

UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN
INFORMACIJSKE TEHNOLOGIJE

ZAKLJUČNA NALOGA

ZAKLJUČNA NALOGA
**ANALIZA VREMENSKEGA VPLIVA NA
KONCENTRACIJO PRAŠNIH DELCEV V
ZRAKU**

JERNEJ VASIČ

UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN
INFORMACIJSKE TEHNOLOGIJE

Zaključna naloga

**Analiza vremenskega vpliva na koncentracijo prašnih delcev v
zraku**

(Analysis of weather influence on the concentration of particle matter in the air)

Ime in priimek: Jernej Vasič

študijski program: Računalništvo in informatika – 1. stopnja

Mentor: doc. dr. Branko Kavšek

Koper, september 2013

Ključna dokumentacijska informacija

Ime in PRIIMEK: Jernej VASIČ

Naslov zaključne naloge: Analiza vremenskega vpliva na koncentracijo prašnih delcev v zraku

Kraj: Koper

Leto: 2013

število listov: 37

število slik: 4

število tabel: 8

število prilog: 3

število strani prilog: 6

število referenc: 10

Mentor: doc. dr. Branko Kavšek

UDK:

Ključne besede: Prašni delci, PM10, vremenski vpliv, podatkovno rudarjenje

Izvleček:

Zaradi vedno večjega tehnološkega in industrijskega razvoja, se znižuje kakovost zraka, saj predstavlja človekova aktivnost glavni doprinos škodljivih snovi v atmosfero. Vsebnost onesnaževal v ozračju povzroča negativne trajne vplive na okolje in zdravje ljudi. Atmosferski aerosoli so mešanica trdnih in tekočih delcev suspendiranih v zraku in se razlikujejo po obliki, velikosti in kemijski sestavi. Zaradi težnje po varstvu okolja, varovanju in izboljšanju kakovosti zraka, se izvajajo meritve obremenjenosti zraka s prašnimi delci. Namen zaključne naloge je analiza vpliva vremenskih faktorjev na koncentracijo PM10 lebdečih trdnih delcev. Prah velikosti delca, manjšega od 10 mikrometrov, se obravnava kot fini delec ali delec PM10. čeprav vreme ni povzročitelj nastanka teh delcev, pa so prav vremenski vplivi (npr: veter, temperatura) tisti, ki delce PM10 prenašajo iz mest nastanka na mesta, ki so gosteje poseljena. V zaključni nalogi bomo analizirali podatke, zbrane med leti 2007 in 2013 na merilnih postajah v Novi Gorici, v Luki Koper in na Markovcu (gosteje poseljeno področje nad Koprom). Za analizo bomo uporabili algoritme strojnega učenja (linearno regresijo, regresijska drevesa, odločitvena drevesa in odločitvena pravila) ter poskušali tvoriti napovedni model, ki bo prikazoval vpliv vremenskih faktorjev na koncentracijo delcev PM10 v zraku. Hkrati bomo napravili tudi primerjavo med napovednimi modeli za posamezno merilno postajo.

Key words documentation

Name and SURNAME: Jernej VASIČ

Title of final project paper: Analysis of weather influence on the concentration of particle matter in the air

Place: Koper

Year: 2013

Number of pages: 37

Number of figures: 4

Number of tables: 8

Number of appendices:3

Number of appendix pages:6

Number of references:10

Mentor: doc. dr. Branko Kavšek

UDC:

Keywords: Particle matter, PM10, weather influence, data mining

Abstract:

Following the expansion of technological and industrial development, air quality has deteriorated with human activity being the main contributor to pollutants in the atmosphere. The content of pollutants in the atmosphere leads to lasting negative effects on the environment and human health. Atmospheric aerosols are a mixture of solid and liquid particles suspended in the air, and vary in shape, size and chemical composition. Because of our desire to protect the environment, the air and its quality, measurements are carried out on the concentration of particle matter. The purpose of this thesis is to analyze the influence of weather factors on the concentration of suspended particle matter, known as PM10. Powders with a particle size of less than 10 micrometers are considered as particle matter PM10. Although the weather does not cause the formation of these particles, it is known that weather conditions (eg: wind, temperature) can be responsible for PM10 being transferred from its point of origin to places which are more densely populated. We will analyze data collected between 2007 and 2013 at sampling stations in Nova Gorica, in the Port of Koper and in Markovec (a densely populated area of Koper). For the analysis we will use machine learning algorithms (linear regression, regression trees, decision trees and decision rules) and try to create a predictive model that will show the impact of meteorological factors on the concentration of PM10 in the air. At the same time we will compare the predictive models for each measuring station.

Zahvala

Zahvaljujem se mentorju, doc. dr. Branku Kavšku, za strokovno pomoč in usmeritve pri izdelavi zaključne projektne naloge.

Zahvaljujem se mag. Damjanu Juriču za pomoč pri programu za uravnoteženje podatkovnih baz.

Zahvaljujem se tudi svoji družini za vsestransko pomoč na moji izobraževalni poti.

Hvala!

Kazalo

1	UVOD	1
1.1	Pregled vsebine	1
2	OPIS PROBLEMA	2
3	PRIPRAVA PODATKOV	3
3.1	Pretvorba v format .CSV	3
3.2	Razporeditev po letih	6
3.3	Diskretizacija izbranih atributov	7
3.4	Uravnoteženje razrednega atributa	7
3.5	ARFF zapis	9
4	MODELIRANJE	10
4.1	Uporabljeni algoritmi strojnega učenja	10
4.2	Regresijski algoritmi	10
4.2.1	Linearna regresija	10
4.2.2	M5P	10
4.3	Klasifikacijski algoritmi	11
4.3.1	ZeroR	11
4.3.2	Odločitveno drevo (algoritem J4.8)	11
4.3.3	Odločitvena pravila (algoritem PART)	12
4.4	Rezultati	12
5	INTERPRETACIJA REZULTATOV	15
6	ZAKLJUČKI IN NADALJNE DELO	16
	Literatura	18

Seznam kratic

<i>PM10</i>	Particle Matter 10 (prašni delci velikosti manjše od 10 mikrometrov)
<i>WEKA</i>	Waikato Environment for Knowledge Analysis (orodje za podatkovno rudarjenje)
<i>CSV</i>	Comma Separated Values (vrednosti ločene z vejico)
<i>ARFF</i>	Attribute-Relation File Format (podatkovni format programskega orodja WEKA)

1 UVOD

PM10 je kratica za particle matter 10, oziroma delci v zraku, v trdnem ali tekočem stanju, premera 10 mikrometrov ali manj. Ker je izvor PM10 zelo različen, je tudi kemijska struktura prašnih delcev zelo različna. Prašni delci lahko izhajajo direktno iz človeške uporabe, kot so izpušni plini motorjev, elektrarne in podobno ali pa lahko nastane v atmosferi kadar izpušni plini kot so SO_2 in NO_x reagirajo. Količina PM10 v zraku je pomembna saj gre za delce proti katerim človeško telo nima obrambnih mehanizmov, kar naj bi bilo povezano z velikim številom dihalnih bolezni.

