Fakulteta za Računalništvo in Informatiko

Večna pot 113, Ljubljana

1. Seminarska naloga pri predmetu UMETNA INTELIGENCA

Prof. dr. Igor Kononenko  
as. dr. Peter Vračar

Avtorja:

Simon Korošec, 63160171

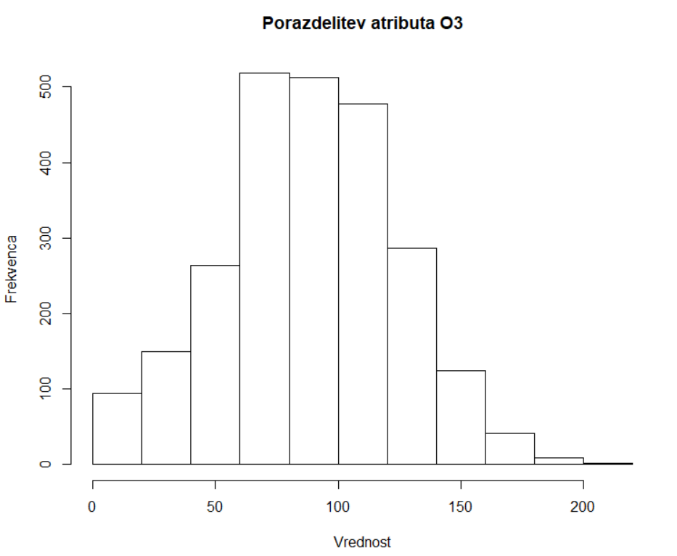
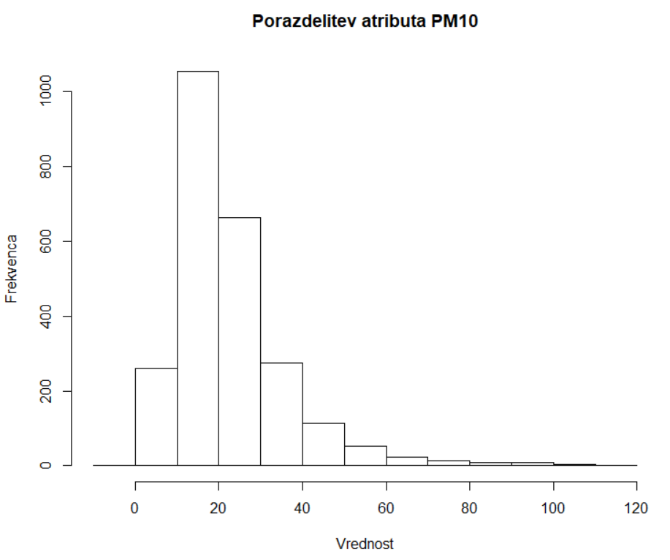
Gašper Smerkolj, 63160285

# Priprava podatkov

Iz atributa datum sva najprej razbrala mesec in nato sva lahko določila letni čas meritve. Mesec sva nato odstranila. Odstranila sva tudi atribut Glob\_sevanje\_min ker so vse njegove vrednosti enake in nam zaradi tega ne nudi dodatnih informacij.

