	35 µg/m ³ (70% GV)	prag procjene ne smije biti prekoračen više od 35 puta u bilo kojoj kalendarskoj godini
			1 godina	28 µg/m ³ (70% GV)	
	donji	kalendarska godina	24 sata	25 µg/m ³ (50% GV)	prag procjene ne smije biti prekoračen više od 35 puta u bilo kojoj kalendarskoj godini
			1 godina	20 µg/m ³ (50% GV)	
PM _{2,5} (6)	gornji	kalendarska godina	1 godina	17 µg/m ³ (70% GV)	-
	donji	kalendarska godina	1 godina	12 µg/m ³ (50% GV)	-
Benzen (C ₆ H ₆)	gornji	kalendarska godina	1 godina	3,5 µg/m ³ (70% GV)	-
	donji	kalendarska godina	1 godina	2 µg/m ³ (40% GV)	-
CO	gornji	kalendarska godina	maksimalna dnevna osmosatna srednja vrijednost	7 mg/m ³ (70% GV)	-
	donji	kalendarska godina	maksimalna dnevna osmosatna srednja vrijednost	5 mg/m ³ (50% GV)	-
SO ₂ Zaštita vegetacije	gornji	kalendarska godina	zimsko razdoblje	zimsko razdoblje (1. listopad do 31. ožujka)	12 µg/m ³ (60% kritične razine za zimsko razdoblje)
	donji	kalendarska godina	zimsko razdoblje	zimsko razdoblje (1. listopad do 31. ožujka)	8 µg/m ³ (40% kritične razine za zimsko razdoblje)
NO _x zaštita vegetacije i	gornji	kalendarska godina		1 godina	24 µg/m ³ (80% kritične razine)

prirodnog ekosustava	donji		1 godina	19,5 µg/m ³ (65% kritične razine)
O₃	gornji	Dnevna 8-satna srednja vrijednost	1 godina	120 µg/m ³ (najviše 25 dana u godini usrednjeno na tri godine)
	donji	od svibnja do srpnja	1 godina	18000 µg/m ³ (prosjeck pet godina, izračunato na temelju 1 satna AOT40)
H₂S	prag	kalendarska godina	1 sat	7 µg/m ³ (GV ne smije biti prekoračena više od 24 puta tijekom kalendarske godine)
			24 sata	5 µg/m ³ (GV ne smije biti prekoračena više od 7 puta tijekom kalendarske godine)
NH₃	prag	kalendarska godina	24 sata	100 µg/m ³ (50% GV) ne smije biti prekoračena 7 puta više u godini

izvor: Izvješće za 2021. Državna mreža za trajno praćenje kvalitete zraka

Na interaktivnoj stranici (Slika 5.) je moguće pratiti pojedine vrijednosti udjela štetnih tvari na bazi unutar 1 (jedan) do 48 (četrdeset i osam) sati. Ne postoji mogućnost praćenje ukupne vizualizacije na temelju izmjerenih vrijednosti svih mjernih mjernih postaja kako bi se napravila usporedba.

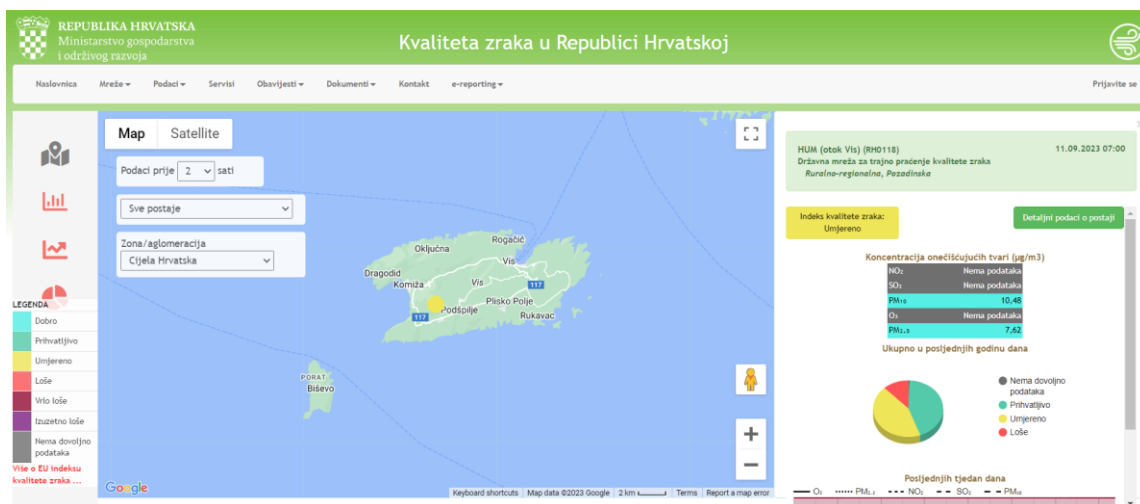
Slika 5. Prikaz naslovne stranice Državne mreže za praćenje kvalitete zraka



izvor: web stranica Državna mreža za trajno praćenje kvalitete zraka, 2023.

Potrebno je višestrukim odabirom na popis postaje doći do podataka. Na nekim mjernim postajama nisu izmjerene sve vrijednosti zbog kvara na uređaju (Slika 6.)

Slika 6. Prikaz izmjerenih podataka za mjernu postaju HUM (Otok Vis)



izvor: web stranica Državna mreža za trajno praćenje kvalitete zraka, 2023.

Ukupne vrijednosti izmjerenih štetnih koncentracija u zraku mogu moguće je vidjeti na kartici (Slika 7.) Podaci koja također po principu višestrukih odabira iz padajućih izbornika prikazuju podatke o koncentraciji onečišivača u zraku za odabrani datum od/do prema izabranoj mjernoj postaji. Nepostojanje podatka o koncentraciji nekog onečišivača označava kako za određenu mjernu postaju u tom radoblju se nisu prikupljali podaci. U padajućem izborniku se mogu izabrati podaci prema vremenu usrednavanja od statnih i dnevnih, izvornih i validiranih podataka.

Slika 7. Prikaz pregleda prikupljenih podataka



izvor: web stranica Državna mreža za trajno praćenje kvalitete zraka, 2023.

2.3. Zakonska podloga praćenja i procjene onečišćenja zraka u RH

Na temelju članka 89. Ustava Republike Hrvatske donesen je Zakon o zaštiti zraka (NN 127/19), koji u svojim odredbama određuje nadležnost i odgovornost za zaštitu zraka, donošenje planova, praćenje i procjenjivanje kvalitete zraka, mjere za sprječavanje i smanjivanje onečišćavanja zraka, izvještavanje o kvaliteti zraka i razmjeni podataka, djelatnost praćenja kvalitete zraka i emisija u zrak, informacijski sustav zaštite zraka, financiranje zaštite zraka, upravni i inspeksijski nadzor. Odredbe koje određuje ovaj Zakon a koje su u skladu sa sljedećim aktima Europske unije:

- Direktiva 2008/50/EZ Europskog parlamenta i Vijeća o kvaliteti okolnog zraka i čistijem zraku za Europu (SL L 152, 11. 6. 2008.),
- Direktiva 2004/107/EZ Europskog parlamenta i Vijeća koja se odnosi na arsen, kadmij, živu, nikal i policikličke aromatske ugljikovodike u vanjskom zraku (SL L 23, 26. 1. 2005.),
- Direktiva Vijeća 2001/81/EZ Europskog parlamenta i Vijeća o nacionalnim vršnim emisijama za određene (pojedine) onečišćujuće tvari (SL L 309, 27. 11. 2001.),
- Direktiva 2003/87/EZ Europskog parlamenta i Vijeća kojom se utvrđuje shema za trgovanje kvotama emisije stakleničkih plinova unutar Zajednice i kojom se izmjenjuje i dopunjuje Direktiva Vijeća 96/61/EZ (SL L 275, 25. 10. 2003.),
- Direktiva 2004/101/EZ Europskog parlamenta i Vijeća kojom se dopunjuje Direktiva 2003/87/EZ kojom se utvrđuje shema za trgovanje kvotama emisije stakleničkih plinova unutar Zajednice, vezano za projektne mehanizme Kyotskog protokola (SL L 338, 13. 11. 2004.),
- Direktiva 2008/101/EZ Europskog parlamenta i Vijeća kojom se izmjenjuje i dopunjuje Direktiva 2003/87/EZ kako bi se zrakoplovne aktivnosti uključile u sustav trgovanja kvotama emisijama stakleničkih plinova unutar Zajednice (SL L 8, 13. 1. 2009.),
- Direktiva 2009/29/EZ Europskog parlamenta i Vijeća kojom se izmjenjuje i dopunjuje Direktiva 2003/87/EZ kako bi se unaprijedio i proširio sustav trgovanja emisijskim jedinicama Zajednice (SL L 140, 5. 6. 2009.),
- Direktiva 2009/30/EZ Europskog parlamenta i Vijeća o izmjenama i dopunama Direktive 98/70/EZ u pogledu specifikacije benzina i dizelskih goriva i plinskog ulja i uvođenju mehanizma praćenja i smanjenja emisija stakleničkih plinova i izmjenama i

dopunama Direktive Vijeća 1999/32/EZ u pogledu specifikacije goriva koje koriste plovila za unutarnju plovidbu i ukidanju Direktive 93/12/EEZ (SL L 140, 5. 6. 2009.),

- Odluka br. 406/2009/EZ Europskog parlamenta i Vijeća s ciljem država članica da smanje svoje emisije stakleničkih plinova kako bi ispunile obveze Zajednice za smanjenje emisija stakleničkih plinova do 2020. (SL L 140, 5. 6. 2009.),
- Odluka Komisije 2011/278/EZ o utvrđivanju prijelaznih pravila na razini Zajednice o usklađenoj besplatnoj dodjeli emisijskih jedinica sukladno članku 10.a Direktive 2003/87/EZ (SL L 130, 17. 5. 2011.),
- Uredba (EZ) br. 1005/2009 Europskog parlamenta i Vijeća o tvarima koje oštećuju ozonski sloj (SL L 286, 31. 10. 2009.),
- Uredba (EZ) br. 842/2006 Europskog parlamenta i Vijeća o određenim fluoriranim stakleničkim plinovima (SL L 161, 14.6.2006.).

Sukladno Zakonu o zaštiti zraka (NN 127/19) između Fonda za zaštitu okoliša i energetske učinkovitost (dalje u tekstu FZOEU) i Državnog hidrometeorološkog zavoda (dalje u tekstu DHMZ), provode se mjerenja kvalitete zraka na postajama Državne mreže za trajno praćenje kvalitete zraka u Republici Hrvatskoj. Uspostavljen je okvir za provedbu ovog zakona kojim se donose Odluke Vijeća za uspostavljanjem uzajamne razmjena informacija i podataka dobivenih od mreža i pojedinačnih postaja koje mjere onečišćenost zraka među državama članicama, te propisuje upitnik za dostavljanje godišnjeg izvješća o procjeni kakvoće zraka. Odluka Komisije 2004/224/EZ kojom se propisuju postupci za dostavljanje informacija o planovima ili programima koji se zahtijevaju prema Direktivi 96/62/EC u odnosu na granične vrijednosti za određene onečišćujuće tvari u vanjskom zraku (SL L 68, 6. 3. 2004.).

3. Uloga i primjena podatkovne znanosti u praćenju kvalitete zraka

3.1. Pojam i razvoj podatkovne znanosti

Jedna od vodećih polja znanosti u medicini, inženjerstvu, ekološkim istraživanja a posebno u današnjim poslovnom okruženju i globalnom društvu je podatkovna znanost. Poduzeća žele unaprijediti korisničko iskustvo, digitalizirati poslovni proces prodaje koristeći velike skupove podataka kako bi iz njih dobili znanje te na kraju unaprijedili ponudu usluga te povećali prodaju. Organizacije koriste podatke kako bi predvidjeli, planirali i organizirali poslovanje te upravljali ograničenim resursima. Utjecaj i primjena podatkovne znanosti se može vidjeti u svakodnevnom životu pojedinca kroz prijedloge na društvenim mrežama, reklame, preporučeni filmovi, knjige, predstave te ostalo čiji algoritmi uče iz naših prijašnjih odabira i preferencija.

U svojoj knjizi autori John D. Kelleher i Brenden Tierney (2018) daju definiciju podatkovne znanosti kao znanosti koja obuhvaća skup načela, definicija, algoritama i procesa za izvlačenje obrazaca iz velikih skupova podataka a koji su prepoznati kao korisni. Dalje navode kako su elementi podatkovne znanosti su strojno učenje (engl. Machine Learning) i rudarenje podataka (engl. Data Mining). Stoga, može se zaključiti kako podatkovna znanost ili znanost o podacima (engl. Data Science) kao multisciplinarno područje obuhvaća različite domene istraživanja i primjene, kao što matematika, statistika, dubinska analiza podataka, umjetnu inteligenciju. Strojno učenje kao jedna od grana umjetne inteligencije (AI) obuhvaća izvođenje zadatka pomoću algoritama koji mogu otkriti obrasce u podacima i učiti iz njih, učenje koje može biti nadzirano (engl. Supervised Learning) i nenadzirano (engl. Unsupervised Learning) te se na temelju toga mogu dobiti predviđanja za donošenje ključnih odluka. Tehnike rudarenje podataka podrazumijevu procese pronalaženja anomalija, obrazaca i korelacija unutar velikih skupova podatka radi predviđanja ishoda.

Termin podatkovna znanost (engl. Data Science) prvi put je korišten početkom 1960-tih godina prošlog stoljeća kako bi opisao novo područje primjene za razumijevanje i interpretaciju velike količine podataka s obzirom na svakodnevni rast količine podataka. Podatkovna znanost se tada koristila u svrhu obrade velike količine podataka u medicini i astronomiji. Statistika i statistički modeli su duboko ukorijenjeni u područje podatkovne znanosti. Sa razvojem interneta i Internet stvari (engl. Internet of Things) dovelo je do eksponencijalnog rasta količine

podataka. Uz takvo rastujuće generiranje podataka došlo je do razvoja podatkovne znanosti kakvu danas poznajemo, te uz statistiku, implementirani su koncepti umjetne inteligencije kao što je strojno učenje. Prva obrada velikih skupova podataka u poslovnom svijetu su bili podaci o navikama kupaca i trendovima. Nakon toga, primjena tehnika i metoda, te analiza i predviđanje čini podatkovnu znanosti neizbježnu kod donošenja važnih odluka za poslovanje. Pojedini autori kroz povijest poput J. W. Tukey (1962) ističu budućnost razvoj analize podataka uz korištenje termina Data Science kao spajanje statistike i računalna jer su se računala u 60-tim 20 st. koristila uglavnom za rješavanje matematičkih problema, dok se statistika računala ručno. Godine 1977. osnovana je internacionalna asocijacija za statističku obradu podataka na računalu, poznata kao IASC (engl. International Association for Statistical Computing) koja je u opisu svoje misije prva spomenula frazu pretvaranja podataka u informacije i znanje. Prva konferencija na temu otkrivanja znanja i rudarenja podatka naziva Knowledge Discovery in Database u organizaciji ACM SIGKDD održana je 1989. Prva naslovnica poslovnog časopisa Business Week na temu otkrivanja znanja iz podataka marketinških kompanija naziva Database Marketing je izašla 1994. Krajem desetljeća javlja se potreba za razvojem novog alata koji će brže i preciznije obraditi ogromne generirane podatke koji kotinuirano rastu. Tako J. Zahavi (1999) u svojem istraživanju Mining Data for Nuggets of Knowledge piše kako je skalabilnost veliki problem u rudarenju podataka. Opisuje kako se konvencionalna statistika može primijeniti samo na malom skupu podataka, ali kako današnje baze podataka mogu pohranjivati milijune redaka podataka sa raznim tipovima varijabla. U tu svrhu potreban efikasniji model za obradu i analiza podataka, koji će prepoznati obrasce i ne-linearne veze te povezanosti između pojedinih varijabla podataka. Potrebno je razviti napredniji alat za rudarenje podataka u stvarnom vremenu na temelju kojeg će se moći donositi odluke.

U 2001.g. razvijena je usluga Software-as-a-Service (SaaS). Kreiranjem ove usluge omogućilo se korištenje aplikacija u infrastrukturi oblaka. Iste godine W.S. Cleveland (2001) u svom radu Data Science: An Action Plan for Expanding the Technical Areas of the field of Statistic opisuje kako razviti znanje o tehničkoj primjeni statistike za analizu podataka. Fokus se stavlja na podatkovnog analitičara i njegovo korištenje. Plan je bio postaviti 6 (šest) glavnih područja koji će se proučavati na fakultetima kao posebna područje izučavanja tehničke primjene statistike u podatkovnoj znanosti prema sljedećim udjelima:

- 25% Multidisciplinirano istraživanje (podrazumijeva kolaboraciju tj. suradnju u analizi podataka sa ostalim sličnim područjima istraživanja)

- 20% Modeli i metode za podatke (obuhvaća statističke metode, tehnike modeliranja, modele probabilističke distribucije i estimacije.
- 15% Računalna obrada podataka (HW, SW, računalni algoritmi)
- 15% Plan obrazovanja (nastavni plan i pristup učenja za podučavanje od nižih razreda osnovne škole do držanja seminara za poslovne korisnike)
- 5% Ocjena izvođenja studija (istraživanje uspješnosti u praksi, upitnici o precipiranom potrebi za znanjem i studije na temu razvoja novih alata)

Nakon što je Plan izašao objavljen je u Data Science Journal. U 2006. g. prva nerelacijska baza Hadoop 0.1.0, besplatnog koda je omogućena za korištenje. Hadoop nerelacijska baza podataka je bila povezna na drugu također otvorenu za korištenju bazu podataka, Nutch. Ovim su riješeni neki od ključnih nedostataka relacijskih orijentiranih baza podataka kao što su neučinkovitost pohrane velike količine podataka kao što su Veliki podaci (engl. Big Data), obrada nestandardnih oblika podataka i određne složene vrste podataka. Apache Hadoop je biblioteka besplatnog koda koja omogućuje pohranu i obradu Velikih podataka.