V zaključni nalogi, bomo opisali aplikacijo algoritmov podatkovnega rudarjenja na problemu analize vpliva vremenskih faktorjev na vsebovanost prašnih delcev v zraku, z namenom preverjanja trenutnega stanja v kasneje definiranih krajih, ter najti morebitne povezave med vremenom faktori in količino prašnih delcev.

1.1 Pregled vsebine

V drugem poglavju bomo na kratko opisali problem s katerim se ukvarjamo. Katere podatke imamo, kje nastajajo ter nekatere probleme povezane z njimi.

V tretjem poglavju v celoti opišemo postopek ki nam je bil potreben za ustvarjanje pogojev za testiranje podatkov. Opišemo tudi razloge za uporabo nekaterih postopkov. V četrtem poglavju opišemo uporabljene algoritme ter izvedemo testiranje in ga ustrezno opišemo ter podpremo z najbolj pomembnimi podatki.

V petem poglavju opišemo naše ugotovitve ter razloge za le-te.

Šesto poglavje je namenjeno zaključku ter možnim izboljšavam opravljenega dela.

2 OPIS PROBLEMA

V primorski regiji imamo več merilnih postaj za merjenje vremenskih podatkov. Le nekatere od njih imajo tudi senzor za merjenje prašnih delcev. V zaključni nalogi smo se osredotočili na merilne postaje v Novi Gorici, v Luki Koper ter na Markovcu. Za merilno postajo v Luki Koper smo se odločili ker je Luka potencialni proizvajalec prašnih delcev (pretovarjanje različnih vrst sipkih tovorov kot npr. železova ruda). Markovec smo izbrali zato, ker je to gosto poseljeno področje, ki se nahaja v bližini Luke Koper. Novo Gorico smo izbrali kot referenčno merilno postajo, za potrebe primerjav. Na podatkih iz te merilne postaje, so delali analize predvsem za ugotavljanje ozona v zraku [7] [8] [9]. V zaključni nalogi, se bomo na merilne postaje sklicevali preprosto z Luka Koper, Markovec ter Nova Gorica. Poleg količine prašnih delcev v zraku, vse tri merilne postaje merijo tudi temperaturo, smer ter moč vetra. Te skupne podatke smo uporabili za preverjanje možne relacije med Luko Koper ter količino prašnih delcev v stanovanjskih delih Kopra, oziroma okolice merilne postaje na Markovcu. Med meritvami ki niso skupne v vseh treh merilnih postajah imamo podatek o padavinah ki je prisoten v Novi Gorici ter na Markovcu, medtem ko se v Luki Koper meri relativna vlaga, Markovška ter Novogoriška merilna postaja merita tudi količino ozona v zraku ter količino NO₂, Novogoriški meri tudi količino CO. Vse meritve v vseh laboratorjih so razporejene po datumih ter v urnem intervalu.

3 PRIPRAVA PODATKOV

3.1 Pretvorba v format .CSV

Vse tri merilne postaje imajo meritve shranjene na približno enak način, to je v formatu "<zaporedna številka>", "<datum>", "<ura>", "<senzor>", "<rezultat meritve>". Merilna postaja v Luki Koper ima tudi dodaten zapis "<datum ura>" postavljen za zapisom "<ura>". Tak format za podatkovno rudarjenje ni primeren, saj ne omogoči jasnega vpogleda v meritve niti za katero meritev gre. Ker so datoteke tako velike da je nemogoče ročno manipulirati vse podatke, smo se odločili implementirati program za dinamično pretvarjanje laboratorijske datoteke v datoteko s primernim formatom za uporabo v podatkovnem rudarjenju.

Primer meritve v Luki Koper 20. oktobra 2010 ob 20:00 uri, na enem od senzorjev.

```
"1043312", "2010 - 10 - 20", "20 : 00 : 00", "2010 - 10 - 20 20 : 00 : 00",  
"5", "11.4000000000"
```

Tabela 3.1 prikazuje kako izgleda celotna meritev (vsi senzori na merilni postaji) v

stolpničastem zapisu. Stolpci se postavijo namesto vejic v vrstičnem zapisu.

1043312	2010-10-20	20:00:00	2010-10-20 20:00:00	5	11.4000000000
1043313	2010-10-20	20:00:00	2010-10-20 20:00:00	3	2.9000000000
1043314	2010-10-20	20:00:00	2010-10-20 20:00:00	4	47.0000000000
1043315	2010-10-20	20:00:00	2010-10-20 20:00:00	6	58.0000000000
1043316	2010-10-20	20:00:00	2010-10-20 20:00:00	13	13.9000000000
1043317	2010-10-20	20:00:00	2010-10-20 20:00:00	7	1008.8000000000

Tabela 3.1: Primer zapisa meritev

Tabela 3.2 prikazuje katera številka predstavlja kateri senzor, oziroma kaj meri določen senzor.

Cifra	Vrednost
1	Temperatura vode
2	Vodostaj
3	Hitrost vetra
4	Smer vetra
5	Temperatura
6	Relativna vlaga
7	Pritisk
11	CO
12	Ozon O3
13	Delci PM10
14	Padavine
15	NO2
16	SO2

Tabela 3.2: Kodiranje senzorjev

Številom 8,9 in 10 ni dodeljen noben senzor, zato smo za ta števila določili: 8=zaporedna številka, 9=datum, 10=ura. Tako definiranje meritev ostane konstantno v

vseh treh laboratorjih. Specifično, tabela 3.1 opisuje meritve ob osmih zvečer, ki se vse zgodijo ob enakem trenutku. Prva vrstica pravi da je bila temperatura 11.4 stopinj celzija, druga da je bila hitrost vestra 2.9 metrov na sekundo, tretja da je veter pihal iz smeri ki se nahaja na kotu 47 stopinj (severovzhodnik), četrta vrstica pravi da je bila relativna vlaga 58 odstotna, peta vrstica pravi da je bila količina prašnih delcev enaka 13.9 na kubični decimeter zraka, zadnja vrstica pa nam pravi da je bil pritisk 1008.8 milibarov (ena atmosfera je 1013.25 milibarov). Količine NO₂, O₃ ter CO, ki se nahajajo bodisi v Novi Gorici bodisi pri obeh merilnih postajah, se merijo na enak način kot prašne delce, padavine se merijo v cm.