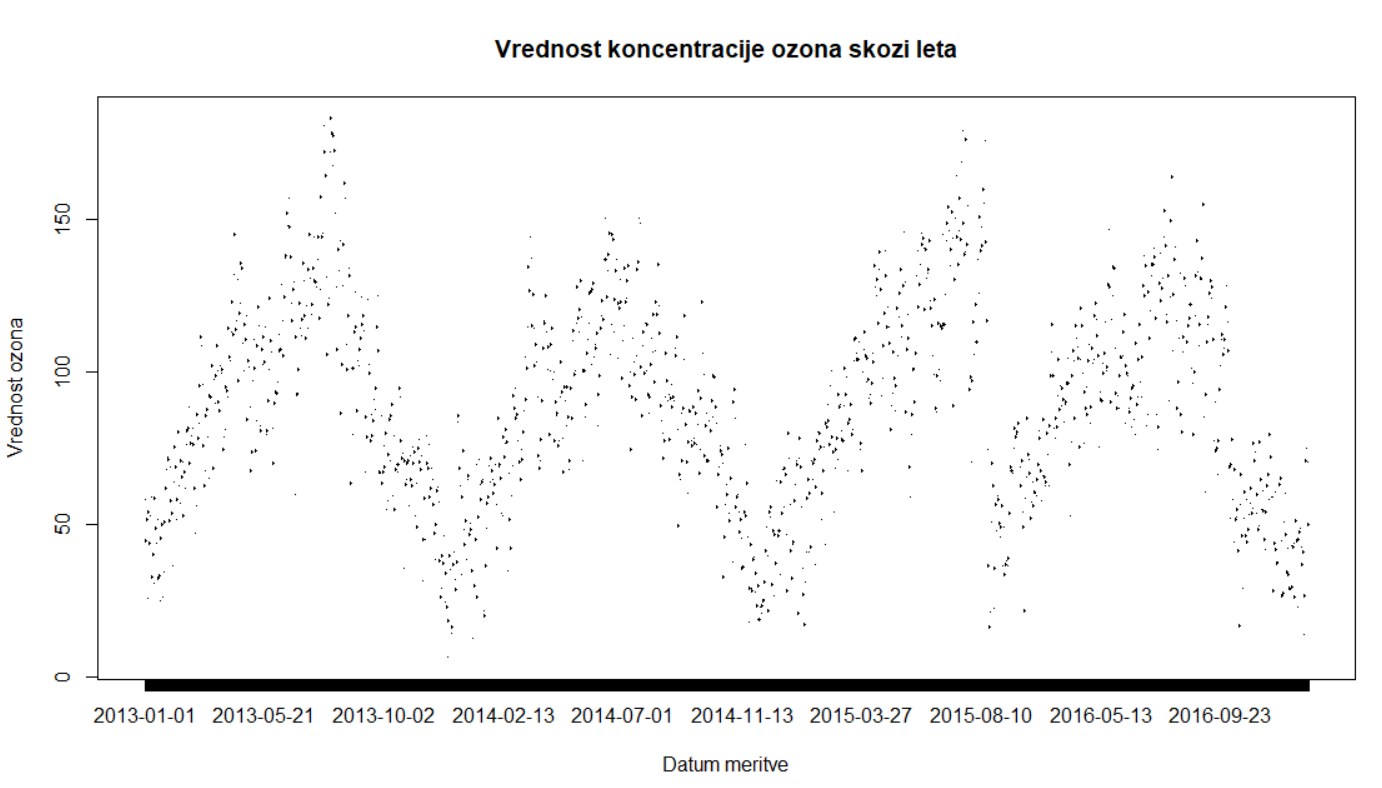
Vizualizacija podatkov

Iz prvih dveh grafov vidimo, da sta oba ciljna atributa porazdeljeni približno normalno.

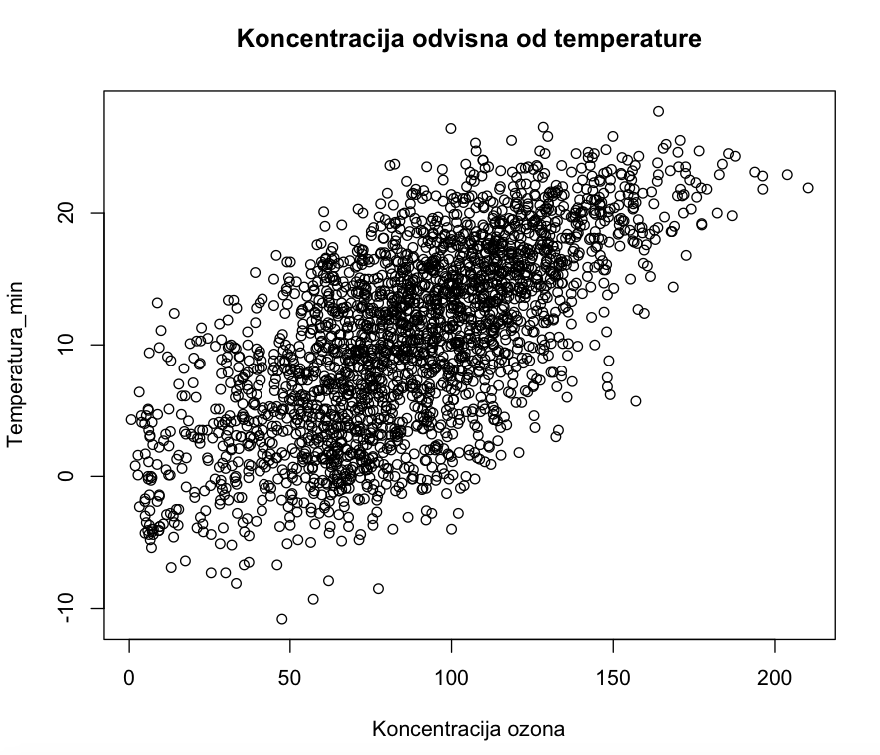


Graf 1 in 2: Porazdelitev ciljnih atributov PM10(levo) in O3(desno).