Sve žvučnije postaje naziv podatkovni znanstvenik (engl. Data scientist) te u 2008. postaje prihvaćen kao dio standard za određenu vrstu posla. U 2009. naziv NoSQL ponovno se predstavlja na organiziranoj diskusiji na temu Open –source, non-relation database. U 2012. potražnja za zanimanje podatkovnog znanstvenika su se povećala za 15%. Dolazi do češćeg organiziranja seminara i konferencija na temu Data Science i Big Data. Tehnologija Data Lake dolazi više do izražaja i njezina prednost u odnosu na skladište podataka (engl. Data warehouse) u pogledu pohrane Velikih podataka, uštede vremena i troškova održavanja, te mogućnost korištenja nerelacijskih baza podataka (NoSQL). Ovo je značajno zbog činjenice koju je objavio IBM u 2013. kako čak 90% podataka u svijetu je stvorena u zadnje 2 (dvije) godine. Prva upotreba tehnika dubokog učenja počela je u 2015., razvojem Googleove glasovna tražilica, aplikacija Google voice, što je za rezultat imalo 49% poboljšanja performance tražilice.

Danas je podatkovna znanost važno područje poslovnog okruženja i akademskog istraživanja te obuhvaća područja istraživanja kao procese strojnog prevođenja i prepoznavanja glasa, robotiku, performanse tražilica te digitalnu ekonomiju. Zadnjih godina godina njezina se primjena proširila na proučavanje pojava, utjecaja i povezanosti prirodnih znanosti kao što je biologija, zatim na područje medicine i informacijskih procesa u medicini, društvenih znanosti.

Podatkovna znanost danas utječe na svakodnevne odluke koje se donose u ekonomiji, financijama, državnim i regulatornim tijelima.

3.2. Područja primjene podatkovne znanosti u meteorologiji i hidrologiji

Meteorologija i hidrologija su dva područja koja su u posljednjih desetljeća često istraživana zbog velikih klimatskih promjena. Posljedice klimatskih promjena osjećaju se u svim dijelovima svijeta i sve snažnije pogađaju određene dijelove zemlje gdje dolazi do ekstremnih vremenskih uvjeta poput poplava, suša koja ostavljaju razarajuće posljedice za stanovništvo, bioraznolikost i gospodarstvo zemlje općenito. Klimatske promjene su ozbiljna prijetnja čovječanstvu i cijelom ekosustvu. Uzrok klimatskih promjena je ljudski faktor. Meteorologija je znanost koja se bavi proučavanjem pojava u zemljinoj atmosferi i promjena vremenskih prilika oko nas. Zrak je jedan od osnovnih uvjeta života a sastoji se od mješavine plinova koji tvore Zemljinu atmosferu, potreban prije svega za disanje i fotosintezu. Suhi zrak sadrži (po obujmu) 78,08% dušika, 20,95% kisika, 0,93% argona i ostalih plinova u manjim količinama, kao što su staklenički plinovi, od kojih su najznačajniji vodena para, ugljikovog dioksid, metan, dušikovi oksidi i ozon. Kada dolazi do neravnoteže plinova u zraku zbog ljudske aktivnosti javlja se zagađenje koje ugrožava kvalitetu zraka.

S druge strane, hidrologija je znanost koja proučava vodu te njezina biološka i kemijska, svojstva i pojave u njoj. Voda je neophodan resurs i također jedan od osnovnih uvjeta na zemlji. Sva živa bića na zemlji se sastoje najvećim dijelom od vode (maseni udio vode do 90%). Vodom je prekriveno oko 70% zemljine površine. Ista ta voda neprekidno isparuje u atmosferu tj. kruži u prirodi pod utjecajem sunčeva zračenja gdje se kondenzira stvarajući kišne ili snježne oblake u atmosferi. Nakon toga vraća se na zemlju u obliku oborina (kiša, snijeg, tuša, rosa, inje i magle). Time uvelike utječe na klimu.

Iz toga razloga, došlo je do povezivanja dvaju grana geofizike pod nazivom hidrometeorologija kako bi se proučavala njihova međusobna veza i mogućnosti primjena u inženjerstvu kao što su kontrole poplava, navodnjavanje te proizvodnu električne energije kao održivog načina života. Proučavanje utjecaja klime na klimatsko okruženje predstavlja izuzetno i kompleksno područje istraživanja. Brojna su akademska i znanstvena istraživanja stručnjaka na području podatkovne znanosti koji se bave utjecajima na štetnih plinova i čestica u zraku na klimatološke promjene, zdravlje ljudi i onečišćavanje okoliša. U prvim redovima znanstvenici i stručnjaci se

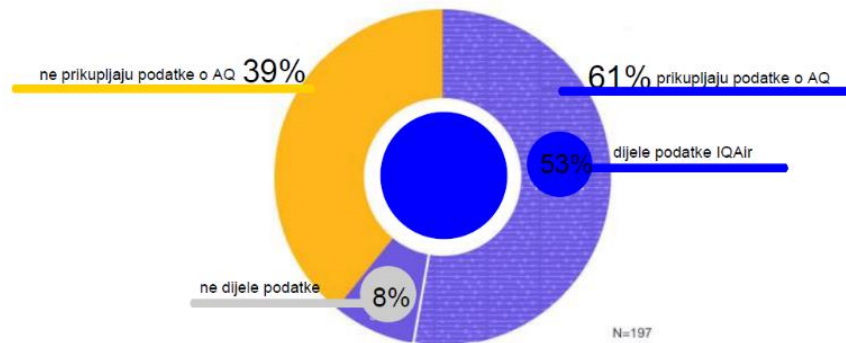
susreću se sa velikim skupovima prikupljenim podataka sa raznih uređaja i senzora za mjerenje, što dovodi do potrebe korištenja sve opsežnijih i kompleksnijih modela i alata za predviđanje budućih scenarija. Primjenjuju se alati umjetne inteligencije, metode strojnog učenja te neuronske mreže kako bi se dobili što precizniji rezultati predviđanja.

Primjena podatkovne znanosti i podatkovne analitike u objema granama se provodi se najčešće kroz razne alate za predikciju, vizualizaciju i analizu prikupljenih podataka, uporabom raznih algoritama i statističkih metoda te alata, tehnologija kako bi predvidjeli buduća događanja. Postoje brojni znanstveni radovi koji su značajno utjecali na svijest i poduzimanja akcija u zaštiti kvalitete zraka i zdravlja ljudi.

3.3. Specifičnosti primjene podatkovne znanosti u prikupljanju i obradi podataka o kvaliteti zraka

Pojavom globalnog zagađenja zraka i okoliša, došlo je do povećane potrebe prikupljanja i praćenja podataka o kakvoći zraka. Kako bi se probudila svijest o posljedicama zagađenja zraka i okoliša, podatke se želi prikazati na način da budu dostupni i vidljivi svima, bilo kad i bilo gdje. Posljednih godina došlo je do stvaranja velikog broja državnih, regionalnih te globalnih otvorenih platformi koje prikupljaju i dijele podatke, analiziraju i vizualiziraju te predviđaju indeks kvalitete zraka. Utjecaj onečišćenja zraka i okoliša izravno utječe na gospodstvo zemalje, na standard života i kvalitetu zdravlja njezinih građana, pa čak međudržavne odnose jer zrak ne poznaje granice. Otvoreni podaci o kvaliteti zraka otvaraju vrata za unaprijeđenje platformi kako bi se donijele odluke za poduzimanje akcija u očuvanju čistog zraka i okoliša te zdravlje ljudi kao i podizanja svijesti na globalnoj razini općenito. U prilog tome govori činjenica kako od ukupno 197 država u svijetu, 61% država mjere kvalitetu zraka na svakodnevnoj bazi, dok samo 53% država dijele podatke o mjerenjima kvalitete zraka svoje zemlje (Slika 8.)

Slika 8. Prikaz udjela u mjerenju i dijeljenju podataka o kvaliteti zraka



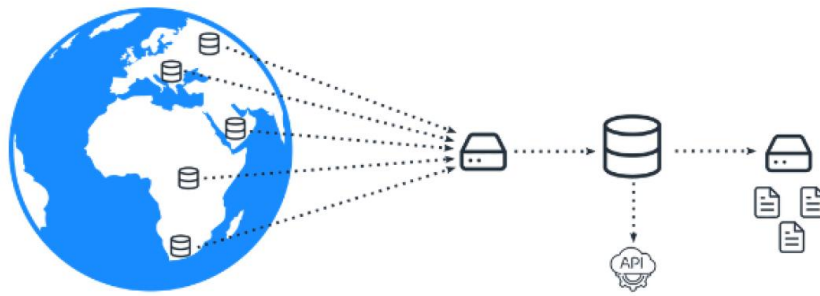
izvor: autorski rad prema OpenQA

Specifična primjena podatakovne znanosti ogleda se u brojnim mogućnostima prikupljanju podataka u stvarnom vremenu sa svih dostupnih izvora kao što su od službenih mjernih postaja, jeftini pametni uređaji kao što su senzori, IoT do algoritama raznih društvenih mreža koji na objavama svojih korisnika bilježe geolokacijske karakteristike po vremenskim serijama. Kolaboracija i dijeljenje podataka o kvaliteti zraka sa ostalim otvorenim i državnim platformama za praćenje kvalitete zraka, sa organizacijama i udrugama za zaštitu zraka i okoliša, korporativnim partnerima, zainteresiranim građanima sa njihovih pametnih uređaja i ostalim. Predikacija na temelju izvornih i agregiranih podataka, traženje povezanosti varijabli pomoću algoritama strojnog učenja, tehikama rudarenja podataka, umjetne inteligencije. Na kraju specifična primjena se ogleda i u vizualizaciji podataka i podatkovna analiza koje doprinose većoj preglednosti i usporedbnoj analizi podataka na jednom mjestu čim pomažu u donošenju mjera i akcija u sprečavanju daljnjeg zagađenja zraka.

Najpoznatija i najveća globalna otvorena platforma za vizualizaciju i praćenje kvalitete zraka u stvarnom vremenu je OpenAQ (Slika9.), koja koristi alate ETL-a³ kako bi dohvaćala podatke o kvaliteti zraka iz različitih izvora diljem svijeta te omogućio pristup i prezentaciju tim istim izvorima na jednom mjestu. Na kraju pohranjuje podatke iz različitih izvora u skladište podataka.

³ Skladište podataka puni se iz različitih izvora, a taj proces naziva se ETL (eng. Extract-Transform-Load) te se odnosi na izvlačenje podataka iz izvora, transformaciju tih podataka te unošenje u skladište podataka (Ćurko i Španić Kezan, 2016)

Slika 9. Prikaz ilustracije rada platforme OpenAQ



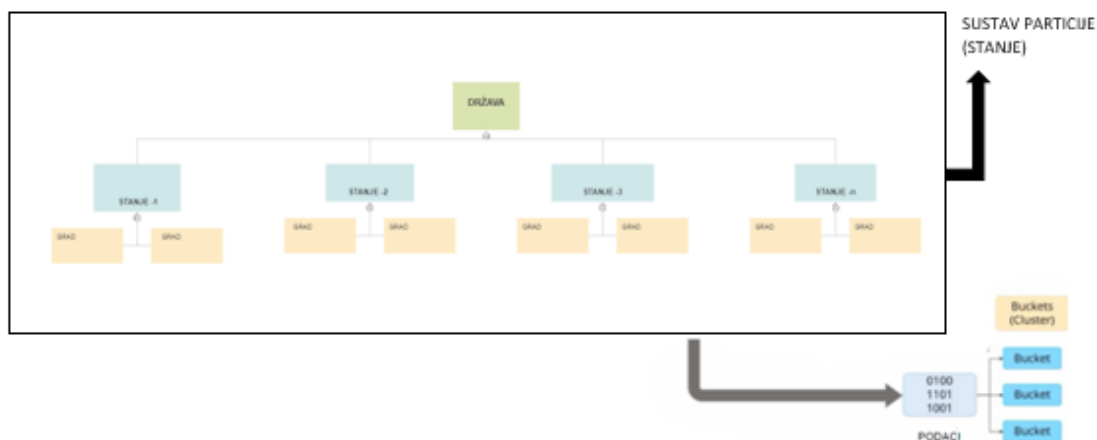
Izvor: autorski rad prema OpenAQ

Putem usluge AWS Athena koji je interaktivni ETL alat nudi pristup bazama podataka putem distribuirane SQL (engl. Structured Query Language) tražilice, postavljanja složenih i kompleksnih SQL upita na baze te kreiranja izvještaja za dobivanje seta podataka o određenim mjerenjima za određenu lokaciju. Primjer postavljanja takvog upita je: `SELECT parameter, value, from_iso8601_timestamp(datetime) AS datetime FROM openaqMeasurements WHERE locationid='2178' AND provider='airnow' AND year='2022' AND month='12'`. Rezultat takvog upita bi bio:

```
"parameter","value","datetime"
"pm10","30.0","2022-12-12 01:00:00.000 -07:00"
"pm10","18.0","2022-12-12 02:00:00.000 -07:00"
"pm10","17.0","2022-12-12 03:00:00.000 -07:00"
"pm10","10.0","2022-12-12 04:00:00.000 -07:00"
"pm10","10.0","2022-12-12 05:00:00.000 -07:00"
"pm10","26.0","2022-12-12 06:00:00.000 -07:00"
"pm10","13.0","2022-12-12 07:00:00.000 -07:00"
"pm10","13.0","2022-12-12 08:00:00.000 -07:00"
"pm10","11.0","2022-12-12 09:00:00.000 -07:00"
"pm10","16.0","2022-12-12 10:00:00.000 -07:00"
"pm10","16.0","2022-12-12 11:00:00.000 -07:00"
"pm10","24.0","2022-12-12 12:00:00.000 -07:00"
"pm10","11.0","2022-12-12 13:00:00.000 -07:00"
"pm10","19.0","2022-12-12 14:00:00.000 -07:00"
"pm10","12.0","2022-12-12 15:00:00.000 -07:00"
...
```

Dobivene rezultate upita na baze bi mogli koristiti kao test podatke za modele predviđanja lebdećih čestica PM10. Kako bi smanjili troškove i vrijeme postavljanja upita na baze podataka The Apache Hive Partition pomaže u optimizaciji objekata na temelju povezanih podataka stvarajući cluster. Na slici 10. Prikazan je model partition-a i bucketing-a.

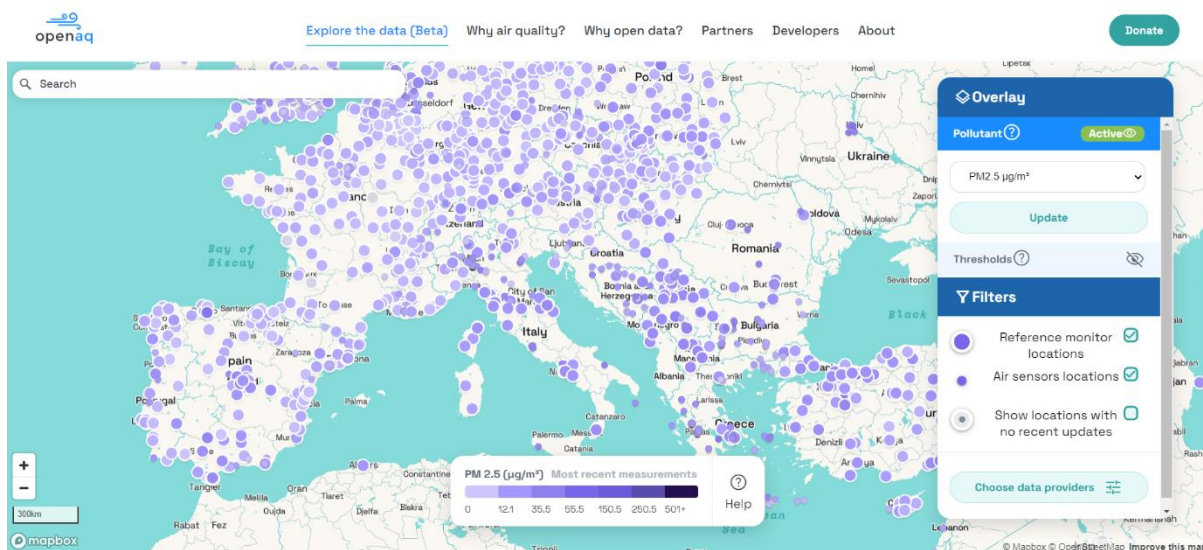
Slika 10. Prikaz Apache Hive Partition i Bucketing modela



izvor: autorski rad prema Apache Hive

Zbog takvog segmentiranja i optimizacije je moguće dobiti brže i uz manje troškova složenije i kompleksnije analize podataka. OpenAQ REST API dopušta do 300 zahtjeva po prozoru od 5 minuta ili otprilike jedan zahtjev po sekundi. Podaci se ažuriraju prema lokaciji svaka 72 sata i organiziraju kako bi se omogućilo filtriranje prema zemlji, lokaciji, godini i mjesecu. Na slici 11. na dashboardu (interaktivnoj stranici) prikazuje mjerenja izvornih podataka plinova i lebdećih čestica u stvarnom vremenu.

Slika 11. Prikaz platforme openAQ prema PM2.5 u svijetu



izvor: platforma openAQ

Podatci se bilježe na raznim uređajima te prikupljene podatke se obrađuju i analiziraju, te se rezultati vizualiziraju. Za potrebe vizualizacije moguće je filtrirati izvore podataka koji mjere štetne tvari u zraku prema službenim mjernim postajama, sensorima i ostalim uređajima.