Ker zgoraj opisani format ni primeren za uporabo v podatkovnem rudarjenju, smo se lotili slednjih postopkov čiščenja ter popravljanja:

- a) Urni format, opisan v tretjem stolpcu, vsebuje tako meritve v polnih urah kot meritve v polovičnih urah, te meritve niso konstantne ampak se pojavljajo od nekega trenutka naprej ter se na nekem trenutku dokončno ustavijo. Poleg tega dejstva, velja tudi da so meritve o prašnih delcev v polovičnih urah kopija tistih v polnih. Zaradi tega, smo se odločili da meritve v polovičnih urah, v celoti odstranimo. To smo naredili tako da smo razdelili stolpec številka 3 v dva stolpca.
- b) S pomočjo sortiranja, smo odstranili vrstice s polurnimi meritvami.
- c) Stolpec z minutami ter sekundami smo v celoti odstranili ter sortirali po datumu ter uri.
- č) Odstranili smo prvi stolpec ki vsebuje zaporedno število ter četrti stolpec, ki vsebuje datum ter uro, saj sta že vsebovana.
- d) V tem trenutku imamo datoteko kjer vrstica izgleda tako:

2008 – 02 – 03, 20, 4, 244

torej datum, ura, število senzorja oziroma meritev, ki je opravljena ter v zadnjem stolpcu, vrednost same meritve. Nad to datoteko smo se lotili razširjanja v primerni format za naslednje korake. To smo naredili tako da smo zagnali program, ki nam je kot izhod kreiral novo datoteko, kjer je vsaka vrstica sestavljena iz 14 mestnega polja, vsako mesto v polju predstavlja eno meritev definirano po zgornjem standardu (13 delci PM₁₀, 5 temperatura ipd.), pretvorjena v niz. V kolikor meritev ni bila izvedena v tisti vrstici, se v polju nahaja vprašaj, v nasprotnem primeru pa se nahaja vrednost meritve. Za zgornjo vrstico, program

vrne slednjo vrstico:

?, ?, ?, ?, 244, ?, ?, ?, 2008 – 02 – 03, 20, ?, ?, ?, ?

- e) Po dobljeni datoteki, smo se lotili čiščenja nepotrebnih podatkov, oziroma odvzeli smo stolpce 1,2,3,11,12,13 saj noben od teh stolpcev nima nobenih meritev v merilni postaji Luka Koper.
- f) Pognali smo drug program, ki ima kot nalogo to da, stisne vrstice z istimi datumom in uro v eno, tako da dobimo novo datoteko kjer imamo vrstico sestavljeno iz vseh meritev ki so bile hkrati merjene.
- g) Datoteko odpremo, ter pojavitve nizov -999 ter -9999 spremenimo v vprašaj, saj so to kontrolne vrednosti v merilni postaji za označitev manjkajoče ali napačne meritve, največkrat zaradi pokvarjene opreme.
- h) Mejne vrednosti pri temperaturi smo popravili po konceptu interpoliranja, oziroma vzeli najbližjo temperaturo v istem dnevu, to je bilo potrebno v dveh primerih, vrednosti -48 ter +89 stopinj celzija.
- i) Napačne vrednosti pri smeri vetra smo interpolirali po istem principu, namreč nekatere smeri vetra so bile večje od 360 stopinj, kar je nemogoče.

Ostali dve bazi podatkov sta kreirani na podoben način, spreminja se le dolžina polj, saj imamo v drugačnih merilnih postajah drugačne meritve, neobstoj stolpca z datumom in uro, ter nepotrebnost čiščenja polurnih vrstic, saj se le-te pojavljajo samo v Luki Koper.

3.2 Razporeditev po letih

V želji po najbolj učinkovitih možnih pravilih, smo se lotili deljenja podatkovnih zbirk po letih, saj je podatek o letnici nepomemben za napovedovanje prihodnih količin PM10 v zraku kajti leto, za razliko od dnevov ter mescev, se v prihodnosti ne bo več pojavilo, kar pomeni, da je njegova informacija enaka nič. Leto 2013 smo izpustili, saj v njem ni bilo dovolj podatkov. V dveh merilnih postajah so podatki od leta 2007, v merilni postaji Luka Koper pa so meritve le od leta 2008 naprej, kar pomeni da smo skupno izluščili 17 podatkovnih baz. V teji fazi so vsi podatki zvezni. Ker linearna

regresija ter njej podobni algoritmi najbolje delujejo na zveznih podatkih, bodo podatki v trenutnem stanju uporabljeni za tako imenovane regresijske algoritme.

3.3 Diskretizacija izbranih atributov

Dva podatka sta bila določena kot primerna za proces diskretizacije, to sta smer vetra ter razredni atribut PM10. Za atribut smer vetra, velja da je merjen v stopinjah, kjer so stopinje razdeljene na urni način, oziroma tako da je ničla enaka vetru ki piha iz severa. Stopinje so od 0 do 360. V merilnih postajah na Markovcu ter v Novi Gorici je bila smer vetra že diskretizirana v 16 skupin. Za diskretizacijo v merilni postaji Luka Koper pa je bilo določenih 8 vrednosti, opisane v tabeli 3.3.

Diskretizirana Vrednost	Od(>)	Do(<=)
0	338	22
45	22	67
90	67	112
135	112	157
180	157	202
225	202	247
270	247	292
315	292	338

Tabela 3.3: Diskretizacija smeri vetra

Razredni podatek PM10 je bil diskretiziran iz več razlogov, prvi je ta da je splošna praksa diskretizirati razredni atribut za grajenje odločitvenih dreves, drugi je ta da v EU velja pravilo da če ima mesto več kot določeno število dnevov z vsaj eno meritvijo kjer je PM10 količina večja ali enaka 50, je količina PM10 previsoka. Zadnji razlog izhaja iz same neučinkovitosti merjenja ki izhaja iz izhlapevanja, v kar se ne bomo spuščali, ker ni smisel te zaključne naloge. Razredni atribut PM10 je bil diskretiziran po slednjem kriteriju: količine od 0 do brez 25 so bile označene kot „nizko“, od 25 do brez 50 kot „srednje“ ter vse višje količine, kot „visoko“.

3.4 Uravnoveženje razrednega atributa

Ker v osnovi, če štejemo posamezne primerke v razrednem atributu, imamo opravka z neuravnoveženo podatkovno bazo, obstaja možnost da je učinkovitost rezultatov so-

razmerna le temu. Z drugimi besedami, v naših zbirkah, se najbolj pogosto pojavlja vrednost „nizko” in zaradi tega so algoritmi bolj nagnjeni k pravilnosti v primerih, ko predvidijo da bo količina PM10 „nizko”. To se je izkazalo tudi kot dejanska težava, saj smo v fazi testiranja opazili, da je bilo veliko učinkovitih pravil za nizko količino PM10, medtem ko je bilo za visoko praktično nemogoče izluščiti karkoli z dovolj veliko natančnostjo.

Iz tega razloga smo se odločili narediti nove, uravnotežene, podatkovne baze. To smo naredili na sledeči način:

Najprej smo sortirali podatke na poglavi PM10 ter nato na podlagi mesca ter dneva. Drugi korak je bil razdeliti podatkovno bazo v tri datoteke na podlagi razrednega atributa PM10. S tem smo dobili tri datoteke, v katerih imamo podatke različne velikosti z istim razrednim podatkom, v katerem so podatki sortirani na podlagi mesca in dneva, ura pa je naključna. Ker smo želeli uravnotežene podatke, smo vzeli najmanjšo od treh datotek, ki se je vedno izkazala imeti vrednost PM10 = „visoko”, ter jo pustili nedotaknjeno. V ostalih dveh datotekah smo vzeli prvih k vrstic v vsakem dnevu, kjer je k odvisen od razlike v količini podatkov med tremi datotekami. Tak pristop smo uporabili, ker smo na tak način obdržali naključnost podatkov, saj ne vemo koliko je podatek ura v prvih k vrsticah posameznega dneva (ure so se pomešale zaradi predhodnega sortiranja). Dobljene zbirke podatkov združimo. Rezultat je podatkovna baza velika približno trikratnik vseh podatkov z razrednim atributom „visoko”. Tak postopek je bil uporabljen za vse podatkovne baze, spreminjalo se je le število k po katerem je bilo izbrano število zapisov na dan, to pa je bilo izbrano naknadno za vsako bazo na podlagi testiranja, oziroma tako, da se je čim bolj približalo številu „visokih” vrednosti.