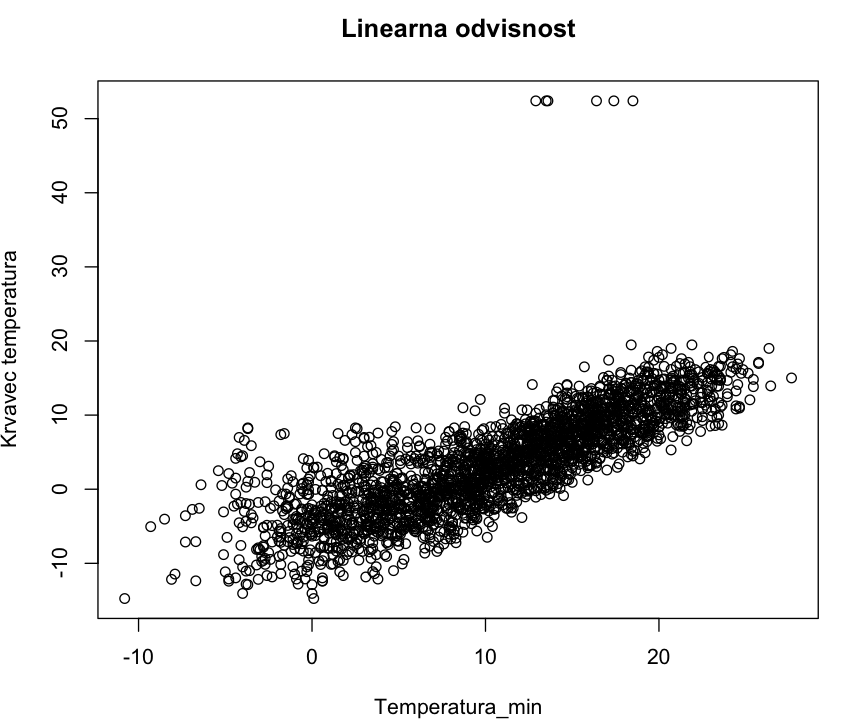
Naslednji graf prikazuje kako se koncentracija ozona približno polovico leta povečuje nato pa se zmanjša, ta vzorec je viden skozi vsa leta.



Graf prikazuje odvisnost med atributoma Temperatura\_min, ter Koncentracijo ozona, le ta je linearna.



Graf, ki prikazuje odvisnost med Krvavec\_min, ter Temperatura\_min in tako dobimo še naslednji atribut ki se delno povezuje z Koncentracijo ozona.



Klasifikacija

Za klasifikacijo atributa PM10 sem izdelal štiri različne modele in vsakega testiral na treh različnih formula. Uporabil sem naslednje modele:

* K – najbližjih sosedov (5, 10, 15, 20, 25),
* Glasovanje (uporabil sem modele "tree", "rf", "rfNear" in "knn" zgrajene z knjižnico CORElearn),
* Bagging in
* Umetna nevronska mreža.

Uporabil sem naslednje formule:

* Celotna množica,
* Atributi, ki jih je predlagala funkcija wrapper in
* Najbolje ocenjeni atributi po metodi ReliefFequalK.

Modele sem nato testiral po metodi prečnega preverjanja in dobil naslednje klasifikacijske točnosti:

|  |  |
| --- | --- |
| Formula 1 | |
| Metoda | Točnost |
| KNN | K-15: 0.9043617 |
| Glasovanje | 0.9075794 |
| bagging | 0.9015211 |
| umetna nevronska mreža | 0.9023585 |
|

|  |  |
| --- | --- |
| Formula 2 | |
| Metoda | Točnost |
| KNN | K-15: 0.9019545 |
| Glasovanje | 0.9075624 |
| bagging | 0.9055796 |
| umetna nevronska mreža | 0.8950826 |
|

|  |  |
| --- | --- |
| Formula 3 | |
| Metoda | Točnost |
| KNN | K-10: 0.9047974 |
| Glasovanje | 0.905973 |
| bagging | 0.9067794 |
| umetna nevronska mreža | 0.9128271 |
|

Klasifikacijska točnost večinskega razreda je znašala 0.868414.

Modele za atribut O3 sem izdelal z naslednjimi metodami:

* Metoda naključnih gozdov,
* Bootsing metoda,
* Metoda podpornih vektorjev.