Platforma hackAIR je nastala u 2016. kao razvojni projekt otvorene platforme za građane u svrhu podizanja svijesti o praćenju kvalitete zraka. Projekt je podržan kroz program EU-a „Platforme kolektivne svijesti za održivost i društvene inovacije” od prosinca 2018. Početna verzija platforme razvijena je 2017., te se pilot projekt provodio u Njemačkoj i Norveškoj. Druga verzija platforme objavljena je 2021 kroz program Horizon 2020. Na platformi se prikupljaju podaci od izvora:

- državnih, područnih mjernih postaja za praćenje kvalitete zraka
- raznih dostupni harverski uređaja kao što su postavljeni senzori, IoT, te ostali pametni uređaji za mjerenje kvalitete zraka
- kolaboracijom i dijeljenjem sa ostalim otvorenim državnim i područnim platformama za praćenje i prikupljanju kvalitete zraka
- na temelju zapažanja tj. percepcije podaci osjetljive skupine stanovništva, zainteresiranih građana, udruga za zaštitu okoliša
- podaci društvenih mreža sa označenim dostupnim geolokacijskim i vremenskom obilježem

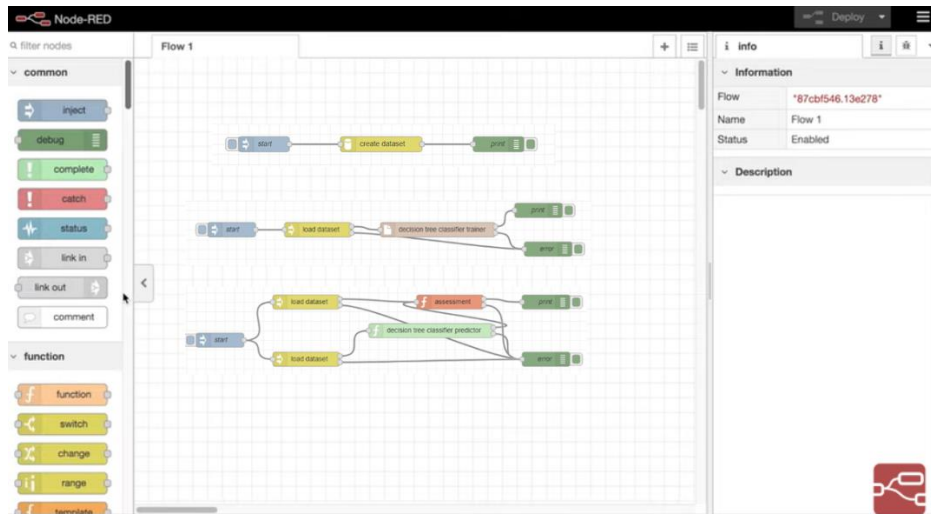
Web aplikacija je razvijena u vizualnom programskom jeziku Node-RED za zajedničko povezivanje više hardverskih uređaja, APIs i online servisa kako bi razmjenivali informacije i podataka, tvoreći tako mrežu aplikacija koje definira kao procese. Neke od njih mogu biti višestruko povezane. Kreiraju se programski blokovi na osnovi tijeka (engl. flow) koji predstavljaju procese koji će se obaviti kada blok dobije potrebnu informaciju na ulaz, a zatim obrađenu informaciju proslijediti preko svog izlaza novom bloku/elementu.

Aplikacija koristi set čvorova (engl. nodes)⁴ pod nazivom Node-RED-contrib-machine learning koji sadrže algoritme strojnog učenja (engl. Machine Learning). Takvi tokovi (engl. flows) se sastoje od ulaznih podataka, kreiranja set podataka, korištenje modela i evaluacije te rezultata modela. Nakon kreiranja seta podataka u csv format koji je dostupan u mapi test za

⁴ Čvorovi (engl. nodes) su osnovna podatkovna struktura koja sadrži podatke i jednu ili više veza s drugim čvorovima. Čvorovi se mogu koristiti za predstavljanje strukture stabla ili povezanog popisa. U takvim strukturama gdje se koriste čvorovi, moguće je prijeći s jednog čvora na drugi čvor.

obučavanje modela klasifikacije podataka koristi se algoritam stablo odlučivanja (decision tree classifier) te evaluacija i prikaz rezultata. Na slici 12. Prikaz tijeka (engl. flow) u Node-RED.

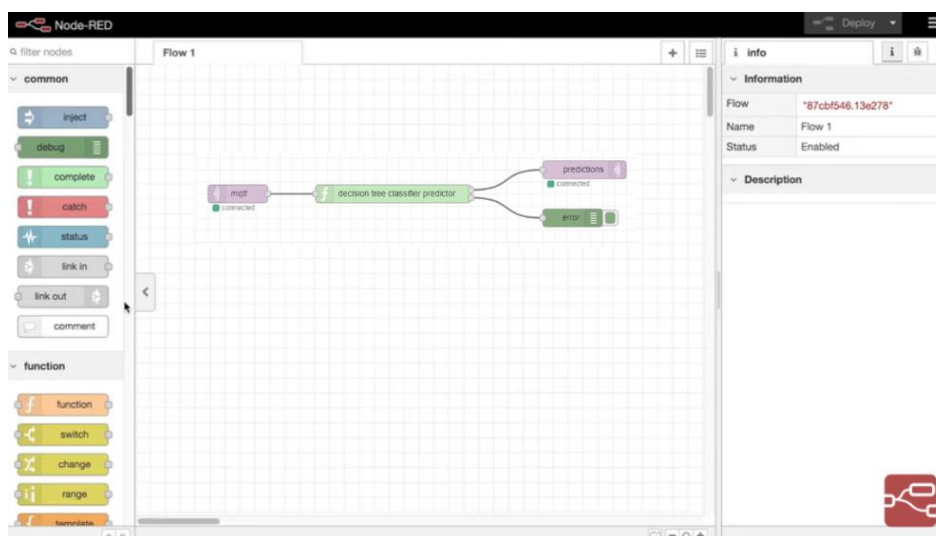
Slika 12. Prikaz tijeka u programskom jeziku Node-RED



izvor: program Node-RED

Nadalje, sljedeći prikaz toka rezultata u stvarnom vremenu, gdje se podaci dobivaju dobivaju preko standardnog protokola MQTT-a preko kojeg se razmjenjuje komunikaciju između senzora, nosivih uređaja i drugi uređaji Interneta stvari (IoT) i platforme za praćenje kvalitete, a koji obično moraju slati i primati podatke putem mreže s ograničenim resursima i ograničenom propusnošću. Nakon, korištenja modela klasifikacije odnosno algoritma stablo odlučivanja, dobiveni rezultati predikcije se šalju natrag uređaju.

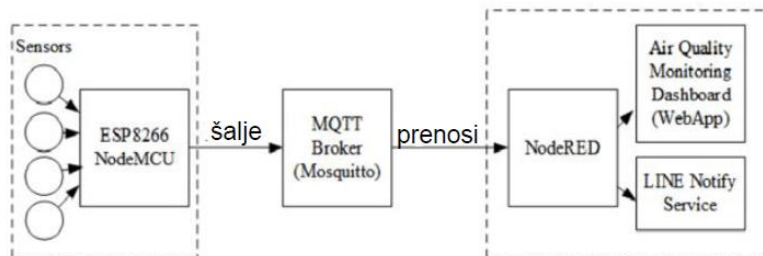
Slika 13. Prikaz slanja podataka predikcije putem protokola MQTT



Izvor: program Node-RED

Na slici 14. je prikaz systemske arhitekture hackAIR platforme koja pokazuje razmjenu podataka između pametnih uređaja uz upotrebu protokola MQTT

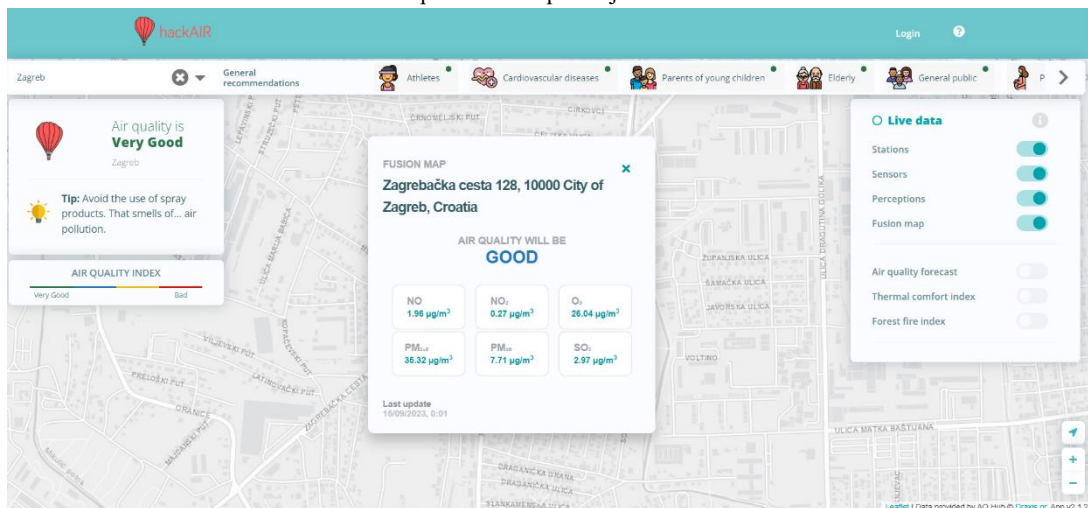
Slika 14. Prikaz systemske arhitekture platforme za praćenje kvalitete zraka



Izvor: autorski rad prema Node-RED

Vizualizacija je omogućena kroz Node-RED dashboard (Slika 15.) koji je namjenjen za prikazivanje predviđanje podataka u stvarnom vremenu kao skup rezultata po pojedinim onečišćivačima u zraku.

Slika 15. Prikaz platforme za praćenje kvalitete zraka hackAIR



Izvor: hackAIRa

4. Rudarenje podataka i strojno učenje

4.1. Prediktivna analitika i rudarenje podataka

Krovni pojam analitika podataka kao grane podatkovne znanosti se odnosi na praksu proučavanja složenih obrazaca iz velike količine podataka, korištenjem statističkih i analitičkih alata za utvrđivanje povezanosti i zavisnosti među podacima, za predviđanje budućih događaja te izvlačenje znanje iz podataka koje nam može pružiti informacije potrebne za izradu strategije i donošenje učinkovitih poslovnih odluka. Za kvalitetnu analitiku podataka potrebno je prikupiti podatke iz odgovarajućih izvora radi preciznije izvedbe.

Prije same izvedbe, korištenjem alata analitike podataka, potrebno je pripremiti podatke, što podrazumijeva čišćenje podataka i predobrada koji predstavljaju ključne pretpostavke za kvalitetnu analitiku. Priprema podataka uključuju pretvaranje sirovih podataka u prikladan format za analizu, uklanjanje pogrešaka, odstupanja i nedosljednosti te poboljšanje kvalitete i pouzdanosti podataka, u suprotnom bez pretprocesa obrade i čišćenja podataka, dobiveni rezultati mogu odvesti u krivom smjeru.

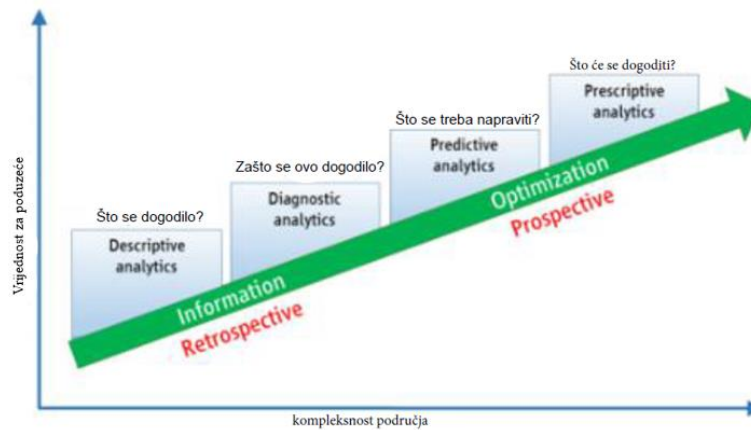
Bez analitike podataka, veliki skupovi podatka ne bi imali nikakve vrijednosti i ne bi bili poslovno upotrijebljivi. Korištenjem alata analitike podatka ti podaci postaju iskoristiv i vrijedan izvor za donošenje važnih strateških odluka.

Postoje četiri ključne vrste analitike podatka koje se koriste pojedninačno ili skupno ovisno o informacijama koje se traže i poslovnoj namjeni:

- deskriptivna – “Što se dogodilo?”
- dijagnostička – “Zašto se ovo dogodilo?”
- preskriptivna – “Što se treba napraviti?”
- prediktivna – “Što će se dogoditi?”

Na slici 16. Prikazane su različite vrste analitike podataka prema vrijednosti za poduzeće i kompleksnosti.

Slika 16. Prikaz različitih vrsta analitike podataka



Izvor: atutorski rad prema ISSA

Deskriptivna analitika koristimo za analizu međusobnih odnosa trenutnih i povijesnih podataka te praćenje trendova⁵. Smatra se najjednostavnijom oblikom analize podataka jer opisuje trendove i odnose, bez naprednije analize. Deskriptiva analitika koriste alate analitike podataka kao MS Excel za statističke analize, tehnike rudarenja podataka. Podaci su prikazani kroz razne alate vizualizacije kao Google Charts i Tableau koji račlanjuju podatke, identificiraju trendove i povezanosti podacima, te koriste vizualizaciju za prikaz informacija. Primjena deskriptivne statistike se koristi u raznim izvješćima kao praćenje broja prometa po web stranicama kao što je broj klikanja, razni financijski izvještaji te ostalo. Na temelju dobivenih informacija donose se poslovne odluke.

Dijagnostička analitika podataka obično slijedi nakon deskriptivne kao logičan slijed koraka analitike a može se provodi ručno, korištenjem algoritama poput linearne regresije ili sa statističkim alatom kao što je MS Excel. Dijagnostička analitika je proces korištenja podataka kako bi odredili uzroke koje dovode do događaja i koeficijente korelacije, odnosno međusobnu povezanosti dviju ili više varijabli. Koristi se najčešće u marketinške svrhe kod ispitivanja potražnje tržišta, ponašanja i preferencija kupaca, zatim kod dijagnostika kod tehničkih

⁵ Vremenski nizovi koji prate kretanje pojave kroz vrijeme

problema hardware komponenti računala, u ljudskim resursima za ispitivanja zadovoljstva radnim mjestom za procjenu fluktuacije zaposlenika.

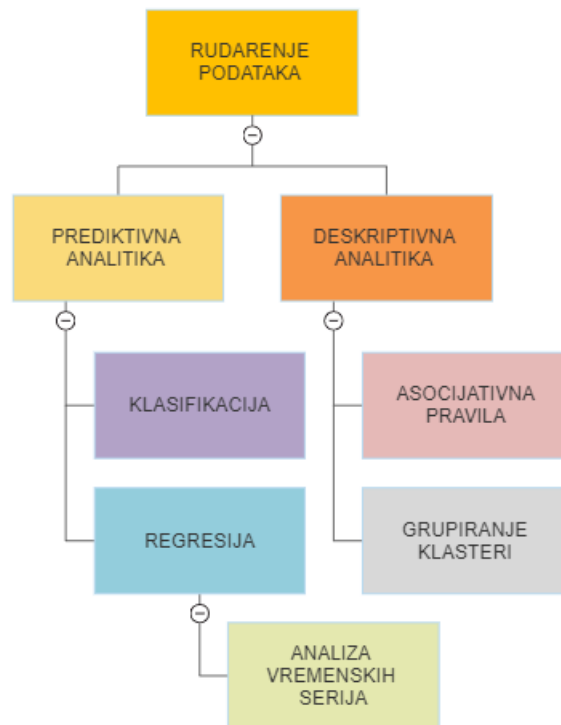
Preskriptivna analitika podataka određuje na temelju svih relevantnih varijabla i inputa najoptimalnije rješenje, što je čini najkorisnijim analitikom za donošenje odluka. Kod preskriptivne analitike koristi se algoritmi strojnog učenja (engl. Machine Learning) kako bi se u što kraćem vremenu obradila ogromna količina podataka. Koristi se kod donošenja investicijskih ulaganja, te povećanje prihoda od prodaje. Ostale primjene su preporučene aktivnosti na društvenih mreža temeljem ponašanja korisnika, otkrivanje finacijskih prevara, poboljšanje performansi alata na temelju aktivnosti korisnika, automatizacija slanja marketinških e-mail kampanja te ostalo.

Prediktivna analitika spada u najnaprednije analitike podataka. Opisuje je proces kojim se izvlače informacije iz velikog skupa podataka. Kako bi ispitala obrasce i trendove u trenutnim i povijesnim podacima koristi razne statističke tehnike od prediktivnog modeliranja, strojnog učenja, umjetne inteligencije do rudarenja podacima (engl. Data Mining) kako bi odredila koji od tih obrazaca će se ponavljati čime se predviđa budući ishod. Prediktivna analitika i rudarenja podataka koriste se na način da tehnikom rudarenjem podataka se otkrivaju brzo i precizno skriveni obrasci među podacima trenutnim i povijesnim podacima, uči iz iskustva, te donosi informacije o obrascima i vezama među podacima. Prediktivna analitika koristi otkrivene obrasce iz velike količine podataka na temelju kojih predviđa buduće ishode za donošenje starteških odluka. Brojne su primjene korištenja prediktivne analitike uz tehnike rudarenja podataka u raznim poslovnim, znanstvenim, medicinskim, akademskim, ekološkim, državnim te ostalim područjima a neke od njih otvaraju velike mogućnosti i potencijale u budućnosti kao:

- predviđanje ponašanja kupaca u trgovini
- predviđanje presuda na sudovima,
- prepoznavanje i identificiranje ljudi, mjesta, životinja te ostalog iz fotografija, videa,
- predviđanje rizika za oboljenja u zdravstvu,
- detekcija finacijskih prevara u stvarnom vremenu
- detekcija znakova rojenja pčela
- detekcija komunikacije među kitovima

Na slici 17. je prikaz za bolje razumijevanje kako prediktivna analitika koristi tehnike rudarenja podataka može se usporediti sa drugim procesima analitike podataka.