3.5 ARFF zapis

Ker smo želeli uporabiti neuravnotežene podatkovne baze kot testne podatke za pravila kreirana nad uravnoteženimi, smo morali vsako datoteko podkrepiti z svojo pretvorbo v .arff format, saj program WEKA zahteva enako glavo obeh datotek. Slika 3.1 prikazuje glavo podatkovne baze iz merilne postaje Luka Koper leta 2010, v svoji .arff datoteki.



```
lu_2010_full.arff ✕
@relation 'lu_2010_full-weka.filters.unsupervised.attribute.Remove-R2,6,10'

@attribute Hitrost_Vetra numeric
@attribute Temperatura numeric
@attribute Relativna_Vlaga numeric
@attribute Pritisk numeric
@attribute Mesec numeric
@attribute Dan numeric
@attribute Ura numeric
@attribute Smer_VetraD {45.0,90.0,135.0,270.0,315.0,0.0,225.0,180.0}
@attribute PM10D {nizko,srednje,visoko}

@data
0.9,13.7,60,996.1,5,15,21,45.0,nizko
10.1,2.4,44,1005.8,3,10,13,45.0,nizko
10.2,2.5,43,1006.4,3,10,12,45.0,nizko
```

Slika 3.1: Primer glave dokumenta tipa .arff

4 MODELIRANJE

Nad neuravnoteženimi podatkovnimi bazami smo uporabili algoritme Linearna regresija ter M5P, podatki so bili zvezni. Nad uravnoteženimi bazami, smo uporabili algoritme PART ter J4.8, da bi algoritma zgradila model odločanja oziroma pravila, potem pa smo za testiranje uporabili predhodne neuravnotežene podatkovne baze. Vsi algoritmi opisani v tem poglavju, so implementirani v odprtokodnem programskem paketu za podatkovno rudarjenje WEKA [5] [6].

4.1 Uporabljeni algoritmi strojnega učenja

V nadaljevanju, bomo predstavili uporabljene algoritme, njihov koncept ter razlog za njihovo izbiro. Algoritmi so razdeljeni v dve kategoriji, regresijski ter klasifikacijski algoritmi. Prva kategorija je bila uporabljena nad zveznimi atributi, medtem ko je druga imela diskretiziran razred ter atribut smer vetra.

4.2 Regresijski algoritmi

4.2.1 Linearna regresija

Linearna regresija, je algoritem ki nad množico vrednosti najde linearno funkcijo ki se najbolj prilagaja.

4.2.2 M5P

Algoritem M5P [4] je priredba znanega Quinlanovega algoritma M5 (P kot prime) [3]. Algoritem kreira poseben primer regresijskih dreves, oziroma regresijska drevesa z linearnimi modeli kot liste. Izbran je bil kot primer učinkovitega regresijskega algoritma nad zveznimi podatkovnimi bazami z nenormaliziranimi podatki, kar zagotovo omejuje njegovo učinkovanje.

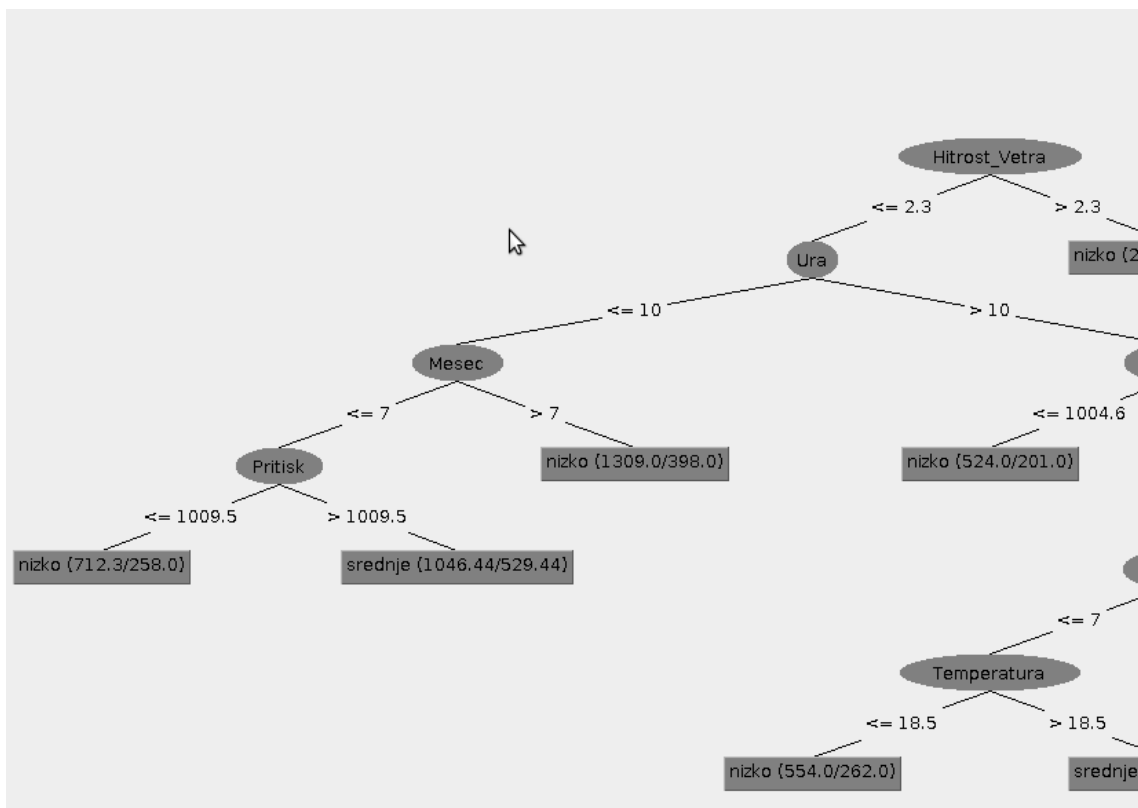
4.3 Klasifikacijski algoritmi

4.3.1 ZeroR

Algoritem ZeroR, je standarden klasifikacijski algoritem v strojnem učenju, ki odloča na podlagi večinskega razreda. Z drugimi besedami, tisti razred ki je skupno najbolj pogost, ga algoritem uporabi za določanje. Ta pristop je seveda zelo neučinkovit, zato velja algoritem kot najnižja sprejemljiva točnost. To torej pomeni, da če od katerega algoritma dobimo pravilnost nižjo kot pri ZeroR, je rezultat nesprejemljiv in je zelo verjetno nekaj močno narobe v našem postopku.