Formule sem pridobil na isti način kot za atribut PM. S prečnim preverjanjem sem nato dosegel naslednje klasifikacijske točnosti:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Formula 1 | |  | Formula 2 | |
| Metoda | Točnost |  | Metoda | Točnost |
| Random Forests | 0.8212146 |  | Random Forests | 0.8022466 |
| Boosting | 0.7117439 |  | Boosting | 0.8002556 |
| SVM | 0.8026336 |  | SVM | 0.787336 |

|  |  |
| --- | --- |
| Formula 3 | |
| Metoda | Točnost |
| Random Forests | 0.7998175 |
| Boosting | 0.797437 |
| SVM | 0.8006426 |

Klasifikacijska točnost večinskega razreda je znašala 0.609328.

Regresija

Za regresijo točnost atributov PM10, ter O3, sem uporabil štiri različne metode na treh različnih formulah. Uporablil sem naslednje modele:

* CoreModel(1=mean predicted value, 2=median predicted value, 3=linear by MSE, 4=linear by MDL, 5=linear as in M5, 6=kNN, 7=Gaussian kernel regression, 8=locally weighted linear regression)
* Regresijska drevesa
* Nevronska mreža
* Linearni model

Uporabil sem naslednje formule:

* Celotna množica,
* Atributi, ki jih je predlagala funkcija wrapper in
* Najbolje ocenjeni atributi po metodi ReliefFequalK.

Točnosti s podanimi modeli za atribut PM10

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Formula 1 | CoreModel |  | Formula 2 | CoreModel |  | Formula3 | CoreModel |
| Metoda |  |  | Metoda |  |  | Metoda |  |
| 1 | 0.7899559 |  | 1 | 0.8174941 |  | 1 | 0.7785541 |
| 2 | 0.8099541 |  | 2 | 0.7822657 |  | 2 | 0.7821141 |
| 3 | 0.8090228 |  | 3 | 0.8184151 |  | 3 | 0.7656336 |
| 4 | 0.863832 |  | 4 | 0.8153686 |  | 4 | 0.7753401 |
| 5 | 0.7479881 |  | 5 | 0.7633269 |  | 5 | 0.7422536 |
| 6 | 0.6894159 |  | 6 | 0.7169512 |  | 6 | 0.7033621 |
| 7 | 0.7433226 |  | 7 | 0.7405943 |  | 7 | 0.7229701 |
| 8 | 1.127.945 |  | 8 | 1.248.134 |  | 8 | 2.314.778 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| lm.model |  |  |  | Nevronske |  |  | Regresijska drevesa |
| Metoda |  |  | Formula2 | 0.8979192 |  | Formula 1 | 0.8274773 |
| Mae | 7.961.599 |  | Formula1 | 0.8609758 |  | Formula 2 | 0.7975253 |
| Rmae | 0.831104 |  |  |  |  | Formula 3 | 0.8278224 |

Točnosti s podanimi modeli za atribut O3

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Formula 1 | CoreModel |  | Formula 2 | CoreModel |  | Formula 3 | CoreModel |
| Metoda |  |  | Metoda |  |  | Metoda |  |
| 1 | 0.5352372 |  | 1 | 0.5583526 |  | 1 | 0.5556219 |
| 2 | 0.5289665 |  | 2 | 0.5420835 |  | 2 | 0.5461032 |
| 3 | 0.5466939 |  | 3 | 0.5180777 |  | 3 | 0.4979781 |
| 4 | 0.523285 |  | 4 | 0.5294955 |  | 4 | 0.503557 |
| 5 | 0.4728432 |  | 5 | 0.5098632 |  | 5 | 0.5006464 |
| 6 | 0.4600129 |  | 6 | 0.4774612 |  | 6 | 0.4776968 |
| 7 | 0.4863083 |  | 7 | 0.4850018 |  | 7 | 0.4774028 |
| 8 | 0.6953557 |  | 8 | 1.140.306 |  | 8 | 0.830242 |
|  |  |  |  |  |  |  |  |
|  | Nevronske |  |  | Regresijska drevesa |  | lm.model |  |
| Formula 2 | 0.9999996 |  | Formula 1 | 0.6622979 |  | mae | 14 |
| Formula 2 | 0.6986606 |  | Formula 2 | 0.6115082 |  | rmae | 0.5029431 |
|  |  |  | Formula 3 | 0.6149047 |  |  |  |

Pri metodi za CoreModel, sem za formule 2 ter 3 uporabil zmanjšano število atributov, s katerimi sem včasih dobil boljše rezultate, drugič pa slabše saj se je model preveč prilagodil samo tem atributom.

Metoda nevronske mreža, sem uporabil malo manj ponovitev, št. Uteži pa je bilo odvisno od velikosti mreže.