Slika 17.. Prikaz tehnika rudarenja podataka u prediktivnoj i deskriptivnoj analitici



Izvor: autorski rad

4.2. Metode rudarenja podataka

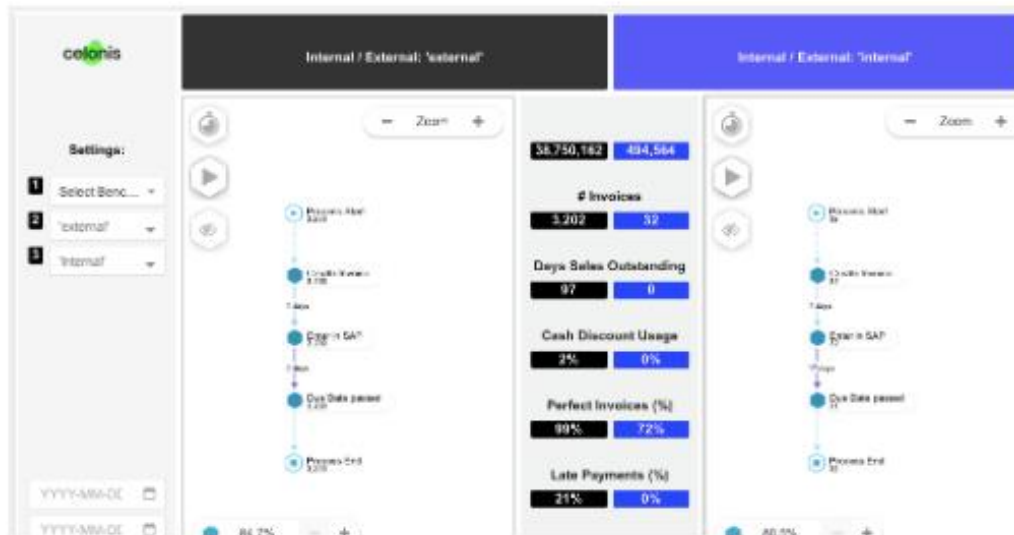
Velike količine podataka koje se generiraju u raznim poslovnim područjima kao što su korporacije, poduzeća, banke, finacijske institucije, marketinške djelatnosti su podaci u kojima je skriveno znanje. Metoda rudarenja podataka se nastoje otkriti nove vrijednosti u podacima kojima bi se otkrile informacije za donošenje važnih odluka za nastavak poslovanja, detekciju rizika u poslovanju, odluke o ulaganju te ostalih vrsta analitika koje se svakodnevno primjenjuju u poslovnim područjima. Tako u svojoj knjizi, autori Garača i Jadrić (2011.) osvrću se na pojavu informacijske krize u 20. stoljeću koja se odnosila na generiranje velike količine podataka koje su posljedica naglog rasta novih tehnologija. Razvijanjem i primjenom raznih metoda, tehnika te analitičkih alata moguće je iz beskorisnih ogromnih količina podataka dobiti korisne informacije za poslovanje. Sve veću primjenu tehnike rudarenja podataka nalazimo i kod analitike podataka u medicinskim ustanovama, u akademskim te znanstvenim istraživanjima, kod demografskih praćenja te javnih tijela i državnih institucija. Postoji više definicija i intepretacija tehnike rudarenje podataka (engl. data mining) te njezinih metoda. Prema članku autora Fayyad, Piatetsky-Shapiro, Smyth (1996.) opisuju nastanak koncepta

riječi za nalaženje korisnih obrazaca među podacima koji predstavljaju vrijedne informacije među kojima su se kroz vrijeme osim termina rudarenja podataka koristili su i ostali termini kao izvlačenje znanja (engl. knowledge extraction), otkrivanje informacija (engl. information discovery), branje ili “žetva” informacija (engl. information harvesting), arheologija (povijesnih) podataka (engl. data archeology), procesuiranje obrazaca podataka (engl. data pattern processing). Termin rudarenje podataka većinom su koristili statističari, podatkovni analitičari te unutar sustava za informacijski management kao MIS (Management Information System). Naziv rudarenje podataka sve se više koristi kod rada sa bazama podataka. Tek od 1989. počinje se koristiti naziv rudarenje podataka na raznim radionicama za otkrivanje znanja iz baza podataka, te počinje dobivati na popularnosti kao službeni naziv korištenjem u područjima zajedno sa umjetnom inteligencijom i strojnim učenjem. Autori Berry i Linoff (2004.) u svom su radu naveli važnu ulogu rudarenja podatka u skladištima podataka. Ističu kako podaci pohranjeni u skladištu podataka bez inteligente obrade su od male ili nikakve koristi za organizaciju. Analitičkom obradom podataka mogu se pronaći obrasci ponašanja neke pojave, odnosa između varijabli, napraviti predviđanja za buduće događaje, razviti nove ideje na temelju dobivenih znanja iz informacija. Korištenjem analitičkih alata i tehnike rudarenje podataka moguće je iskoristiti ogromne beskorisne količine velikih podataka kako bi se iz njih dobilo znanje za donošenje optimalnijih poslovnih odluka i najboljih rješenja. Nadalje, moglo bi se zaključiti kako je tehnika “rudarenje podataka korak u procesu otkrivanju znanja iz baza podataka koji se sastoji od primjene analize i algoritama nad podacima kako bi se otkrili brojni korisni obrasci/modeli podataka” (Fayyad, et.al., 1996, 40). Također, kako bi sve saželi osnovna definicija rudarenje podataka je “proces otkrivanja zanimljivih uzoraka, modela i drugih vrsta znanja u velikim setovima podataka” (Han, et. al., 2006, 2). Tehnike i koncepti rudarenja podataka daju osnovu za automatizaciju gradnje modela procesa. Metodologija rudarenja procesima ili procesno rudarenje (engl. process mining) koristi razne tehnike rudarenja podataka unutar algoritama rudarenja procesa kao što su grupiranje ili klasteri, klasifikacija, rudarenje sekvencijalnih obrazaca za analizu zapisa događaja i izvlačenja vrijednih uvida u poslovne procese. Poduzeća sve više žele pratiti stanje i odvijanje procesa kao što je korištenje podataka u stvarnom vremenu, zapisi o incidentima.

Praćenje se nastoji smanjiti nepotrebni koraci unutar radnih procesa kako bi se ubrzalo poslovanje. Zbog toga se naziva još poslovno procesno rudarenje (engl. business process mining). Stoga, definicija procesnog rudarenja ili poslovno procesno rudarenje “je tehnologija i analitička disciplina za procesnu inteligenciju i razvijanje razumijevanje procesa tvrtke

vođeno podacima” (Dilmegani 2023.) Koraci u procesnom rudarenju su: paniranje, priprema podatka, proces otkrivanja i vizualizacije, proces analize te procjena procesa (Slika 18.)

Slika 18. Prikaz procjene procesa u alatu Celonis

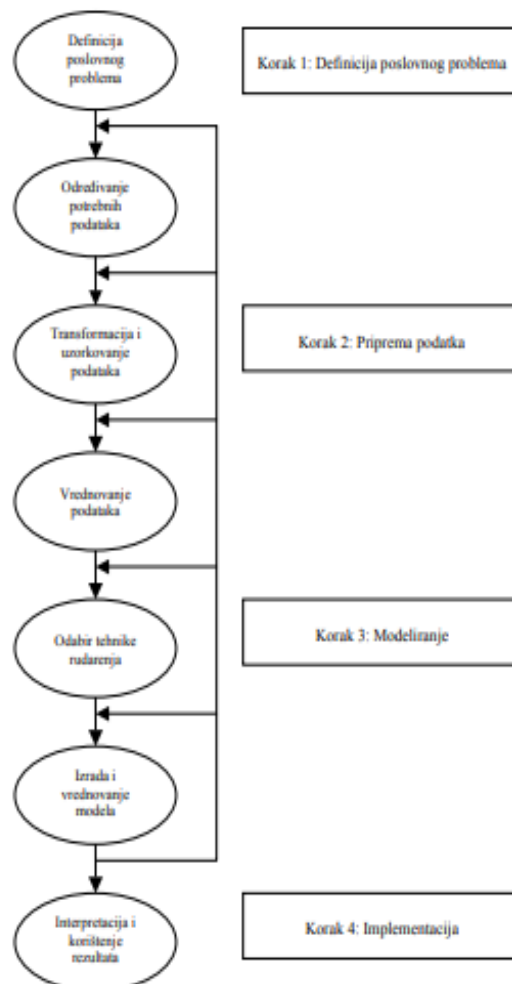


Izvor: Celonis

Primjena procesnog rudarenja koriste sva područja poslovanja najviše u IT uslugama automatizacije, bankarskom poslovanju, poboljšanje procesa organizacija (audit), korisnička podrška te ostalo. Postoje razni alati na tržištu za procesno rudarenje najpoznatiji su među komercijalnim alatima IBM Process Mining, Aris Software AG, Celonis, UiPath Process Analyzer, Minit Microsoft te besplatni kao Apromore, Disco, Process Street.

Metodologija rudarenja podataka se može gledati kao proces. Na slici 19. Prikaz procesa rudarenja podataka..

Slika 19. Prikaz procesa rudarenja podataka



izvor: Pejić Bach 2005

U svom radu Pejić Bach (2005.) navodi 4 (četiri) koraka u procesu metodologije rudarenje podataka :

1. Definicija poslovnog problema
2. Priprema podataka
3. Modeliranje
4. Implemetacija

Definicija poslovnog problema započinje prepoznavem problem koji treba riješiti te cilj koji je potrebno postići na što odgovora analitika podataka (deskriptivna, dijagnostička, preskriptivna,

prediktivna). Temeljem analitike odabrati će se najbolja metoda rudarenja podataka. Korak pripreme podataka obuhvaća:

- određivanje potrebnih podataka – (povijesnih ili podataka u stvarnom vremenu) koji će se koristiti za izradu modela
- transformacija podataka – ukloniti šum podataka, nekozistentne podatke, duplikate, odabrati prikladan oblik i formu za unos u alate rudarenja podataka
- uzorkovanje podataka – odabrati relevantne podatke koji se odnose na cilj zadatka i čiji uzorak će odgovarati modelu. Takav uzorak se dijeli na dva dijela - za testiranje modela i za izradu modela kako bi se na taj način utvrdila efikasnost modela
- vrednovanje podataka – izuzeti netipične podatke poput ekstremnih vrijednosti, “dirty data” koje se odnose na nepostojeće podatke, netočne vrijednosti te definicije.

Modeliranje je osnovni proces gdje se metode rudarenja podataka primjenjuju da bi se pronašle veze i obrasci među podacima, kako bi se moglo predvidjeti budući ishod, izgradnjom modela. Evaluacija modela se odvija u fazi provjere efikasnosti i skalabilnosti modela. Implementacija rezultata u obliku donošenja strateških odluka ili korištenje modela predikcije u poslovanju ovisi o interpretaciji značaja dobivenog znanja odnosno informacija.

Ovisno o tipu i vrsti podatka koji se pohranjuju u organizaciji primjenivati će se različiti koncepti i metode rudarenja podataka. Postoje 6 (šest) osnovnih metoda rudarenja podataka koje se primjenjuju za rješavanja određenih problema:

- klasifikacija (engl. classification)
- grupiranje (engl. clustering)
- regresija (engl. regression)
- asocijativna pravila (engl. association rules)
- detekcija izdvajanje (engl. outlier detection)
- sekvencijalni obrasci (engl. sequential patterns)
- predikcija (engl. prediction)

Klasifikacija (engl. classification) sadrži važne i relevantne informacije o podacima i metapodacima⁶ (engl. metadata). Klasificira i dijeli podatke u različite razrede. Primjene klasifikacije podataka se mogu naći za različite tipove podataka npr. kod otvorenih podataka su multimedija, tekstualni podaci, geo podaci, podaci vremenskih nizova, web itd. Metoda

⁶ Metapodaci su podaci o podacima, sastavni dio su podatka i zajedno tvore jednu cjelinu

klasifikacije se primjenjuje za sve vrste baza kao objektno-orijentirane, transakcijske, relacijske baze te ostale.

Grupiranje (engl. clustering) predstavlja grupiranje dijelova informacija u grupe povezanih objekata. Tzv. klasteri se odnose na skrivene obrasce, predstavlja nenadzirano učenje, budući da uči iz tih obrazaca. Postoje razni primjeri primjene kao u sustav kao što je CRM (Customer relationship management), analitika web stranica, rudarenje teksta, razna znanstvena istraživanja te ostalo.

Regresija (engl. Regression) identificira i analizira veze između varijabli te njihovu međuovisnost određivanjem zavisne i nezavisne varijable. Cilj je regresije pomoću vrijednosti nezavisne varijable X prognozirati kretanje zavisne varijable Y , koja je s njom u uzročno-posljedičnoj vezi. Primjeri korištenja metode mogu se naći u poslovnim primjenama gdje se treba otkriti uzročno-posljedična veza između dviju ili više varijabli kao kretanja potražnje tržišta, preferencije kupaca u odnosu na ponudu proizvoda te ostalo.

Asocijativna pravila (engl. association rules) pomažu otkriti poveznicu između dvije ili više varijabli tražeći skrivene obrasce ponašanja u setu podataka tj. ispitivanje povezanosti između dviju ili više varijabli. Utvrđuje se smjer (pozitivna ili negativna veza) i intezitet te veze. Asocijativna pravila provjeravaju vjerovatnosti obrazaca prema općem obliku izvršne naredbe ako-onda (engl. if-then). Naredbe se provode samo ako je zadovoljena prva naredba ako (engl. if). Upotreba asocijativni pravila se obično provodi nad zdravstvenim podacima i podacima o prodaji proizvoda kad se želi utvrditi koeficijenta (intezitet) korelacije tj. povezanosti dviju ili više varijabli u setovima podataka.

Detekcija izdvajanja (engl. outlier detection) je metoda promatranja seta podatka kojom se utvrđuju oni podaci koji se izdvajaju od drugih po svojem obrascu ponašanja i ne pripadaju u niti jedan obrazac. Primjena ove metode najučinkovitija u različitim područjima informatičke sigurnosti kao neovlašteni upadi, detektiranje anomalija i neobičajenih ponašanja, u bankarskom poslovanju za detekciju financijskih prijevара, detekcija prevara kreditnim i debitnim karticama.

Sekvencijalni obrasci (engl. sequential patterns) je metoda za evaluaciju sekvencijalnih podataka kako bi otkrili sekvencijalne obrasce. Ova metoda procjenjuje određene kriterije kao što su učestalost pojavljivanja, vrijeme trajanja ili vrijednost u skupu nizova kako bi našao zanimljive i korisne skrivene obrasce. Primjer korištenja ove metode jesu praćenje potrošnje

kupaca, otkrivanje obrazaca u transakcijskim podacima, za potrebe radova akademskog istraživanja te ostalo.

Predikcija (engl. prediction) je koja kombinira navedena obilježja prethodnih metoda kako bi pronašla obrasce u podacima te izgradila model koji će predvidjeti buduće ishode. Postoje brojne prediktivne metode, koje variraju od jednostavnih kao što je drvo odlučivanja (engl. decision tree) i linearne regresije do umjetnih neuronskih mreža (engl. artificial neural networks). Većina metoda rudarenja podataka uče iz iskustva. Neuronske mreže (engl. Neural network) ili stablo odlučivanja (engl. Decision tree) uče iz iskustva više od tisuća primjera.

4.3. Strojno učenje

Strojno učenje (engl. machine learning) je grana umjetne inteligencije (AI) i računalne znanosti koja se fokusira na upotrebu podataka i algoritama za oponašanje načina na koji ljudi uče, postupno poboljšavajući njegovu točnost. Moglo bi se još reći “na osnovnoj razini, strojno učenje je kategorija umjetne inteligencije koja omogućava računalima da uče samostalno. Sve se svodi na prilagodbu radnji kako bi popravili točnost a koja se mjeri brojem pokušaja za dobivanje ispravnog rezultata (Alzubi et.al., 2018,1). Vrste strojnog učenja (Dalbelo Bašić et.al. 2017, 27):

- nadzirano učenje (engl. supervised learning) ako je diskretna/nebrojčana: *klasifikacija*, ako je kontinuirana/brojčana vrijednost: *regresija*
- nenadzirano učenje (engl. unsupervised learning) ako su dani podaci bez ciljane vrijednosti, treba naći pravilnost u podacima:
 - *grupiranje (engl. clustering)*
 - *procjena gustoće (engl. density estimation)*
 - *smanjenje dimenzionalnosti (engl. dimensionality reduction)*
- Podržano/ojačano učenje (engl. reinforcement learning), učenje optimalne strategije na temelju pokušaja s odgođenom nagradom.

Strojno učenje (engl. machine learning), dubinsko učenje (engl. deep learning) i neuronske mreže (engl. neural network) su sve grane umjetne inteligencije. Neuronske mreže su zapravo grana strojnog učenja, dubinsko učenje grana neuronskih mreža. Strojno učenje i duboko učenje se u primjeni koriste naizmjenično, međutim postoje razlike među njima.