4.3.2 Odločitveno drevo (algoritem J4.8)

Ta algoritem je implementacija znanega algoritma C4.5 [2] v programskem jeziku Java ki je uporabljena v weki. Temelji na gradnji odločitvenega drevesa po konceptu največje entropije, oziroma ob vsakem koraku določi atribut ki najbolj razdeli drevo (kreira najnižje drevo). Drevo je možno učinkovito stiskati z namenom preglednosti, slika 4.1 prikazuje primer takega drevesa. Drevo predstavlja desni izsek odločitvenega drevesa za leto 2010 v merilni postaji Luki Koper ob faktorju združevanja $M = 300$. Zgrajena so bila bolj učinkovita drevesa, najbolj primerno, je dodano v prilogi B.



Slika 4.1: Primer odločitvenega drevesa

4.3.3 Odločitvena pravila (algoritem PART)

Algoritem PART [1] kreira odločitvena pravila, tipa če X potem Y, kjer je X odločitveni pogoj Y pa vrednost razrednega atributa ki pravilo določi. To naredi tako da uporablja koncept „deli in vladaj” pri kreaciji odločitvenih dreves, nato eliminira uporabljene zapise in rekurzivno kreira nova pravila, dokler ne ostaja več zapisov za uporabiti. Rezultat takega pristopa so izjemno berljiva ter intuitivna pravila za zunanega opazovalca, kar je v našem primeru posebej pomembno saj je cilj naloge nuditi uporabniku razumljive razloge za naše trditve. Slika 4.2 prikazuje primer treh pravil za leto 2010 v merilni postaji na Luki Koper ob faktorju združevanja $M = 20$.

```
Ura <= 11 AND  
Temperatura > 10.3; visoko (30.0/14.0)  
  
Ura > 11 AND  
Mesec > 3 AND  
Dan <= 21 AND  
Pritisk > 1004.7 AND  
Hitrost_Vetra > 0.1 AND  
Dan > 14; nizko (58.0/31.0)  
  
Ura > 11 AND  
Mesec > 3 AND  
Mesec <= 9 AND  
Dan <= 16 AND  
Hitrost_Vetra <= 1.5; srednje (39.0/20.0)
```

Slika 4.2: Primer pravil algoritma PART

4.4 Rezultati

Kot že predhodno povedano, testiranja so bila opravljena nad zveznimi podatkovnimi bazami in sicer z linearno regresijo ter z algoritmom M5P. Nad neuravnoteženimi diskretiziranimi podatkovnimi bazami, smo uporabili algoritme ZeroR, J4.8 ter PART. Nad uravnoteženimi diskretiziranimi podatkovnimi bazami nismo uporabili algoritma ZeroR ker bi bila nihanja pri pravilnosti nerealistično velika, oziroma direktno odvisna od večinskega razreda. Se pravi, za primer, če je baza imela eno vrednost „nizko” večjo od ostalih je pravilnost lahko bila tudi višja od 60%, hkrati pa bi z dodajanjem dveh vrednosti „visoko” dosegali pravilnosti okoli 10%. Zaradi tega se predpostavlja da je večinski razred 34% (saj so baze uravnotežene), medtem ko je bila v neuravnoteženih nihajoča med 45% in 55%. Zaradi te razlike v pravilnosti algoritma ZeroR, je bila zahtevana minimalna pravilnost v ostalih algoritmih veliko manjša. Tabeli 4.1 ter 4.2 prikazujeta najbolj indikativne podatke. Prva prikazuje učinkovanje algoritma PART tako da predstavi pravilnost odločitvenih pravil ter njihovo število za dano leto,

oziroma njegovo stisnjeno verzijo (faktor stiskanja je v oklepajih). Faktor stiskanja je bil izbran tako da se je pravilnost zmanjšala čim manj ob doseganju čim večje intuitivnosti pri pravilih. Druga tabela predstavlja korelacijo vremenskih spremenljivk s količino prašnih delcev v zraku. Vse klasifikacijske točnosti (v tabelah 4.1 in 4.2) so ocenjene z metodo desetkratnega prečnega preverjanja [10].

	Markovec
2007	53.7641/51.9147(10)/109/41
2008	58.059/58.0835(10)/192 /74
2009	53.537/49.0999 (10)/167 /54
2010	58.1882/52.8995 (10)/175 /56
2011	59.7906/ 53.5961(10)/215 /77
2012	56.9097/ 56.6195(10)/167 /91
	Nova Gorica
2007	64.9207 /59.6175 (10)/ 216/27
2008	65.9189/ 60.742(20)/297/40
2009	55.141/ 56.4779(10)/156 /46
2010	61.8934 / 56.7793(10)/197 /44
2011	65.727/ 64.7997(10)/212 /79
2012	53.227/ 50.6058(10)/225 /86
	Luka Koper
2008	68.235 / 63.4077(10) / 209 / 89
2009	59.7885/59.0682(10)/133 / 49
2010	58.3442 /55.8736(10) /164 /56
2011	48.2361/49.899(10) /166 / 65
2012	50.8849 /46.1931(10) /164 /53

Tabela 4.1: Pravilnost algoritma PART ter število pravil

Markovec	Linearna Regresija	M5P
2007	0.3661	0.5997
2008	0.3486	0.4363
2009	0.462	0.5548
2010	0.5354	0.6787
2011	0.4998	0.572
2012	0.4566	0.5886
Nova Gorica	Linearna Regresija	M5P
2007	0.5404	0.634
2008	0.5998	0.659
2009	0.3037	0.3463
2010	0.5288	0.5547
2011	0.6088	0.7864
2012	0.4798	0.7852
Luka Koper	Linearna Regresija	M5P
2008	0.1028	0.7937
2009	0.394	0.7002
2010	0.3166	0.538
2011	0.457	0.6632
2012	0.3817	0.6324

Tabela 4.2: Faktor povezanosti pri Linearni Regresiji in M5P

5 INTERPRETACIJA REZULTATOV

Izkazalo se je, po pričakovanjih, da sta algoritma M5P ter Linearna regresija, neprimerni za naš primer, saj sta imela težave z uteževanjem posameznih atributov. Kljub temu da je dobljena formula povsem neintuitivna ter neuporabna, pa smo dobili zanimivo razliko med dvema algoritmoma glede faktorja povezanosti ter relativne napake. Izkazalo se je, da je algoritem M5P veliko bolj sposoben najti korelacijo med atributi ter da obstaja velika razlika med posameznimi leti, kar je mogoče pripisati vremenskim faktorjem prej kot človeškim.

V diskretiziranih neuravnoteženih bazah, smo dobili največjo pravilnost pri algoritmu J4.8, kljub temu se je algoritem PART dobro odrezal ter zaostal za okoli dva odstotka povprečno. Izkazalo se je, da obstaja velika razlika v pravilnosti določanja, med posameznimi merilnimi postajami. Merilna postaja Luka Koper je v vsakem letu bila tista z najvišjimi pravilnostmi, ne glede na algoritem. V Luki Koper so bile pravilnosti vedno nad 70%, ter dvakrat celo pri 80%. Drugi dve merilne postaje sta imeli pravilnosti bolj nizke, med 62% pa do 75%. Po subjektivni oceni avtorja, je razlog takšne razlike v pravilnosti, povzročena od oddaljenosti merilne postaje Luka Koper od mestnega jedra ter z njim povezanimi človeškimi viri prašnih delcev. Z drugimi besedami, ker je število povročiteljev manjše (ne glede na dejansko količino) si je mogoče razložiti da je lažje predvideti prihodnje količine.

Namen te zaključne naloge, je bil nuditi čim bolj razumljivo ter intuitivno predstavo faktorjev ki vplivajo na prašne delce, za to nalogo smo se odločili uporabiti pravila kreirana z algoritmom PART nad uravnoteženimi podatkovnimi bazami.

6 ZAKLJUČKI IN NADALJNE DELO

Po pričakovanjih so rezultati analize vremenskih vplivov na koncentracijo prašnih delcev v zraku za merilne postaje Nova Gorica, Markovec in Luka Koper pokazali, da vreme nima velikega vpliva na koncentracijo teh delcev v zraku. Glavni vzrok temu vidimo predvsem v tem, da vremenski vplivi kot taki niso neposredni povzročitelj nastanka prašnih delcev, ampak le potencialni prenašalec le-teh.

Dodatna težava, s katero smo se soočili pri analizi meritev, je bila v veliki neuravnovešenosti odvisne spremenljivke (diskretizirane vrednosti količine prašnih delcev). Visokih koncentracij prašnih delcev je bilo - po pričakovanjih - malo, zato so v prvotnih modelih (pred uravnoteženjem razreda) prevladovala pravila, ki so napovedovala predvsem nizke in pa deloma srednje koncentracije prašnih delcev, kar pa ni bil namen raziskave. Razredno spremenljivko smo tako morali uravnotežiti kar pa je posledično prineslo nižje klasiﬁkacijske točnosti modelov.

Poleg tega so se, zaradi različnih senzorjev, na različnih merilnih postajah beležile različne količine in smo morali, če smo hoteli primerjati modele na različnih merilnih postajah med seboj, morali izbrati podmnožico količin, ki so se beležile na vseh treh postajah. Problem je nastal, ker so se ravno tiste količine, ki so bile merjene samo na določeni merilni postaji, pokazale za „pomembne“.

Kljub vsem težavam in nizkim napovednim točnostim celotnih modelov, pa smo vseeno uspeli izluščiti nekaj „zanimivih“ napovednih pravil z relativno visoko napovedno točnostjo, ki napovedujejo visoko koncentracijo prašnih delcev v zraku. Naše ugotovitve so tako pokazale (glej Prilogo C), da so z visoko koncentracijo prašnih delcev v zraku povezani sledeči atributi: za merilno postajo Markovec predvsem NO₂, v Novi Gorici poleg NO₂ še CO, v Luki Koper (kjer so se merili res samo vremenski vplivi), pa ni nekega prevladujočega faktorja, ampak nastopajo specifične kombinacije izbranih vremenskih faktorjev. Poleg tega se v Novi Gorici in na Markovcu v modelih velikokrat pojavlja tudi ozon kot vplivni faktor - to je bilo tudi pričakovati, saj so ravno prašni delci eden glavnih povzročiteljev povišanih koncentracij ozona v zraku.

V zaključni nalogi smo se omejili skoraj izključno na klasiﬁkacijske napovedne modele. Regresijske smo samo poskusili in še to na neuravnoveženih podatkih ter nenormal-

iziranih atributih. V nadaljnjem delu bi bilo potrebno te regresijske modele „pognati“ na ustrezno uravnoteženih in normaliziranih podatkih. Tako bi bili rezultati primerljivi z modeli, ki smo jih dobili v tej nalogi (klasifikacijski napovedni modeli). Poleg tega bi bilo zbrati še več podatkov, ki jih trenutno vse merilne postaje ne merijo, saj so se samo vremenski podatki izkazali za neustrezne, če želimo napovedovati količino prašnih delcev v zraku z relativno visoko napovedno točnostjo.

Literatura

- [1] E. FRANK in I. WITTEN Generating Accurate Rule Sets Without Global Optimization. V *Fifteenth International Conference on Machine Learning*, 1998, 144-151.
- [2] R. QUINLAN, *Programs for Machine Learning*, Morgan Kaufmann, 1993. (*Citirano na strani 12.*)
- [3] R. QUINLAN, *Learning With Continuous Classes*, World Scientific. 2006 (*Citirano na strani 11.*)
- [4] Y. WANG in I. WITTEN Inducing Model Trees for Continuous Classes. V *9th European Conference on Machine Learning*, 1996, 128-137. (*Citirano na strani 10.*)
- [5] E. FRANK in I. WITTEN, M. Hall, Practical Machine Learning Tools and Techniques, Third Edition. Morgan Kaufmann, 2011 (*Citirano na strani 10.*)
(*Citirano na strani 10.*)
- [6] E. FRANK in I. WITTEN, M. Hall, G. Holmes, B. Pfahringer, P. Reutemann The WEKA Data Mining Software: An Update SIGKDD Explorations, 2009 (*Citirano na strani 10.*)
- [7] D. ŠUC in D. VLADUŠIČ I. Bratko. V *Qualitatively faithful quantitative prediction*, Proceedings of the eighteenth International Joint Conference on Artificial Intelligence, 2003. 1052-1057 (*Citirano na strani 2.*)
- [8] D. ŠUC in I. BRATKO. Improving numerical prediction with qualitative constraints Lecture notes in computer science, Lecture notes in artificial intelligence, vol. 2837, 2003. 1052-1057 (*Citirano na strani 2.*)
- [9] D. ŠUC in I. BRATKO. AI Magazine. vol 24, no. 4, 2003. 107-119 (*Citirano na strani 2.*)
- [10] R. KOHAVI. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, Morgan Kaufmann, 1995. 1137-1143 (*Citirano na strani 13.*)

Priloge

A: REZULTATI NA NEURAVNOTEŽENIH PODATKOVNIH BAZAH

Tabele A.1, A.2 ter A.3 opisujejo izmerjeno točnost pri podanem algoritmu, glede na ostale podatke. Z drugimi besedami, opisuje odstotek primerov, ki so bili pravilno klasificirani za izbrano lezo in algoritem. Tabela A.1 Opisuje merilno postajo Nova Gorica, tabela A.2 merilno postajo Markovec, tabela A.3 pa merilno postajo Luka Koper.