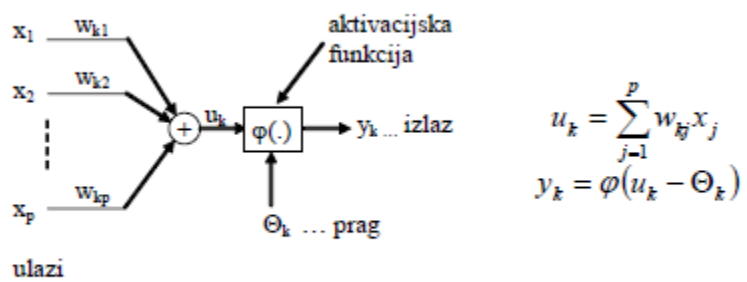
4.3.1. Neuronske mreže

Za umjetne neuronske mreže (engl. artificial neural networks) se najjednostavnije može reći kako se sastoji od seta ili niza algoritama, koje nastoje prepoznati temeljne odnose i veze u skupu podataka kroz proces koji oponaša način na koji funkcionira ljudski mozak. Zbog jasnije slike, to bi značilo kako se umjetna neuronska mreža sastoji od međusobno povezanih jedinica, simbolično nazvanih neurona. Razne tehnologije poput umjetne inteligencije (engl. artificial intelligence), strojno učenje (engl. machine learning) i duboko učenje (engl. deep learning) koriste umjetne neuronske mreže kako bi prepoznali obrasce u podacima te dobili ishode. Osnovnu jedinicu moždane kore ljudskog mozga čine neuroni koje možemo smatrati nekom vrstom jednostavnog računala premda je znatno sporiji od modernih računala. Ljudski mozak ipak nadoknađuje brzinu velikim brojem neurona (oko 10 milijardi neurona i 60 000 milijardi međuspojeva) te na mozak možemo gledati kao na složeni sustav paralelnih računala. Znanje se ne dobiva rođenjem već se mozak razvija učenjem temeljem našeg iskustva. Na sličan način funkcioniraju i neuronske mreže tako da se znanje uči, te veze između jedinica ili neurona se koriste za spremanje znanja. Zaključuje se kako je “neuronska mreža samo zbrika međusobno povezanih jedinica, svojstva mreže su određena njezinom topologijom i svojstvom neurona”. (Russell, et.al. 2005, 728). Početka razvoja ideje modeliranja umjetne neuronske mreže počinje 1943. kada su McCulloch i Pitts predstavili jednostavni matematički model neuronske mreže. Osnovni model na kojem su građene moderne neuronske mreže nazvan Perception, Frank Rosenblatt (1958) je predstavio kao jednostavni neuron koji dijeli inpute u dvije kategorije. Model Perception je korišten kod linearne metode klasifikacije pod nadziranom učenjem, pa je nazvan kao Linear Binary Classifier. Model je bio kritiziran, što je dovelo do usporavanja razvoja modela. Danas, razvojem digitalizacije i tehnologije, neuronske mreže se koriste znatno više za predviđanje budućih ishoda u različitim poslovnim područjima, medicinskim, znanstvenim i akademskim istraživanjima. Primjena algoritama umjetnih neuronskim mreža koriste se u raznim poslovnim problema kao predviđanje budućih kretanja prodaje, zadržavanja kupaca, predviđanje potražnje tržišta, verifikacija podataka, predviđanje rizika u bankarskim i financijskim institucijama te ostala područja poslovanja.

Proces učenja počinje ulazom skupa podataka, što ima za cilj pripremiti ih za konkretne zadatke. Tijekom faze učenja u neuronskoj se mreži formira matrica težina i parametri

aktivacijske funkcije na osnovi zadanih vrijednosti neurona na ulaznom sloju i očekivane (točne) vrijednosti neurona na izlaznom sloju. Inputi predstavljaju set vrijednosti koji trebamo za predikciju finalnog outputa. Elementi modela neurona se sastoje od skupa sinapsi tj. ulaza (inputs) od kojih svaki ima svoju jačinu tj. težinu (weights). Na slici 20. Prikaz osnovnog modela NN.

Slika 20. Prikaz matematičkog izraza i modela osnovnog umjetnog neurona

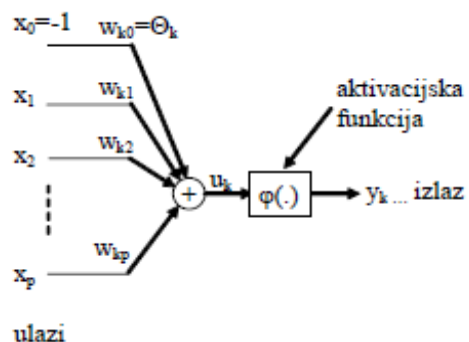


izvor: autorski rad prema FER-u

S

Prag odnosno Bias se može prikazati kao dodatni ulaz iznosa -1 i težine Θ_k . Težine postavljaju standard za jačinu signal direktno utječe na konačni ishod. Dodaju prag s vrijednošću -1 koje neuronska mreža dosad nije imala. Na slici 21. Prikaz osnovnog NN sa dodatnim ulazom.

Slika 21. Prikaz modela osnovnog umjetnog neurona sa dodatnim ulazom iznosa -1 i težine praga Θ_k



Izvor: autorski rad prema FER-u

Dodatna težina informacija je potrebna kako bi se se ista dijelila širenjem naprijed (engl. Forward Propagation) što znači podaci putuju kroz mrežu kako bi generirali ishod (output), tj.

pomicanje podataka na s lijeva (ulazni sloj) na desno (izlazni sloj). Podaci se kreću prema procesu obrade na skrivenom sloju (engl. hidden layer). Tijekom širenja naprijed, predaktivacija (funkcija izračuna ponderirane sume) i aktivacija odvijaju se na svakom čvoru skrivenog i izlaznog sloja neuronske mreže. Aktivacijska funkcija primjenjuje se na temelju ponderiranog zbroja kako bi neuronska mreža tekla nelinearno koristeći navedeni prag. Za razliku od “širenja unatrag” (engl. backpropagation) koji se koristi u strojnom učenju i rudarenju podataka za točnost predviđanja, podaci se putuju s desna (izlazni sloj) na lijevo (ulazni sloj). Neuronska mreža se može razumjeti skupom povezanih ulazno/izlaznih čvorova. Točnost čvora izražava se kao funkcija gubitka ili stopa pogreške. Širenje unatrag izračunava nagib funkcije gubitka drugih težina u neuronskoj mreži. Kako bi se neuronska mreža mogla trenirati, oslanja se i na “širenjem naprijed” i “širenjem natrag”. Arhitektura neuronske mreže određuje način na koji su jedinice (neuroni) povezani.

Postoje 3 (tri) vrste veza mreže neurona, odnosno layer-a mreže koje dijelimo na mreže bez povratnih veza koje mogu biti jednoslojne i višeslojne, mreže sa povratnim vezama te ljestvičaste mreže.

Tablica 3. Tri vrste veza mreža neurona

<u>MREŽE (LAYERS)</u>
Bez povratnih veza (freeforward networks)
<i>Jednoslojne mreže (Single-Layer)</i>
<i>Višeslojne mreže (Multi-Layer)</i>
S povratnim vezama (recurrent networks)
Ljestvičaste mreže (lattice structures)

Izvor: autorski rad

Jednoslojna mreža bez povratnih mreža (engl. single layer freeforward networks) se sastoji od jednog neurona (izlazni sloj) što znači da nema povratnih veza na ulaz jer nema procesuiranja. Koristi se za linearnu asocijativna memoriju gdje su ulazni uzorak (ključ) i odziv (zapamćeni uzorak) u linearnoj kombinaciji tj. dimenzija ulaznog i izlaznog vektora (ključ i zapamćeni uzorak) je jednaka. Višeslojna mreža bez povratnih mreža (multi-layer freeforward networks) predstavlja mrežu gdje osim ulaznog i izlaznog sloja, postoji jedan ili više skrivenih slojeva neurona. Takva mreža predstavlja duboku neuronsku mrežu (engl. deep neural network).

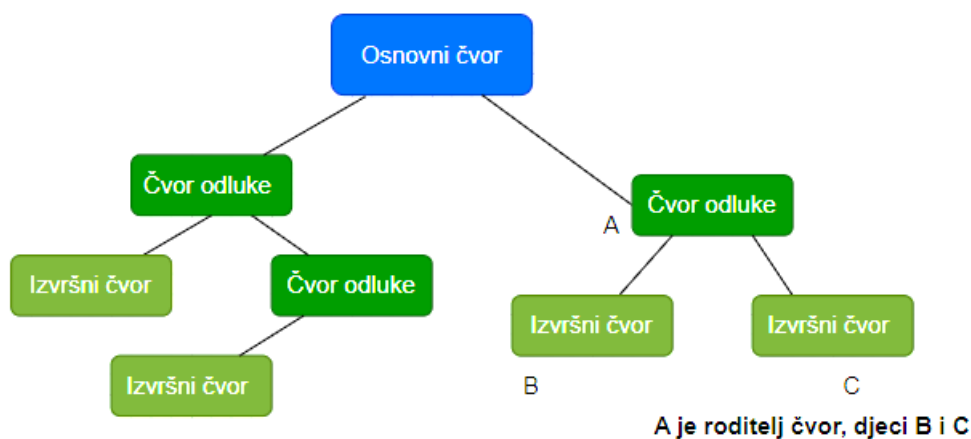
Kada je svaki neuron na svakom sloju povezan smatra se kako je mreža potpuno povezana. Mreže s povratnim vezama (engl. recurrent networks) predstavljaju mreže koje mogu imati

skriveni neuroni te više su složenije od prethodnih. Primjer su takve mreže je gdje na ulazu jednog neurona postoji izlaz drugog neurona. Takva mreža je nelinearna s diskretnim sustavom. Ljestvičaste mreže (engl. lattice networks) je mreža neurona bez povratne veze gdje je svaki ulaz spojen na sve neurone u polju. Zbog toga može imati jednodimenzionalno, dvodimenzionalno ili višedimenzionalno polje.

4.3.2 Stablo odlučivanja

Stablo odlučivanja (engl. decision tree) je algoritma nadziranog učenja (engl. supervised learning) koji se obično koristi u strojnom učenju za modeliranje i predviđanje ishoda na temelju ulaznih podataka, odnosno može se koristiti za rješavanje problema regresije i klasifikacije. Struktura prezentira stablo gdje svaki osnovni čvor (engl. root node) testira atribut, svaka grana, čvor odluke (engl. decision node) odgovara vrijednosti atributa, a svaki list (leaf node) predstavlja konačnu odluku ili predviđanje, izvršni čvor (Slika 22.).

Slika 22. Prikaz hijerarhijske strukture stable odlučivanja

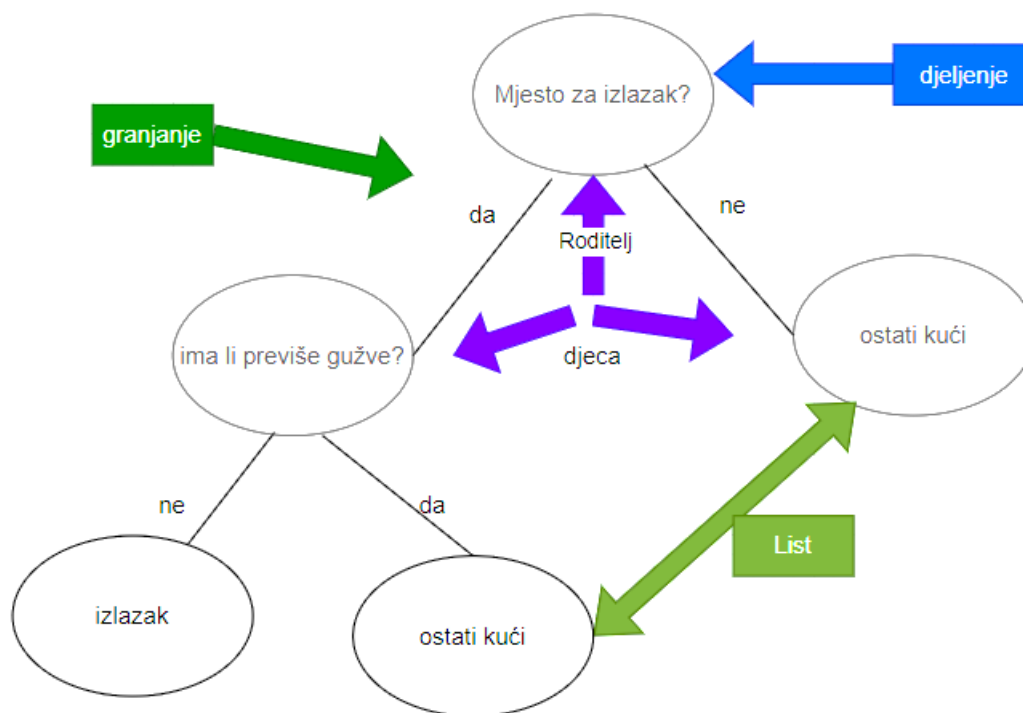


izvor: autorski rad

Struktura stabla odlučivanja zasniva se na logičkom modelu učenja koji pomaže u predviđanju temeljem danih inputa odnosno podataka. Podaci mogu biti kvalitativna ili kvantitativna. Logiku stabla odlučivanja možemo koristiti kao pomoć u svakodnevnim odlukama koje

donosimo u životu kao odluke o kupnji proizvoda, odabir mjesta putovanja, prihvaćanje novog posla, odluke koje donose poduzeća o ulaganju u nove projekte, lansiranju novog proizvoda, banke osiguravajuće kuće u odobravanju kredita, visine premije osiguranja. Kako bi se bolje razumijela logika algoritma, tj. način dijeljenja i granjanja, te moglo intepretirati rezultate stable odlučivanja potrebno je promotriti sljedeći dijagram (Slika 23.)

Slika 23. Prikaz strukture stable odlučivanja



izvor: autorski rad

Osnovni čvor roditelj (engl, parent node) dijeli se na podčvorove ili djecu (engl. child node). Osnovni čvor se dijeli engl. split) a podčvorovi (djeca) se dalje granjaju (engl. branch) na listove sve dok se ne prestanu granjati, odnosno kada se dobije decision node ili child node sa dva ishoda ili lista (leaf nodes) više što predstavlja konačni rezultat odluke, za ili protiv. Stabla odlučivanja se najčešće upotrebljavaju za rješavanje prediktivnih problema kod kojih se predviđa vrijednost binarne ciljne varijable tj. slučaj kada ciljna varijabla ima dva ishoda 0 ili 1.

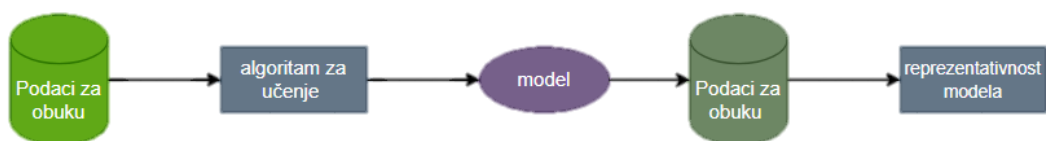
Cilj klasifikacije podataka je izraditi algoritam kojim će se moći predvidjeti s visokom vjerojatnošću da će se dogoditi jedan od mogućih ishoda. Najčešće se koristi za dva moguća ishoda.

Klasifikacija je proces u dva koraka (Danilo, 2014, 7-9):

- korak učenja (trenirati model podacima)
- korak predviđanja (testiranje točnosti modela)

strojnom učenju. Na slici 24. Prikazani su koraci u stablu učenja.

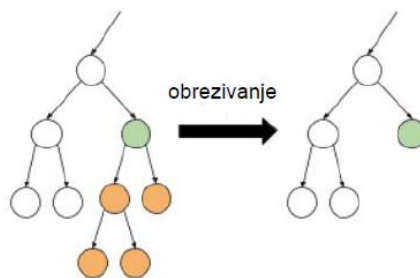
Slika 24. Prikaz koraka u stablu odlučivanja



Izvor: autorski rad

U koraku učenja, model se razvija na temelju danih podataka o obuci. U koraku predviđanja, testira se točnost modela kako bi se na temelju njega mogla raditi predviđanja. Cilj korištenja stabla odlučivanja je stvoriti model koji se može koristiti za predviđanje klase ili vrijednosti ciljne varijable učenjem jednostavnih pravila odlučivanja izvedenih iz prethodnih podataka (podaci obuke). Kod tehnike stabla odlučivanja kao i kod modela ostalih algoritama u strojnom učenju može se dogoditi da u koraku učenja, model u čijem setu podataka za obuku postoji šum podataka, odnosno prenapučenost podataka. U tom slučaju će točnost modela će više odgovarati rezultatom podacima za obuku (engl. training data) nego setu podataka (set data) za model predviđanja. To se zove engl. overfitting. Kako bi se riješio takav problem potrebna je tehnika koja se koristi u stablu odlučivanje naziva obrezivanje (engl. pruning) kako bi se simbolično odrezale one grane koje su od manjeg značaja za klasifikaciju podataka odnosno za ishod rezultata. Cilj je održati model optimalnim za točniji ishod (Slika25.).

Slika 25. Prikaz obrezivanja u stablu odlučivanja



izvor: autorski rad

4.3.3. Linearna regresija

Regresijski model ima veliku primjenu u predviđanju koji utvrđuje postojanje međuovisnosti između dviju ili više varijabli. Utvrđivanje međuovisnosti započinje određivanjem zavisne ili regresand varijable i nezavisne regresorske varijable. Pomoću nezavisne varijable X opisuje, kontrolira te predviđa kretanje zavisne varijable Y , koja se nalazi s njom u uzročno posljedničnoj vezi, što bi značilo da određuje koliko promjena varijable X (predikator) utječe na varijablu Y (kriterij). Regresijski model između zavisne varijable Y i p nezavisnih varijabli X_1, \dots, X_p , gdje t predstavlja vrijeme, ima oblik (Abraham et.al., 1983, 9-10);

$$y_t = f(x_t; \beta) + e_t \quad (1)$$

gdje $f(x_t; \beta)$ je matematička funkcija od p zavisne varijable $x_t = (x_{t1}, \dots, x_{tp})$ i nepoznatih parametra $\beta = (\beta_1, \dots, \beta_m)$. Pretpostavljamo kako su parametri nepoznanica, funkcija modela je poznata.

Kod stohaističkih ili probabilističkih modela u , e_t predstavlja pogrešku, odnosno slučajnu varijablu za koju se pretpostavlja:

1. ako je vrijednost $E(e_t) = 0$ i njezina varijanca $V(e_t) = \sigma^2$ je kontanta i ne ovisi o t ,
2. ako je $Cov(e_t, e_{t-k}) = E(e_t e_{t-k}) = 0$, za sve t i $k \neq 0$, varijable nisu linearno povezane
3. te ako je standardna greška normalne distribucije tada pretpostavka 2.