	2007	2008	2009	2010	2011	2012
ZeroR	49.2719	49.698	50.0608	50.9515	45.5861	45.9376
J4.8	66.8768	66.9296	74.8055	76.5699	74.6414	72.7197
PART	65.4858	65.4505	71.4876	72.8473	71.5752	71.4506

Tabela 1: Merilna postaja Nova Gorica

	2007	2008	2009	2010	2011	2012
ZeroR	59.3777	60.0661	59.8884	57.6904	51.367	54.4916
J4.8	71.0183	68.1679	69.181	70.9806	67.1921	67.1261
PART	69.7563	68.4739	70.0685	70.0473	66.2315	65.3004

Tabela 2: Merilna postaja Markovec

	2008	2009	2010	2011	2012
ZeroR	49.698	50.0608	50.9515	45.5861	45.9376
J4.8	70.3835	80.086	70.5903	80.2596	78.0335
PART	69.2684	79.354	71.2792	78.7875	77.594

Tabela 3: Merilna postaja Luka Koper

B: PRIMER ODLOČITVENEGA DREVESA

Slika B.1 prikazuje stisnjeno odločitveno drevo, zgrajeno z algoritmom J4.8, nad neuravnoteženo podatkovno bazo iz leta 2012 za merilno postajo Luka Koper. Faktor stiskanja je 130. Levi del oklepajev označuje število testiranih primerov, desni pa število napačno klasificiranih primerov.

J48 pruned tree

```
-----
Pritisk <= 1004.1: nizko (834.38/105.45)
Pritisk > 1004.1
|  Temperatura <= 21.21
|  |  Hitrost_Vetra <= 2.001667
|  |  |  Dan <= 18
|  |  |  |  Ura <= 17
|  |  |  |  |  Mesec <= 6: nizko (909.73/256.42)
|  |  |  |  |  Mesec > 6
|  |  |  |  |  |  Relativna_Vlaga <= 73.823333: nizko (283.68/41.17)
|  |  |  |  |  |  Relativna_Vlaga > 73.823333
|  |  |  |  |  |  |  Temperatura <= 16.415: nizko (373.89/101.22)
|  |  |  |  |  |  |  Temperatura > 16.415: srednje (133.32/51.32)
|  |  |  |  |  Ura > 17
|  |  |  |  |  |  Relativna_Vlaga <= 93.123333: nizko (408.68/172.68)
|  |  |  |  |  |  Relativna_Vlaga > 93.123333: srednje (162.27/73.0)
|  |  |  |  Dan > 18
|  |  |  |  |  Dan <= 26
|  |  |  |  |  |  Mesec <= 9
|  |  |  |  |  |  |  Mesec <= 3
|  |  |  |  |  |  |  |  Mesec <= 2
|  |  |  |  |  |  |  |  |  Relativna_Vlaga <= 82.443333: nizko (132.9/63.67)
|  |  |  |  |  |  |  |  |  Relativna_Vlaga > 82.443333: visoko (146.99/40.75)
|  |  |  |  |  |  |  |  Mesec > 2: srednje (160.95/65.0)
|  |  |  |  |  |  |  Mesec > 3: nizko (301.26/49.47)
|  |  |  |  |  |  |  Mesec > 9: srednje (474.0/262.0)
|  |  |  |  |  Dan > 26: nizko (601.89/210.42)
|  |  |  |  Hitrost_Vetra > 2.001667: nizko (1610.52/299.79)
|  |  Temperatura > 21.21
|  |  |  Relativna_Vlaga <= 44.825
|  |  |  |  Temperatura <= 29.888333: nizko (658.67/173.14)
|  |  |  |  Temperatura > 29.888333
|  |  |  |  |  Pritisk <= 1010.578333: srednje (187.76/82.48)
|  |  |  |  |  Pritisk > 1010.578333: nizko (132.54/68.23)
|  |  |  Relativna_Vlaga > 44.825
|  |  |  |  Temperatura <= 25.913333
|  |  |  |  |  Relativna_Vlaga <= 58.845: nizko (383.55/160.67)
|  |  |  |  |  Relativna_Vlaga > 58.845: srednje (399.62/126.02)
|  |  |  |  Temperatura > 25.913333: srednje (348.41/105.89)
```

Number of Leaves : 20

Size of the tree : 39

C: REZULTATI ALGORITMA PART

Priloga C opisuje določena pravila izluščena s pomočjo algoritma PART. Modeli niso popolni, temveč obsegajo le pravila ki so bila označena kot najbolj „zanimiva“. Pravilo je bilo označeno kot zanimivo, kadar je bila relacija med vsemi meritvami in napačnimi meritvami 7:1 ali višja. Zaželjena je bila relacija 10:1 ali višja. Zaradi težavnosti določanja pravil za srednjo količino prašnih delcev, smo si dovolili majhna odstopanja. Oklepaji v pravilu označujejo omenjeno relacijo. Priloga vsebuje pravila za vsa leta.

MERILNA POSTAJA: MARKOVEC

- Ura > 7 AND NO₂ > 66.8: visoko (97.96/11.95)
- Temperatura <= 24.2 AND NO₂ > 31.55 AND Smer Vetra = 315.0 AND Mesec <= 2: visoko (10.21/1.21)
- Smer Vetra = 315.0 AND Mesec <= 4: visoko (16.27/2.67)
- NO₂ > 18.8 AND Ozon <= 7.8 AND Mesec > 10: visoko (49.35/4.69)
- NO₂ > 48.5 AND Mesec <= 4 AND Temperatura > 2.6: visoko (261.1/22.73)
- Mesec <= 3 AND Padavine <= 0.1 AND NO₂ > 46.65: visoko (155.92/13.92)
- NO₂ > 51.35 AND Hitrost Vetra <= 1.2: visoko (180.52/24.6)
- NO₂ > 27.8 AND Padavine > 0 AND Padavine <= 0.2: srednje (16.61/2.75)
- Ura <= 7 AND Smer Vetra = 112.5 AND NO₂ > 6.85: srednje (20.53/3.7)
- Ura <= 6 AND Smer Vetra = 135.0 AND Ozon <= 45.8: srednje (19.73/2.78)
- NO₂ <= 25.4 AND Temperatura <= 16.7 AND Smer Vetra = 202.5 AND Mesec <= 4: nizko (17.68/2.25)
- Ura <= 7 AND Ozon > 13.7 AND NO₂ <= 15.25 AND Ozon <= 96.85 AND Smer Vetra = 112.5: nizko (48.66/1.49)

- Ura ≤ 7 AND Ozon ≤ 95.4 AND Smer Vetra = 157.5 AND Dan ≤ 27 AND NO2 ≤ 14.7 : nizko (67.67/11.78)
- Ura ≤ 7 AND NO2 ≤ 31.05 AND Smer Vetra = 90.0: nizko (52.13/7.0)
- NO2 ≤ 20.15 AND Ura ≤ 7 AND Ura > 0 AND Mesec > 3 AND Temperatura ≤ 19.9 AND Smer Vetra = 135.0 AND Mesec ≤ 9 : nizko (27.5/1.0)
- Ura ≤ 7 AND Mesec > 3 AND Ozon > 52.05 AND NO2 ≤ 13.15 AND Temperatura ≤ 18.5 AND Mesec ≤ 10 : nizko (161.59/20.94)
- NO2 ≤ 24.35 AND Ura ≤ 7 AND Smer Vetra = 67.5 AND Hitrost Vetra > 3 : nizko (25.81/3.02)
- Temperatura ≤ 6.5 AND Ozon ≤ 3.15 : nizko (18.24/1.89)
- Ura ≤ 7 AND NO2 ≤ 21.9 AND Mesec > 11 : nizko (41.02/1.47)
- Temperatura ≤ 27.3 AND Hitrost Vetra ≤ 4.8 AND Smer Vetra = 90.0 AND NO2 ≤ 13.8 : nizko (11.69/0.41)
- Smer Vetra = 67.5 : nizko (12.63/1.54)