Prema prirodi pogreške e , zavisne varijable y je sama po sebi slučajna varijabla. Oblik regresijskog modela može se onda izraziti pretpostavka prema uvjetnoj distribuciji od y_t , za $x_t = (x_{t1}, \dots, x_{tp})$;

1. očekivana vrijednost $E(y_t | x_t) = f(x_t; \beta)$ ovisi o nezavisnoj varijabli x_t ; i parametrima β i $V(y_t | x_t) = \sigma^2$
2. ako su zavisne varijable y_t i y_{t-k} za različita vremenska razdoblja (ili subjekti) nisu linearno povezani $\text{Cov}(e_t, e_{t-k}) = E[(y_t - f(x_t; \beta)) (y_{t-k} - f(x_{t-k}; \beta))] = 0$,
3. uvjet za x_t, y_t slijedi vrijednost normalne distribucije $f(x_t; \beta)$ i varijancu σ^2 , to je definirano pravilom $N(f(x_t; \beta), \sigma^2)$

čim se pretpostavlja da srednja vrijednost je uvjetno distribirana varijablom y_t što je funkcija nezavisne varijable x_t . Ova veza nije deterministička, što znači za svaki x_t , odgovara y_t raspršeno oko koeficijenta varijacije. Linearna kombinacija ulaznih varijabli što implicitno postavlja model gdje je izlazna varijabla linearno ovisno o ulaznim varijablama. Jednostavni oblik modela linearne regresije je;

$$y_t = \beta_0 + \beta_1 x_t + e_t \quad (2)$$

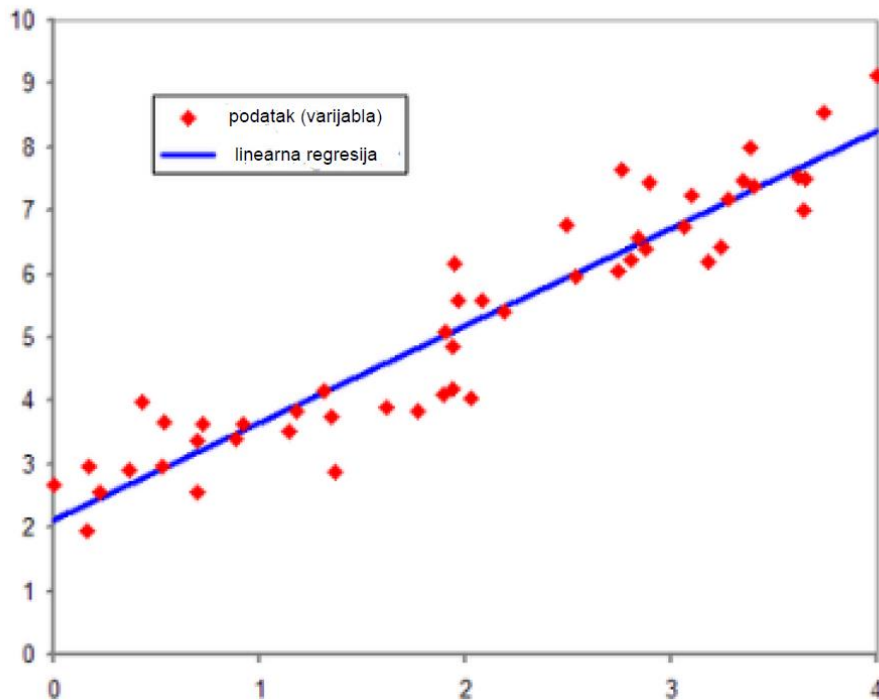
gdje je β_0 konstantni član modela koji pokazuje kolika je očekivana vrijednost zavisne varijable Y , kada nezavisna varijabla poprimi vrijednost nula. Koeficijent regresije β_1 pokazuje za koliko jedinica će se promijeniti očekivana vrijednost zavisne varijable Y ako se nezavisna varijabla X poveća za 1 (jednu) jedinicu varijable. Ta metoda ocjenjivanja parametara naziva se metoda najmanjih kvadrata.

Nakon ocjene parametra, procjenjuje se apsolutna i relativna reprezentativnost modela. Najčešće korištena apsolutna mjera je standardna greška regresije koja pokazuje prosječno odstupanje od stvarnih regresijskih vrijednosti zavisne varijable Y . Model se ocjenjuje i relativnom mjerom koja se naziva koeficijent varijacije regresije koja pokazuje udio standardne greške regresije u aritmetičkoj sredini zavisne varijable. Pomoću njega se može ocjenjivati reprezentativnost dvaju ili više regresijskih modela.

Pretpostavke za analizu varijacije za linearni model naveden je u prethodnom odlomku i predstavlja temeljnu analizu varijance. Na kraju reprezentativnost modela se mjeri još jednom relativno mjerom koja se naziva koeficijent determinacije koji pokazuje ukupni postotak odstupanja varijable od njenog prosjeka protumačeno regresijskim modelom.

Koeficijent determinacije se kreće u intervalu $0 \leq R^2 \leq 1$. Kada su njegove vrijednosti bliže jedinici pokazuju viši stupanj reprezentativnosti promatranog regresijskog modela. Na slici 26. je prikazan pripadajući graf linearne kombinacije svih varijabli.

Slika 26. Prikaz primjera grafa linearne regresije sa jednom nezavisnom varijablom



izvor: autorski rad prema linear regression

Prikaz grafa linearne regresije prikazuje idealan pravac kroz ishodište (linearnu kombinaciju). Za daljnu obradu u ovom radu, zaključuje se na temelju gore navedenih pretpostavki, kako postoji nekoliko uvjeta kako bi model bio pouzdan to su:

- linearnost -varijable imaju linearni odnos, promjena nezavisne varijable X određuje promjenu zavisne varijable Y)
- Nezavisnost - opažanja nezavisnih (prediktorskih) varijabli ne utječu na opažanja zavisnih varijabli (kriterij)
- Normalnost i homoskedastičnost (engl. homoscedasticity) – greške su u modelu normalno distribuirane i varijanca je konstantna (količina nezavisne varijable X nema utjecaja na prosječno kvadratno odstupanje standardne greške regresije)

Razlikujemo 4 (četiri) regresije prema broju ulaznih (nezavisnih, prediktorskih) varijabli X i broju izlaznih (zavisnih, kriterij) varijabli Y.

Tablica 4. Vrste regresije

$x_1=y_1$ jednostavna regresija
$x_1=y_1,y_2$ multivarijantna jednostavna regresija
$x_1, x_2 = y_1$ višestruka regresija
$x_1, x_2 = y_1,y_2$ multivarijantna višestruka regresija

Izvor: autorski rad

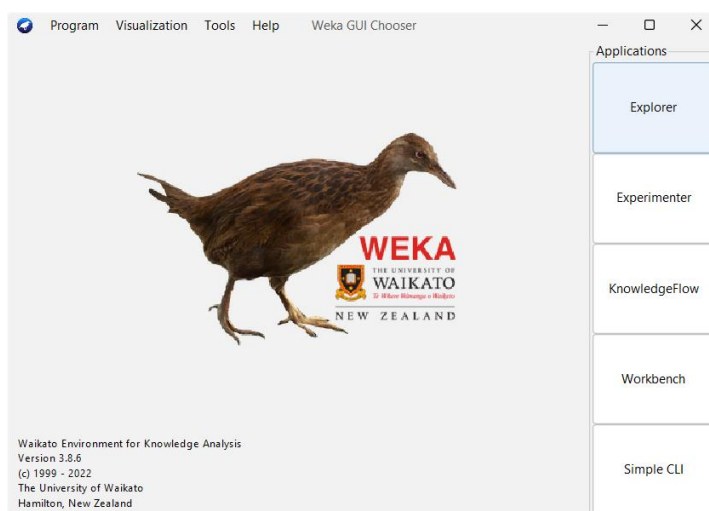
Algoritam linearna regresije baš kao i stablo odlučivanja spada u algoritme nadziranog učenja (engl. supervised. learning). Predviđa kontinuirane i diskontinuirane varijable u područjima primjene i istraživanja kao što su marketing, prodaja, cijena proizvoda, visina plaće, utvrđivanje dobi te ostalo.

5. Korištenje alata Weka u analizi i predviđanju kvalitete zraka

5.1. Metodologija istraživanja

Za potrebe ovog rada koristiti će se izvori sekundarnih službeni podatci Izvješća o praćenju kvalitete zraka na postajama Državne mreže za trajno praćenje kvalitete zraka u 2021., Ministarstva gospodarstva i održivog razvoja, izvješća nastavnog zavoda za javno zdrastvo Splitsko-dalmatinske županije, Odjel za ispitivanje zraka, tla i buke, koji sadrže podatke iz 2021. godine o izmjerenim podacima kvalitete zraka u Hrvatskoj koji su zabilježeni na mjernim postajama. Koristiti će se statistički podaci o broju stanovnika sa popisa iz 2021, Državnog zavoda za statistiku RH te statistički podaci o broju registriranih vozila, po vrsti vozila i vrsti goriva u 2021.godini, Centar za vozila Hrvatske. Podaci o ukupnom broj vozila u prometu u Republici Hrvatskoj koji su prikupljeni na stanicama za tehnički pregled prema godišnjem izvješću, Državnog zavoda za statistiku RH za 2021. Također, za potrebe ovog rada koristiti će procjena o broju zagađivača (tvornice, rafinerije, odlagališta te ostalo) na području Republike Hrvatske. Za izgradnju ovog modela koristi će se strojno učenje u besplatni softverski alatu Weka (Waikato Environment for Knowledge Analysis), licenciran pod GNU općom javnom licencom. Razvijen je na Sveučilištu Waikato, Novi Zeland i uz knjigu "Data Mining: Praktični alati i tehnike strojnog učenja". Sadrži zbirku algoritama strojnog učenja za zadatke rudarenja podataka u ažuriranom, interaktivnom sučelju. Za ovaj rad koristiti će se verzija 3.8.6.. (Slika 27.).

Slika 27. Prikaz sučelja Weka alata za strojno učenje



Izvor: Weka

5.2. Rezultati istraživanja

5.2.1. Deskriptivna analiza varijabli

Prikazani su u tablici 5. atributi, njihovi opisi, format i modaliteti.

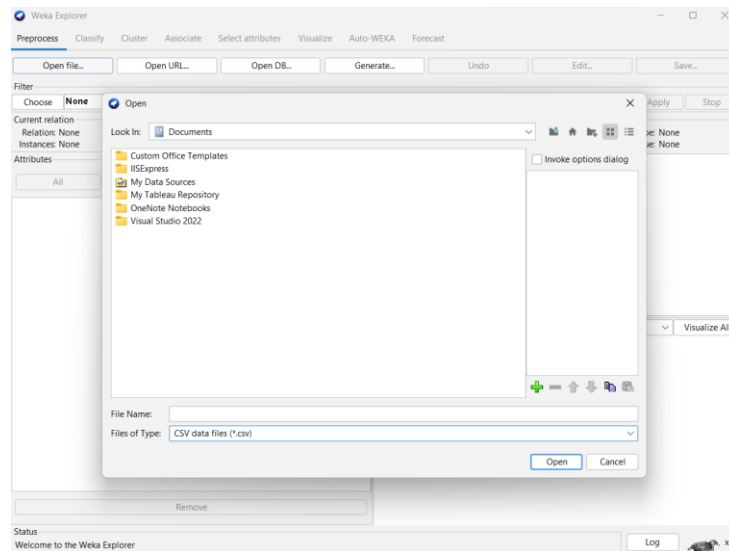
Tablica 5. Popis atributa u bazi kvaliteta zraka HR.arff

Naziv atributa	Opis atributa	Numerički/Nominalni	Modaliteti atributa (nominalnih i binomnih)	Najmanja i najveća vrijednost numeričkih atributa
Mjerna Postaja	Naziv mjerne postaje za kvalitetu zraka u Republici Hrvatskoj	Nominalni	Zagreb, Velika Gorica, Osijek, Rijeka, Desnič, Varaždin, Koprivnica, Kutina, Sisak, Slavonski Brod, Karlovac, Opuzen, Pula, Polača, Parg, Plitvička jezera, Vis, Višnjan, Dugi otok, Split	
SO2	Sumporov dioksid	Numerički		25,170
SO2 24 sata	Sumporov dioksid	Numerički		9,33
NO2	Dušikov dioksid	Numerički		27,304
CO	Ugljični monoksid	Numerički		0.5, 4.7
O3	Prizemni ozon	Numerički		75,174
H2S	Sumporovodik	Numerički		2, 17.23
NH3	Amonijak			11.4, 184
C6H6	Benzon	Numerički		4.3, 50.3
PM10	Lebdeće sitne čestice	Numerički		72,588
PM10 24 sata	Lebdeće sitne čestice koncentracija 24 sata	Numerički		29,166
PM2,5	Lebdeće sitne čestice	Numerički		36,324
broj stanovnika	Ukupan broj stanovnika na području mjernih postaja	Numerički		1.389,767.131
broj regvozila	Ukupan broj vozila koji su u prometu (dizel, bezin, hibrid) na području mjernih postaja	Numerički		23.571,403.418
zagadivaci	Postoji industrija, rafinerija, odlagalište području mjerenja vrijednost je 1, suprotno je 0	Numerički		0,1
Ishod	Ukoliko je koncentracija štetnih tvari u zraku premašena i/ili na području mjerenja postoji.	numerički		0,1

Izvor: autorski rad

Prije rada u alatu za strojno učenje Weka, prikupljeni podaci su se pripremili za izgradnju modela pod nazivom zrakRHkvaliteta.csv. Na početkom ekranu alata odabirom Preprocess, open file unose se set podataka u cvs (comma separated values) formatu nad kojima će model učiti i predviđati točnost modela. (Slika 28.)

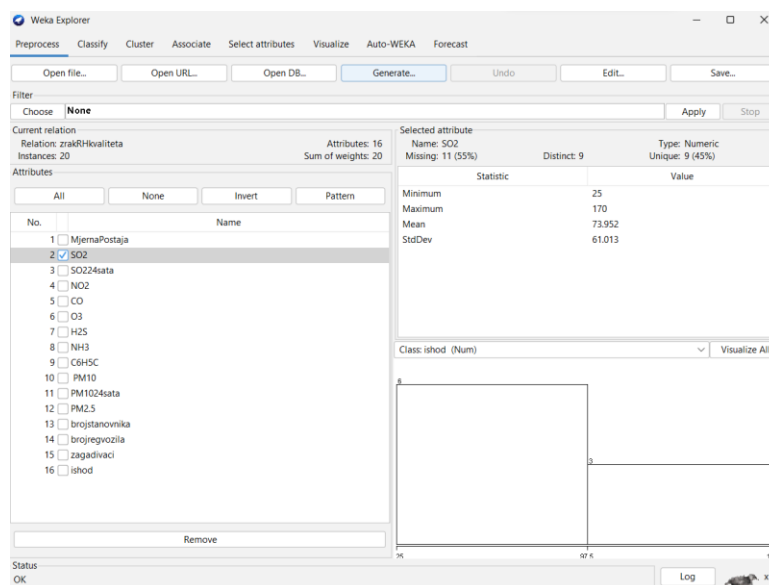
Slika 28. Prikaz preprocess u alatu Weka



Izvor:Weka

Nakon učitavanja podataka na sučelju se pokazuje set podataka za gradnju modela koji sadrži 16 atributa i 20 instanci. Za svaki pojedinačni atribut (varijablu) dobije se prikaz statističke analize kao što je najveća i najmanja vrijednost, uz pripadajuću vizualizaciju (stupičasti grafovi). Na slici 29. je prikazan set podataka za učenje.

Slika 29. Prikaz preprocessa prepoznavanja karakteristika unesenog seta podataka



izvor:Weka

Format atributa je numerički i nominalan. Nedostaju podaci za pojedine vrijednosti štetnih tvari u zraku kod nekih mjernih postaja zbog kvarova na uređaju ili nedostatnog obuhvata za mjerenje (uvjetna ocjena; obuhvat podataka < 85% , nedostatan obuhvat; obuhvat < 75%).

U tablici 6. je prikazana statistička analiza atributa u preprocesu za atribute u setu podataka zrakRHkvaliteta.arff:

Tablica 6. Statistička analiza atributa

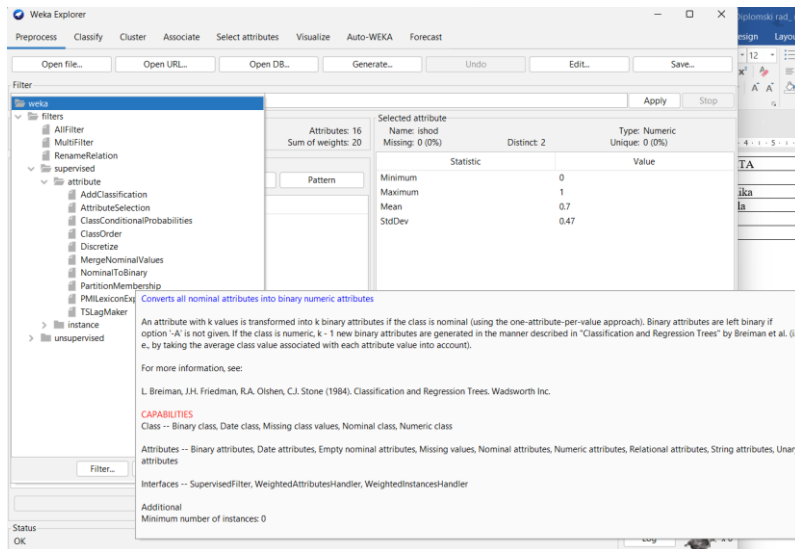
Red.br	Atribut	Prosječna vrijednost	Standardna devijacija
1.	MjernaPostaja	/	/
2.	SO2	73.952	61.013
3.	SO224SATA	17.172	8.348
4.	NO2	110.34	72.984
5.	CO	2.1	1.452
6.	O3	134.133	22.618
7.	H2S	9.077	7.672
8.	NH3	97.7	122.047
9.	C6H5C	26.36	19.156
10.	PM10	266.267	148.268
11.	PM1024SATA	90.467	36.043
12.	PM2.5	144.462	90.969
13.	brojstanovnika	74.801	168.502
14.	brojregvozila	132.744	88.991
15.	zagadivaci	0.3	0.47
16.	Ishod	0.7	0.47

Izvor: autorski rad

5.2.2. Izrada modela predviđanja razine onečišćenja zraka

Na setu podataka primjenili smo algoritam linearne regresije (engl. linear regression) kako bi na temelju procjene reprezentativnosti ili točnosti modela predvidjeti kvalitetu zraka u Republici Hrvatskoj. Regresijskom analizom želimo ispitati međuovisnost svakog atributa, utvrditi postojanje veze kako bi se pomoću nezavisne varijable moglo prognozirati kretanje zavisnih varijabla u ovom slučaju predviđanje kvalitete zraka. Također analizirati ćemo povezanosti (korelaciju) između varijabli (atributa), te njihov smjer (pozitivni ili negativni) i intezitet. Prije primjene algoritma linearne regresije bilo je potrebno primijeniti filter Nominal to Binary na atribut MjernePostaje (nominalni, opisni) kod nadziranog učenja kako bi vrijednosti svih podataka u skupu zbog potrebe analize bile numerički atributi. Prikaz filtera na slici 30.