MERILNA POSTAJA: NOVA GORICA

- CO > 1.41 : visoko (176.92/17.87)
- CO > 0.455 AND Mesec > 8 : visoko (13.55/2.21)
- NO2 > 38.05 AND Hitrost Vetra ≤ 0.7 AND Temperatura > 16.2 AND Ozon ≤ 26.7 : visoko (47.31/5.2)
- Ura ≤ 7 AND Ozon ≤ 0.9 : visoko (30.68/4.3)
- Padavine ≤ 0 AND Ura > 7 AND NO2 > 63.45 AND Hitrost Vetra ≤ 1.3 : visoko (334.21/26.89)
- Temperatura > 28.7 : visoko (21.76/1.97)
- CO > 0.56 AND Mesec ≤ 11 AND Temperatura > 10 AND Temperatura ≤ 14.9 : visoko (100.09/15.48)
- CO > 1.015 AND NO2 > 33.85 : visoko (485.98/84.55)
- Temperatura ≤ 12.6 AND NO2 > 41.45 AND Ura > 17 : visoko (374.65/39.02)

- $\text{NO}_2 > 28.65$ AND $\text{Mesec} \leq 3$ AND $\text{Mesec} > 1$ AND $\text{Ura} \leq 15$: visoko (182.04/16.87)
- $\text{Mesec} > 2$ AND $\text{Ozon} \leq 7.9$ AND $\text{Ura} \leq 5$: visoko (28.89/2.8)
- $\text{NO}_2 > 32.35$ AND $\text{Mesec} \leq 3$ AND $\text{Ura} > 18$: visoko (206.66/13.48)
- $\text{NO}_2 > 32.35$ AND $\text{Mesec} \leq 3$ AND $\text{CO} > 0.905$ AND $\text{Mesec} > 1$: visoko (81.34/6.55)
- $\text{Ozon} \leq 15.65$ AND $\text{Ura} > 9$ AND $\text{NO}_2 > 24.25$ AND $\text{Hitrost Vetra} \leq 0.7$: visoko (188.3/28.56)
- $\text{Hitrost Vetra} \leq 0.8$ AND $\text{CO} > 0.64$: visoko (48.75/3.93)
- $\text{Ura} \leq 7$ AND $\text{NO}_2 \leq 20.65$ AND $\text{Temperatura} > 19.9$: srednje (20.52/3.1)
- $\text{Mesec} > 4$ AND $\text{Ozon} > 134.7$ AND $\text{Ozon} \leq 150.55$: srednje (23.15/2.46)
- $\text{Mesec} > 4$ AND $\text{Ozon} > 25.3$ AND $\text{Ozon} \leq 131.3$ AND $\text{NO}_2 \leq 35.3$ AND $\text{Temperatura} \leq 30.1$ AND $\text{Smer Vetra} = 270.0$ AND $\text{Hitrost Vetra} \leq 1.3$: srednje (12.85/2.25)

MERILNA POSTAJA: LUKA KOPER

- $\text{Hitrost Vetra} \leq 1.915$ AND $\text{Mesec} > 9$ AND $\text{Temperatura} > 11.415$: visoko (32.0/3.0)
- $\text{Hitrost Vetra} \leq 0.50678$: visoko (104.61/11.41)
- $\text{Temperatura} > 25.291667$ AND $\text{Pritisk} \leq 1014.255$ AND $\text{Mesec} > 8$ AND $\text{Pritisk} > 1012.63$: visoko (17.07/0.07)
- $\text{Mesec} \leq 2$ AND $\text{Dan} > 23$: visoko (128.0)
- $\text{Ura} \leq 8$ AND $\text{Mesec} > 2$ AND $\text{Ura} > 2$ AND $\text{Relativna Vlaga} \leq 48$: nizko (76.33/6.0)
- $\text{Hitrost Vetra} \leq 1.4$ AND $\text{Mesec} \leq 2$ AND $\text{Pritisk} > 1025.1$ AND $\text{Pritisk} \leq 1033$ AND $\text{Ura} > 8$: visoko (44.59)
- $\text{Smer Vetra} = 270.0$ AND $\text{Mesec} \leq 10$ AND $\text{Hitrost Vetra} > 1.3$: visoko (12.06/0.06)
- $\text{Hitrost Vetra} > 1.2$ AND $\text{Smer Vetra} = 315.0$ AND $\text{Hitrost Vetra} > 1.8$: visoko (22.19/2.19)

- Ura > 10 AND Relativna Vlaga > 34 AND Pritisk > 1010.4 AND Hitrost Vetra <= 1.8 AND Temperatura <= 26.9 AND Mesec > 1 AND Temperatura > 9 AND Mesec <= 5: visoko (133.5/9.75)
- Smer Vetra = 135.0 AND Ura <= 8: nizko (25.0/3.0)
- Mesec <= 5 AND Ura <= 2 AND Relativna Vlaga > 56: visoko (45.0/4.0)
- Mesec <= 5 AND Ura > 8 AND Pritisk > 1001 AND Mesec > 3 AND Pritisk <= 1008.5: visoko (24.0)
- Pritisk > 1009.1 AND Mesec <= 9 AND Mesec > 8: visoko (36.0)
- Hitrost Vetra <= 0.4 AND Mesec > 10 AND Pritisk > 1006.5: visoko (115.0/19.0)
- Mesec > 3 AND Mesec <= 9 AND Hitrost Vetra > 0.6 AND Ura > 15 : nizko (26.0/2.0)
- Relativna Vlaga > 81 AND Mesec > 3 AND Ura <= 8: nizko (10.0)
- Relativna Vlaga > 81 AND Mesec <= 7 AND Ura > 1 AND Pritisk > 1011.1 : visoko (23.0)
- Relativna Vlaga > 71 AND Ura <= 0: visoko (11.0)
- Mesec <= 3 AND Hitrost Vetra <= 2 AND Ura > 11: visoko (268.0/26.0)
- Hitrost Vetra <= 1.5 AND Mesec <= 3 AND Ura <= 4 : visoko (88.0/6.0)
- Smer Vetra = 90.0 AND Pritisk <= 1014 AND Mesec > 1 AND Hitrost Vetra <= 1.5 AND Pritisk > 1009.4: srednje (24.04/3.04)
- Mesec > 4 AND Pritisk > 1018.973333: srednje (24.23/2.1)
- Mesec > 3 AND Hitrost Vetra > 2.313333 AND Smer Vetra = 90.0: nizko (37.36/3.23)
- Relativna Vlaga <= 24 AND Hitrost Vetra > 2.7: nizko (23.13)
- Mesec > 4 AND Ura <= 9 AND Ura > 2 AND Smer Vetra = 90.0 AND Temperatura <= 20.4 AND Hitrost Vetra > 0.7 AND Pritisk <= 1021.2: nizko (97.86/3.0)