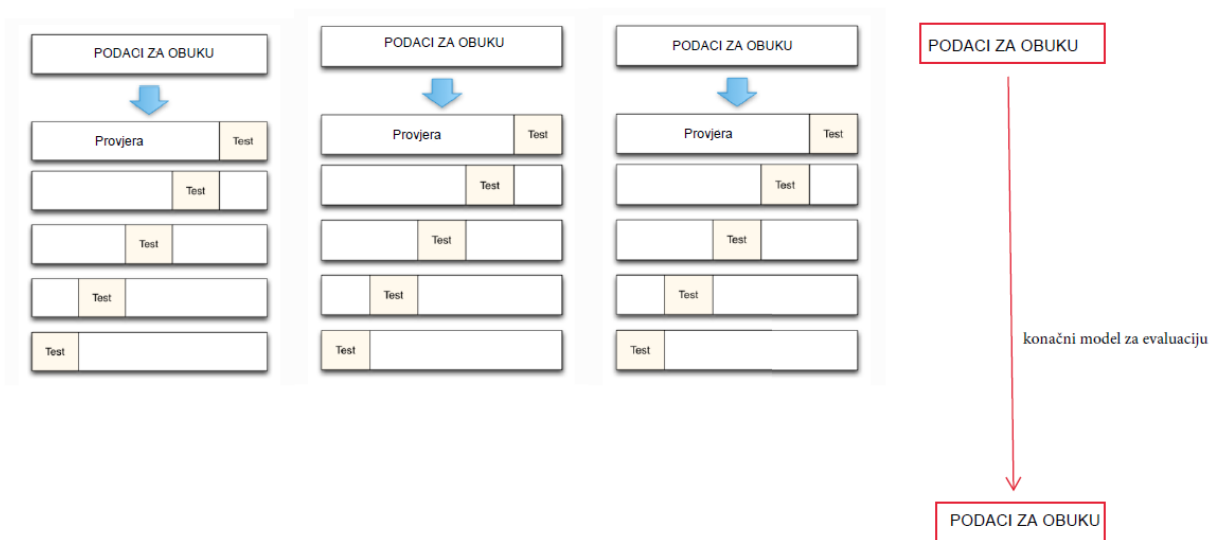
Slika 30. Prikaz primjene filtera Nominal to Binary kod nadziranog učenja



izvor: Weka

Nakon izmjene atributa MjernaPostaja u numerički atribut, dobiveni su podaci o prosječnoj vrijednosti koja iznosi 0.95 te standardnoj devijaciji koja iznosi 0.224. Primjenom filtera set podataka za gradnju modela sadrži 20 atributa i 34 instanci Nakon što se prilagodio set podataka za analizu, za provjeru (test) odabrana se metoda 10-fold cross validation koja predstavlja unakrsnu provjeru seta podataka za procjenu. Metoda dijele set podataka na dijelove radi validacije modela. Metoda pokazuje najeveću prednost korištenja kod organičenog seta podataka. Na slici 31. prikaz procesa korištenja 10-fold cross validation za testiranje podataka.

Slika 31. Prikaz korištenja k-fold cross validation



Izvor: autorski rad prema k-fold cross validation

Rezultati dobivenog modela su :

=== Classifier model (full training set) ===

Linear Regression Model

ishod =

```
-0.0309 * MjernaPostaja=Dugi otok, Varazdin, Visnjan, Plitvicka jezera, Pula, Parg, Vis,
Polaca, Opuzen, SlavonskiBrod, Sisak, Velika Gorica, Osijek, Rijeka, Desnic, Koprivnica,
Karlovac, Zagreb +
0.2092 * MjernaPostaja=Plitvicka jezera, Pula, Parg, Vis, Polaca, Opuzen, SlavonskiBrod,
Sisak, Velika Gorica, Osijek, Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
0.6068 * MjernaPostaja=Pula, Parg, Vis, Polaca, Opuzen, Slavonski Brod, Sisak, Velika
Gorica, Osijek, Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
0.1051 * MjernaPostaja=Parg, Vis, Polaca, Opuzen, SlavonskiBrod, Sisak, Velika
Gorica, Osijek, Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
-0.0744 * MjernaPostaja=Vis, Polaca, Opuzen, SlavonskiBrod, Sisak, Velika
Gorica, Osijek, Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
0.0327 * MjernaPostaja=Polaca, Opuzen, SlavonskiBrod, Sisak, Velika
Gorica, Osijek, Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
0.1096 * MjernaPostaja=Opuzen, SlavonskiBrod, Sisak, Velika
Gorica, Osijek, Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
-0.0057 * MjernaPostaja=Velika Gorica, Osijek, Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
-0.0693 * MjernaPostaja=Osijek, Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
0.1625 * MjernaPostaja=Rijeka, Desnic, Koprivnica, Karlovac, Zagreb +
-0.0264 * MjernaPostaja=Desnic, Koprivnica, Karlovac, Zagreb +
-0.1743 * MjernaPostaja=Koprivnica, Karlovac, Zagreb +
0.0164 * MjernaPostaja=Karlovac, Zagreb +
0.1254 * MjernaPostaja=Zagreb +
0.002 * SO2 +
-0.0037 * SO224sata +
0.0732 * CO +
0.0047 * O3 +
0.0026 * C6H5C +
0 * PM10 +
0.0013 * PM1024sata +
0.0003 * PM2.5 +
-0.0002 * brojregvozila +
-0.126 * zagadivaci +
-0.9737
```

Time taken to build model: 0.01 seconds

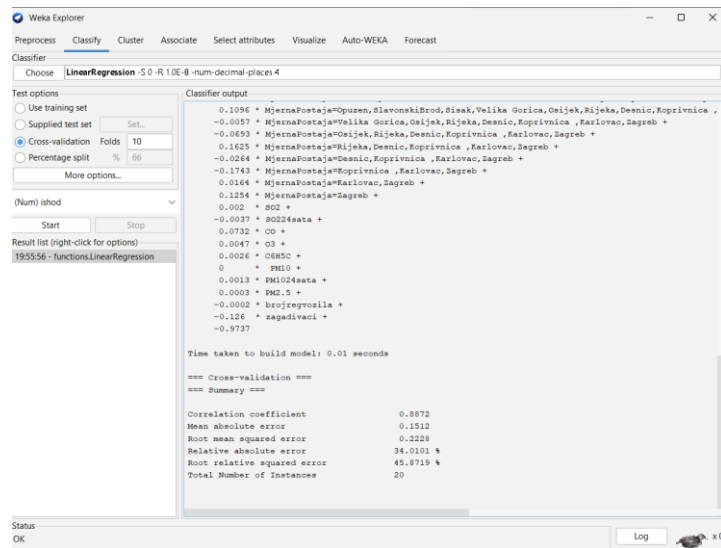
=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.8872
Mean absolute error	0.1512
Root mean squared error	0.2228
Relative absolute error	34.0101 %
Root relative squared error	45.8719 %
Total Number of Instances	20

Rezultat pokazuje kao sumu regresijskih metrika za ocjenu modela. Apsolutna pogreška modela - Mean Absolute error (MAE) mjeri prosjek greške između nezavisne i zavisne varijable. Za model ona iznosi 0.1512 što pokazuje minimalnu grešku. Nadalje, Root mean squared error (RMSE) predstavlja standarnu devijaciju, odnosno standardnu grešku regresije koja mjeri prosječno odstupanje stvarnih od regresijskih vrijednosti zavisne varijable (prediktora). RMSE iznosi 0.2228. Koeficijent korelacije iznosi 0.8872 koji pokazuje pozitivnu korelaciju jakog inteziteta što bi značilo da rast jedne varijable prati rast druge druge. Tako u modelu su prikazane pozitivne i negativne veze (negativni predznak). Iznad rezultata prikazana je jednažba modela linearne regresije. Na slici 32. prikaz rezultata modela u Weka.

Slika 32. Prikaz rezultata korištenjem algoritma Linear Regression



izvor: Weka

Također, ostali algoritmi kao neuronske mreže Multilayer Perceptron. Ovdje se također koristi metoda testa provjere modela 10-fold cross validation, čime se sriječava overfitting ali i bolje određuju optimalni parametri za učenje modela.

Rezultat algoritma Multilayer Perceptron također daje bolje rezultate o točnosti tj reprezentativnosti modela, čime je apsolutna pogreška MAE i standardna greška RMSE još manja, što govori u prilog modelu (Slika 33.)

```

Class
  Input
  Node 0

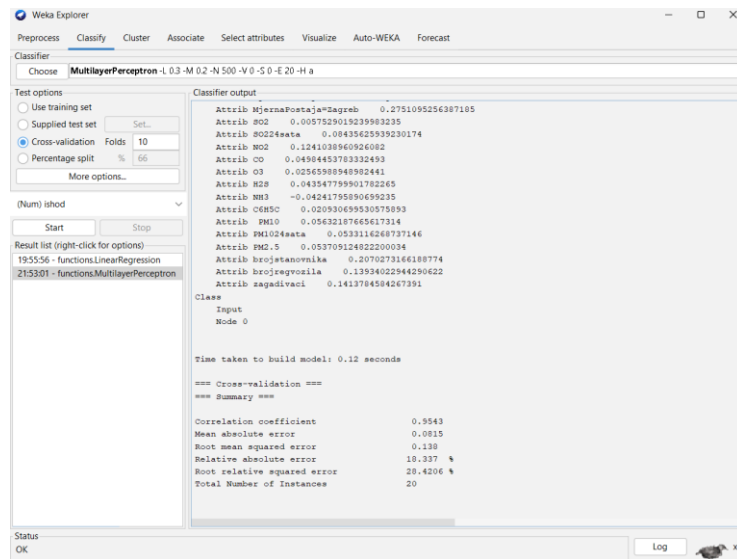
Time taken to build model: 0.12 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient          0.9543
Mean absolute error            0.0815
Root mean squared error        0.138
Relative absolute error        18.337 %
Root relative squared error     28.4206 %
Total Number of Instances      20
    
```

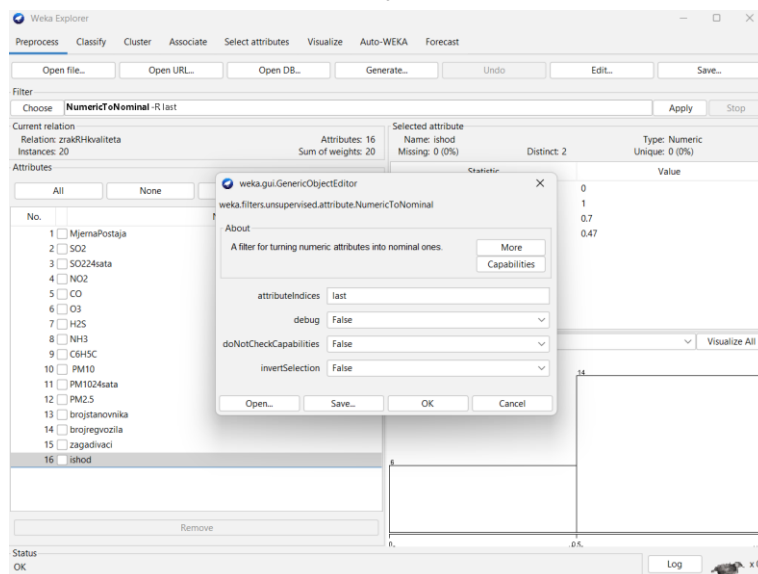

Slika 33. Prikaz korištenja algoritma Multilayer Perceptron



Izvor: Weka

Kao i prethodni algoritmi stablo odlučivanja rješava problem regresije i klasifikacije, pa se provjera modela tj. njegova točnost može provjeriti. Također, će se koristiti 10-fold cross validation zbog navedenih prednosti u prethodnom algoritmu. Kod stabla odlučivanja potrebno je koristiti filter na Y varijabli (prediktoru) u ovom slučaju ishodu filter Numeric to Nominal budući da stablo odlučivanja J48 (C4.5) zajedno koristi kombinaciju numeričkih i nominalnih podataka (Slika 34.)

Slika 34. Prikaz korištenja filter Numeric to Nominal



izvor: Weka

Dobiveni su rezultati o reprezentativnosti model.

Number of Leaves : 2

Size of the tree : 3

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	17	85	%
Incorrectly Classified Instances	3	15	%
Kappa statistic	0.5833		
Mean absolute error	0.2786		
Root mean squared error	0.3762		
Relative absolute error	63.3239	%	
Root relative squared error	79.6742	%	
Total Number of Instances	20		

=== Detailed Accuracy By Class ===

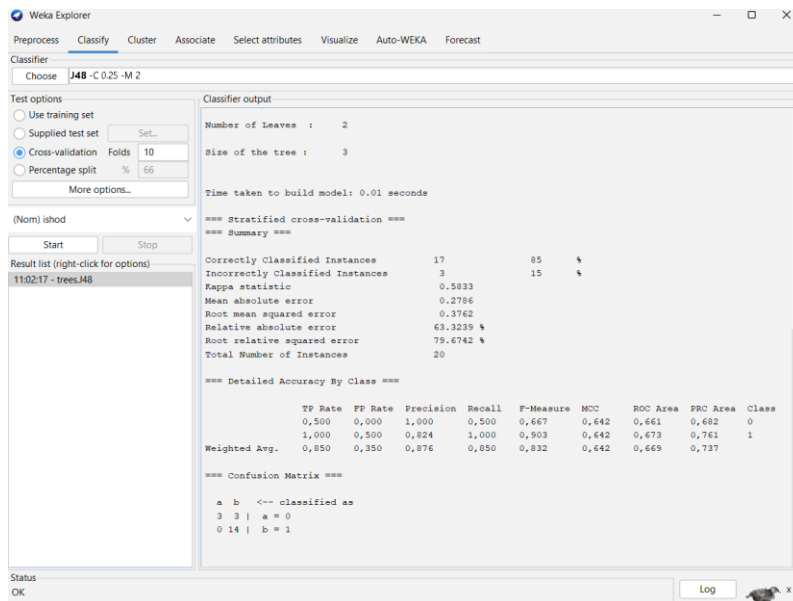
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,500	0,000	1,000	0,500	0,667		0,642	0,661	0
	1,000	0,500	0,824	1,000	0,903		0,642	0,673	1
Weighted Avg.	0,850	0,350	0,876	0,850	0,832		0,642	0,669	

=== Confusion Matrix ===

```
a b <-- classified as
3 3 | a = 0
0 14 | b = 1
```

Rezultat MAE iznosi 0.2786 i RMSE iznosi 0.3762 što je opet dosta zadovoljavajuće za interpretaciju o točnosti podataka. Pravilno je klasificirano 85% podataka, a netočno 15%. Konfuzijska matrica se intepretira ovisno o kontekstu podataka pa tako element na poziciji *aa* i element na poziciji *bb* su točne klasifikacije. Element na poziciji *aa* (poznat kao i true positive) u modelu predstavlja kako 3 varijable ne utječu na kvalitetu zraka, dok element na poziciji *bb* (poznat kao i true negative) u modelu predstavlja 14 varijabli koje utječu na predviđanje kvalitete zraka. Element na poziciji *ab* i element na poziciji *ba* su netočne klasifikacije. Element na poziciji *ab* (poznat kao i false negative) predstavlja 3 varijable koji su modelom pogrešno klasificirano. Element na poziciji *ba* (poznat kao i false positive) koji iznosi 0 što znači nema pogrešno klasificiranih varijabli. Prikaz rezultata algoritma J48 u Weki (Slika 35.).

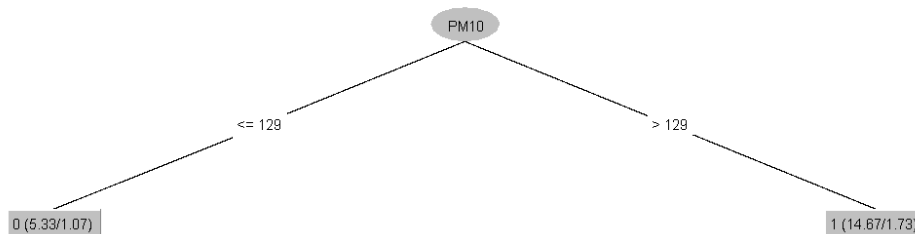
Slika 35. Prikaz rezultata korištenjem algoritma J48 (C4.5)



izvor: Weka

Na slici 36. Prikazan je rezultat modela stablu odlučivanja.

Slika 36. Prikaz vizualizacije rezultata modela u stablu odlučivanja



Izvor: Weka

Stablo odlučivanja sastoji iznosi ukupno: osnovnog čvora (root node) a sastoji od dva unutarnja čvora (decision node), te se može intepretirati kako ukoliko lebdeće čestice PM10 iznose manje od ili jednako 129 $\mu\text{g}/\text{m}^3$ godišnje kvaliteta zraka će biti zadovoljavajuća u Republici Hrvatskoj, ovisno o kretanju PM10, 5 varijali neće zagađivati zrak, dok je 1 pogrešno klasificirana. Ali ukoliko je veća od 129 $\mu\text{g}/\text{m}^3$ doći će do zagađenje zraka jer 14 varijabli utječe na zagađenje dok za 2 varijable se neće utjecati negativno na kvalitetu zraka.

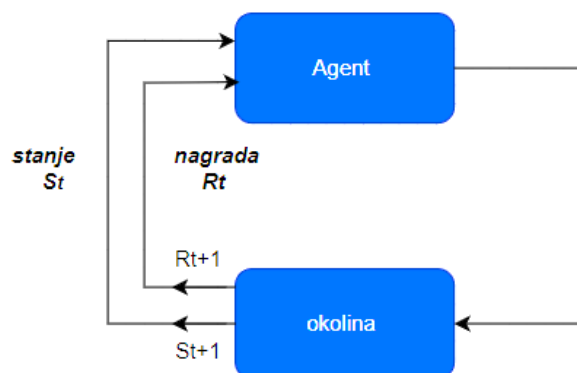
5.3. Predviđanja budućih trendova rudarenja podataka i strojnog učenja o kvaliteti zraka

S obzirom kako je onečišćenje zraka globalno predstavlja problem za zdravlje ljudi. Sve više stručnjaka u znanstvenim istraživanjima traže nove načine za najbolju analizu i predviđanje štetnih tvari u zraku u stvarnom vremenu. Iako posljednjih godina došlo je do znatnog poboljšanja u performansama i različitostima modela za predviđanje kvalitete zrake, mogućnost za točno predviđanje ostaje još na željama. Iako je analiza vremenskih serija ima važnu ulogu u predviđanju kvalitete zraka, neki napredniji algoritmi kao što su neuronske mreže i algoritam Arima (The Autoregressive Integrated Moving Average) koji je poznat za simulaciju i analizu.

Osim nadziranog i nenadziranog učenja, pojačano ili ojačano učenje (engl. reinforcement learning) se više koristi učenju koristi rješenja za samokonfiguriranje i optimizaciju iz povratnih informacija okoline. Pojačano učenje je strategija sa odgodom nagrade. Može se promatrati kao automatizirani sustavi temeljeni na agentima, ciljno usmjerenom učenju i donošenju odluka. Okolina je sve s čime agent stupa u interakciju i nije u potpunosti pod kontrolom agenta.

U svakom koraku izračuna t , reprezentacija okoline koju prima agent naziva se stanje, predstavljeno sa $s_t \in S$, gdje je S skup svih mogućih stanja (Chang et al., 2019, 1-2). Na slici 37. je prikazano kretanje pojačanog učenja.

Slika 37. Prikaz rada pojačanog učenja



izvor: autorski rad prema Chang et al. 2019

Agent (strojno učenje) regira i prima informaciju iz okoline. Poduzima akciju gdje koristi sve moguća stanja (operacije). Agent odlučuje sam o akcijama tj. tehnika i radnja koje će koristiti.

Sve radnje koje koje poduzima agent direktno utječe na okružje. Stanje(S) predstavlja sva moguća stanja u kojima se nalazi agent. Nagrada (R) predstavlja odgovor okruženja tj. okoline kojom se ocjenjuje ili validira svaka radnja koju poduzima agent. Tako se trenira tj. uči agent kako je jedini kriterij odnosno uvjet nagrada koju prima.

Novi trendovi su zasigurno kombinacija neuronskih mreža, analiza vremenskih serija i ARIMA kao rješavanje nedostataka za ARIMA za “hvatanje” nelinearnih veza u obrascima. Model RL-ARIMA-NN je model arhitekture koja je zasnovana na tri modula: pojačano ili ojačano učenje, “širenje naprijed” neuronske mreže i vremenske serije.

6. Zaključak

U ovom radu pokazalo se kako se primjenom odgovarajućih prediktivnih algoritama strojnog učenja kao što je linearna regresija, može izgraditi model kojim će se utvrditi povezanosti i međuovisnost za predikciju kvalitete zraka za RH u budućnosti. Za gradnju modela podataka potrebno je su nezavisne varijable: broja automobila i ostalih vozila u pokretu prema potrošnji dizela, bezina i hibrida, broj stanovništva što pokazuje koliko je područje urbano i aktivno, postojanje aktivnih odlagališta, opasnih nesaniраних materijala, ostataka bivših tvornica sa opasnim tvarima po okoliš i zrak te ostale ljudske aktivnosti za koje smatramo da utječu na kvalitetu zraka i zagađenje okoliša. Primjenom odrađenih algoritama u kojima su ugrađene statističke i matematičke metode možemo dobiti rezultate predikcije za kvalitetu zraka kojima se mogu donositi odluke za kreiranje propisa i regulacije poput ograničenja prometa vozilima na benzin i dizel u središtima urbanih područja i centrima a posebno kad su procjene koncentracija štetnih tvari u zraku najveće, praćenje i kontrole aktivnih odlagališta, ostataka tvornica na kojima se nalaze opasni materijali, praćenje i kontrola industrijskih postrojenja. Sufinanciranje i prelazak na električne automobile što podrazumijeva sprječavanje zagađenja zraka i okoliša, kontrola građevinske dozvola o gradnji visoko energetske stambene i ostalih građevina na dopuštenim lokacijama, moderniziranje proizvodnje što rezultira smanjenjem štetnih plinova industrija i rafinerija u zraku, propisno upravljanje odlagalištima te veća edukacija stanovništva upravljanju zagađivačima i svijest o utjecaju zagađenja zraka i okoliša na ljudsko zdravlje.

Hrvatska mora prestati provoditi tek puku statističku i izvještajnu metodu o štetnim koncentracijama u zraku već početi koristiti mogućnosti tehnologije kao umjetne inteligencije, a naročito treba koristiti strojno učenje za izgradnju prediktivnih modela kao i tehnike dubokog učenja, kojom će se ispitati svaka uzročno-posljednična veza svih faktora koji utječu na kvalitetu zraka koja direktno utječe na onečišćenje okoliša te zdravlje njezinih stanovnika. Potrebno je raditi analize na dnevnoj bazi sa podacima u stvarnom vremenu kako bi pravodobno donosila odluke koje će značiti umanjivanje štetnih koncentracija zbog prometa u zraku u urabanim područjima. Potrebno je koristiti vlastitu infrastrukturu praćenje i analize kvalitete zraka, povezati sustave obavještanja posebno za osjetljive skupine ljudi. Potrebno je prikupljati podatke sa raznih pametnih uređaja kao što su senzori, pametni mobiteli te razni IoT uređaji. Prediktivnom analitikom i rudarenjem podataka može se utjecati na promjene u modernizaciji infrastrukture te proizvodnje kako bi iste uskladili sa ekološkom održivošću.

Ovim radom pokazalo se kako se linearnom regresijom, stablom odlučivanja te neuronskom mrežom unatoč malom setu dostupnih podatka o izmjerenim štetnim koncentracijama sa državnim mreža za praćenje kvalitete zraka i podacima o stanovništvu, broju vozila u prometu, te postojanje zagađivača na mjernom području, može dobiti model za predviđanje kretanje kvalitete zraka u Hrvatskoj.

Literatura

1. Pejić Bach, M. (2005). Rudarenje podataka u bankarstvu. (Zbornik ekonomskog fakulteta u Zagrebu)
2. Garača Ž., Jadrić M (2011.) Rudarenje podataka: različiti aspekti informacijskog društva (Ekonomski fakultet u Splitu)
3. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996): From Data Mining to Knowledge Discovery in Databases, AI Magazine, Vol. 17, 3, 37–54
4. Dilmegani C. 2023. What is Process Mining in 2023? dostupno na <https://research.aimultiple.com/process-mining/#why-use-process-mining> pristupano na 09.09.2023.
5. Berry, M., and Linoff, G. 2004. Data Mining Techniques for Marketing, Sales and Customer Support (2nd ed.). Wiley, New York.
6. Han J, Kamber M (2006). „Data mining, Concepts and Techniques “. San Francisco: Elsevier Inc.
7. DHMZ dostupno https://meteo.hr/kz/modeliranje/izvjesce_2021_kvaliteta_zraka.pdf, pristupano na dan 30.08.2023
8. DHMZ dostupno na https://meteo.hr/kz/modeliranje/ocjena2016kvalitete_zraka.pdf, pristupano na dan 30.08.2023
9. DHMZ dostupno https://meteo.hr/kz/modeliranje/kvaliteta_oborine_2017-2022.pdf, pristupano na dan 30.08.2023.
10. Državni zavod za statistiku dostupno na <https://podaci.dzs.hr/2022/hr/29136>, pristupano na dan 30.08.2023
11. Amit Kumar Y., Hasmat M., S.S. Chandel, Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models (2014) Centre for Energy and Environment, National Institute of Technology, Hamirpur, 177005 Himachal Pradesh, India
12. Danilo C. Decision tree algorithm Weka tutorial Web Mining e Information Retrieval (2014) dostupno na http://art.uniroma2.it/basili/didattica/WmIR_15_16/DecTree_WekaTutorial.pdf pristupano na 02.09.2023.
13. Kelleher J.D, Tierney B., Data Science (2018) dostupno na <https://books.google.hr/books?id=UlpVDwAAQBAJ&printsec=frontcover&hl=hr&so>

- urce=gbs_ge_summary_r&cad=0#v=onepage&q&f=false pristupano na dan 30.08.2023
14. Nastavni zavod za javno zdravstvo Splitsko-dalmatinska županija <https://nzjz-split.hr/servisi/izvjesca-o-kvaliteti-zraka/> pristupano na dan 30.08.2023
 15. Centar za vozila Hrvatske dostupno na <https://cvh.hr/gradani/tehnicki-pregled/statistika> pristupani na 30.08.2023.
 16. Izvješće pučke pravobraniteljice 2021., Analiza stanja ljudskih prava i jednakosti Hrvatskoj, 224. Pravo na zdrav život i zdrav okoliš, dostupno na <https://www.ombudsman.hr/hr/download/izvjesce-pucke-pravobraniteljice-za-2021-godinu/?wpdmdl=13454&refresh=6247119d58b0c1648824733>), pristupano na dan 30.08.2023.
 17. M.Mirković, Industrijska politika, op.cit., 179-185; D.Mišić, op.cit., 147-148.
 18. Z. Šimončić-Bobetko, Razvoj tekstilne idustrije u Hrvatskoj između dva svjetska rata (1918-1941), 101-204
 19. Z. Šimončić-Bobetko, Cementna industrija u Hrvatskoj, PP 2(1), 97-167 (1983)
 20. Što su PM čestice dostupno na <https://evomed.hr/sto-su-pm-cestice/> pristupano na dan 30.08.2023
 21. John W. Tukey, The Future of Data Analysis (1962), dostupno na <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-33/issue-1/The-Future-of-Data-Analysis/10.1214/aoms/1177704711.full> pristupano na dan 30.08.2023.
 22. Jacob Zahavi, Mining Data for Nuggets of Knowledge (1999) dostupno na <https://knowledge.wharton.upenn.edu/article/mining-data-for-nuggets-of-knowledge/> pristupano na dan 30.08.2023.
 23. W.S.Cleveland, Dana Sience: An Action Plan for Expanding the Techical Areas of Fields of Statistic dostupno na <https://docplayer.net/17350243-Data-science-an-action-plan-for-expanding-the-technical-areas-of-the-field-of-statistics.html> pristupano na 30.08.2023.
 24. Ćurko K., Španić Keznan M. (2016), Skladištenje podataka: put do znanja i poslovne inteligencije, Sveučilište u Zagrebu, Ekonomski fakultet.
 25. Russell S., Norvig P. (2005), AI a modern approach, Learning 2.3 (2005): 4. dostupno na https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf pistupano na 30.08.2023.

26. Alzubi J., Nayyar A., Kumar A. (2018), Machine Learning from Theory to Algorithms: An Overview dostupno na <https://iopscience.iop.org/article/10.1088/1742-6596/1142/1/012012/pdf> pristupano na 02.09.2023.
27. Dalbelo Bašić B., Šnajder J.(2018) FER, Uvod u strojno učenje dostupno na https://www.fer.unizg.hr/_download/repository/SU-2017-01-Uvod.pdf pristupano na 02.09.2023.
28. Abraham, B., Ledolter J., (1983), Statistical Methods for Forecasting, Univerisity of Iowa,
29. Chang , Wei laReinforcement Learning for Improving the Accuracy of PM2.5 Pollution Forecast Under the Neural Network, Department of Computer Science and Information Engineering Department, National Taiwan University, Taipei, Taiwa dostupno na https://www.researchgate.net/publication/334854393_Reinforcement_Learning_for_Improving_the_Accuracy_of_PM25_Pollution_Forecast_Under_the_Neural_Network_Framework pristupano 02.09.2023

Popis slika

Slika 1. Prikaz kvalitete zraka za Zagreb	13
Slika 2. Prikaz praćenje kvalitete zraka za Osijek na dan 01.09.2023.	14
Slika 3. Prikaz povezanosti stanovništva, njihovih aktivnosti na stvaranje opasnih razina u zraku	16
Slika 4. Prikaz mjernih postaja za praćenje kvalitete zraka u Republici Hrvatskoj	17
Slika 5. Prikaz naslovne stranice Državne mreže za praćenje kvalitete zraka na European Environment Agency	22
Slika 6. Prikaz izmjerenih podataka za mjernu postaju HUM (Otok Vis) IQAir	23
Slika 7. Prikaz pregleda prikupljenih podataka	23
Slika 8. Prikaz udjela u mjerenju i dijeljenju podataka o kvaliteti zraka	31
Slika 9. Prikaz ilustracije rada platforme OpenAQ	32
Slika 10. Prikaz Apache Hive Partition i Bucketing modela	33
Slika 11. Prikaz platforme openAQ prema PM2.5 u svijetu	33
Slika 12. Prikaz tijeka u programskom jeziku Node-RED	35
Slika 13. Prikaz slanja podataka predikcije putem protokola MQTT	35
Slika 14. Prikaz systemske arhitekture platforme za praćenje kvalitete zraka	36
Slika 15. Prikaz platforme za praćenje kvalitete zraka hackAIR	36
Slika 16. Prikaz različite vrste analitike podataka	38
Slika 17. Prikaz tehnika rudarenja podataka u prediktivnoj i deskriptivnoj analitici	40
Slika 18. Prikaz procjene procesa u alatu Celonis	42
Slika 19. Prikaz procesa rudarenja podataka	43
Slika 20. Prikaz matematičkog izraza i modela osnovnog umjetnog neurona	48

Slika 21. Prikaz modela osnovnog umjetnog neurona sa dodatnim ulazom iznosa -1 i težine praga Θ_k	48
Slika 22. Prikaz hijerahijske strukture stable odlučivanja, dostupno na	50
Slika 23. Prikaz strukture stable odlučivanja	51
Slika 24. Prikaz koraka u stablu učenja	52
Slika 25. Prikaz obrezivanja u stablu odlučivanja	53
Slika 26. Prikaz primjera grafa linearne regresije sa jednom nezavisnom varijablom	55
Slika 27. Prikaz sučelja Weka alata za strojno učenje	57
Slika 28. Slika preprocesa u alatu Weka.	59
Slika 29. Prikaz preprocesa prepoznavanja karakteristika unesenog seta podataka Weka alata za strojno učenje	59
Slika 30. Prikaz primjene filtera Nominal to Binary kod nadziranog učenja	61
Slika 31. Prikaz korištenja k-fold cross validation	61
Slika 32. Prikaz rezultata korištenjem algoritma Linear Regression	63
Slika 33. Prikaz korištenja algoritma Multilayer Perceptron	64
Slika 34. Prikaz korištenja filter Numeric to Nominal	64
Slika 35. Prikaz rezultata korištenje J48 (C4.5)	66
Slika 36. Prikaz rezultata u stablu odlučivanja	66
Slika 37. Prikaz rada pojačanog učenja	68

Popis tablica

Tablica 1. Prikaz referentnih metoda mjerenja	20
Tablica 2. Gornji i donji pragovi procjene koncentracije po onečišćujućoj tvari	21
Tablica 3. Tri vrste veza mreža neurona	49
Tablica 4. Vrsta regresije	56
Tablici 5. Popis atributa u bazi kvalitetazrakaHR.arff	58
Tablica 6. Statistička analiza atributa	60

Životopis kandidata

Julia Ilić

Datum rođenja: 08.07.1979.

Mjesto rođenja: Split, Hrvatska

E-adresa: juliailic@gmail.com

LinkedIn: <https://www.linkedin.com/in/julia-ilić-/>

OBRAZOVANJE I OSPOSOBLJAVANJE

Sveučilište u Splitu, Ekonomski fakultet

bacc.oec, smjer Financije, MMP

01/09/2011 – 01/09/2016, Split, Hrvatska

Učilište Mencl, Split

ECDL

01/10/2005 – 30/06/2006

TEČAJEVI

Sveučilište u Zagrebu, Sveučilišni računski centar - Srce

Napredni MS excel

Croz d.o.o.

Napredni SQL

RADNO ISKUSTVO

Poslovni IT tajnik

Agencija za podršku informacijskim sustavima i informacijskim tehnologijama - APIS IT

19/04/2019 – Trenutačno, Zagreb, Hrvatska

Poslovni analitičar/ Business Analyst

Agencija za podršku informacijskim sustavima i informacijskim tehnologijama - APIS IT

28/10/2018 – 19/04/2019, Zagreb, Hrvatska

Voditelj u mreži

Hrvatska demokratska zajednica - HDZ

01/01/2014 – 28/10/2018 Omiš, Hrvatska

Administrator-blagajnik

Hrvatska demokratska zajednica -HDZ

01/07/2003 – 01/01/2014

Drugi jezici

Engleski

Slušanje C2

Čitanje C1

Govorna interakcija C2

Govorna produkcija C2

Pisanje C1

Njemački

Slušanje B2

Čitanje B2

Govorna interakcija B2

Govorna produkcija B2

Pisanje B1

DIGITALNE VJEŠTINE

Business intelligence (BI)

- SAP Business Objects
- Poznavanje DWH principa skladištenja podataka
- Data Science & Data Analytics
- Rad u WEKA softveru
- Poznavanje ETL procesa
- rad sa mrežnom analitičkom obradom podataka, OLAP kočke
- Poznavanje rada u Tableau programu
- Rad sa bazama podataka DB2 SQL i ORACLE SQL

Programski jezici

- Python (Basic)
- HTML 5 & CSS
- Node-RED
- VBA basic
- Poznavanje programskih jezika C# (Xamarin), JavaScript, HTML

Modeliranje, simulacija i sistemska dinamika

- Simul8
- Bizagi Process Modeler
- Modelio
- Vensim
- Celonis

Automatizacija

- Microsoft Power Automate