Seminarska naloga

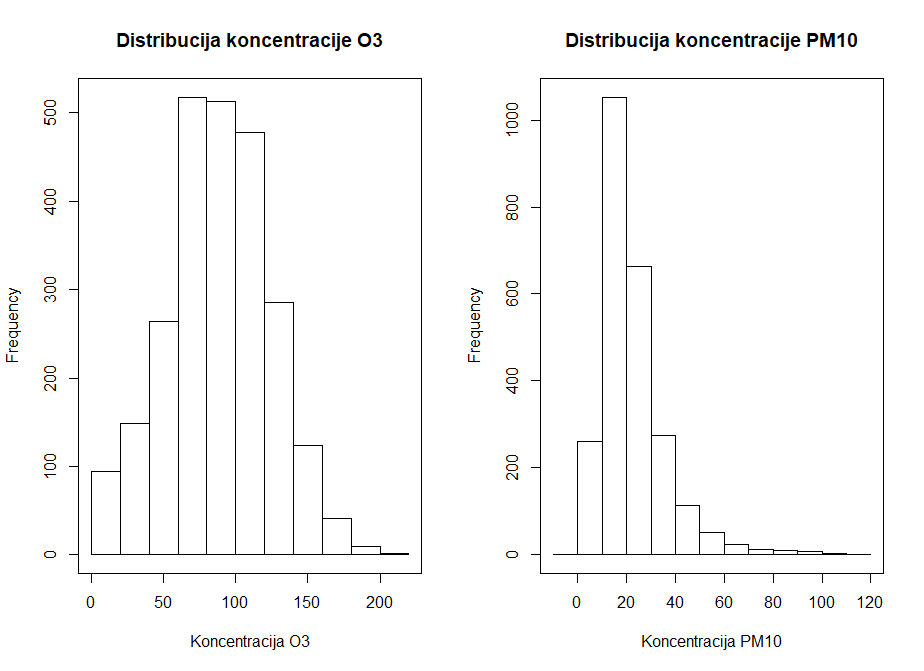
pri predmetu Umetna Inteligenca

Avtorja: Simon Šegatin (63160294), Anže Košir (63160174) Ljubljana, 11. 12. 2017

# Uvod

Seminarsko nalogo sva začela z internetno raziskavo o variaciji koncentracije O3 in PM10 čez različna časovna obdobja. To nama je pokazalo, da tako koncentracija trdih delcev PM10 kot ozona O3 najbolj niha na letni ravni. Nekateri članki so omenjali tudi variacijo ozona glede na to ali je bila meritev opravljena na deloven dan ali med vikendom. Na podlagi teh informacij sva se nato odločila, da atribut Datum odstraniva ter dodava atribute Year, Season, Month in Weekday. Ob pregledu obstoječih atributov sva opazila, da je atribut Glob\_sevanje\_min vedno enak 0, zato sva ga odstranila. Dodala sva tudi atributa O3Class in PM10Class, ki predstavljata razrede koncentracije O3 in PM10.

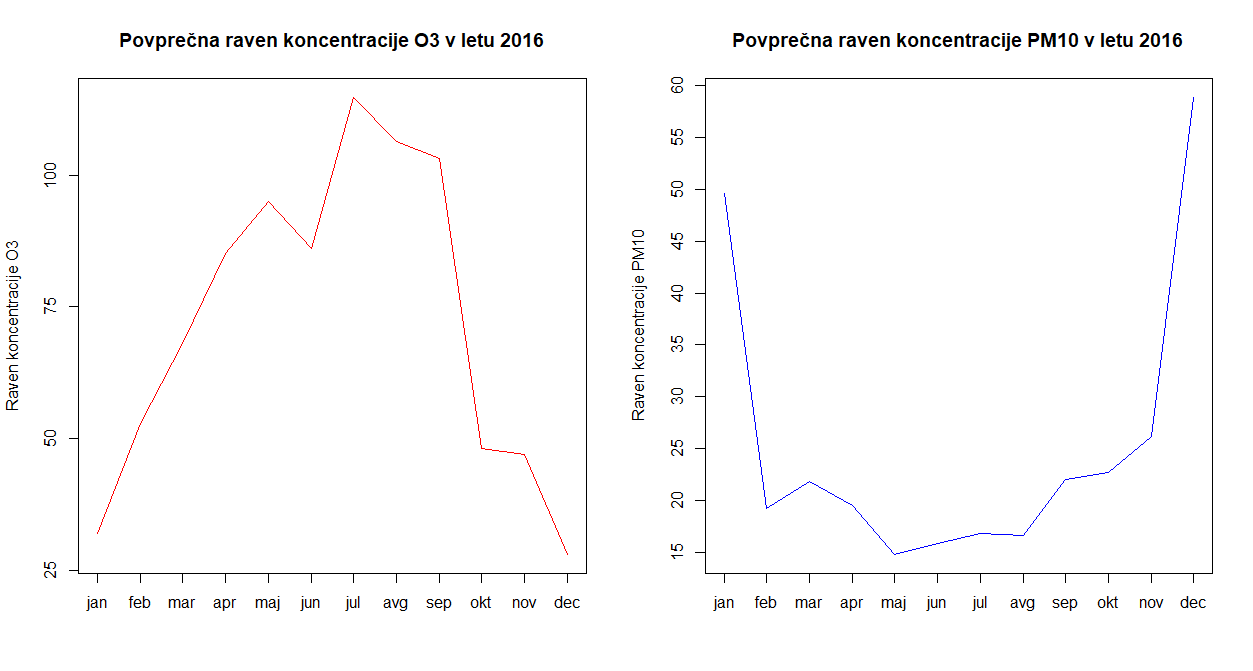
# Vizualizacija podatkov



Slika - Histograma koncentracij O3 in PM10

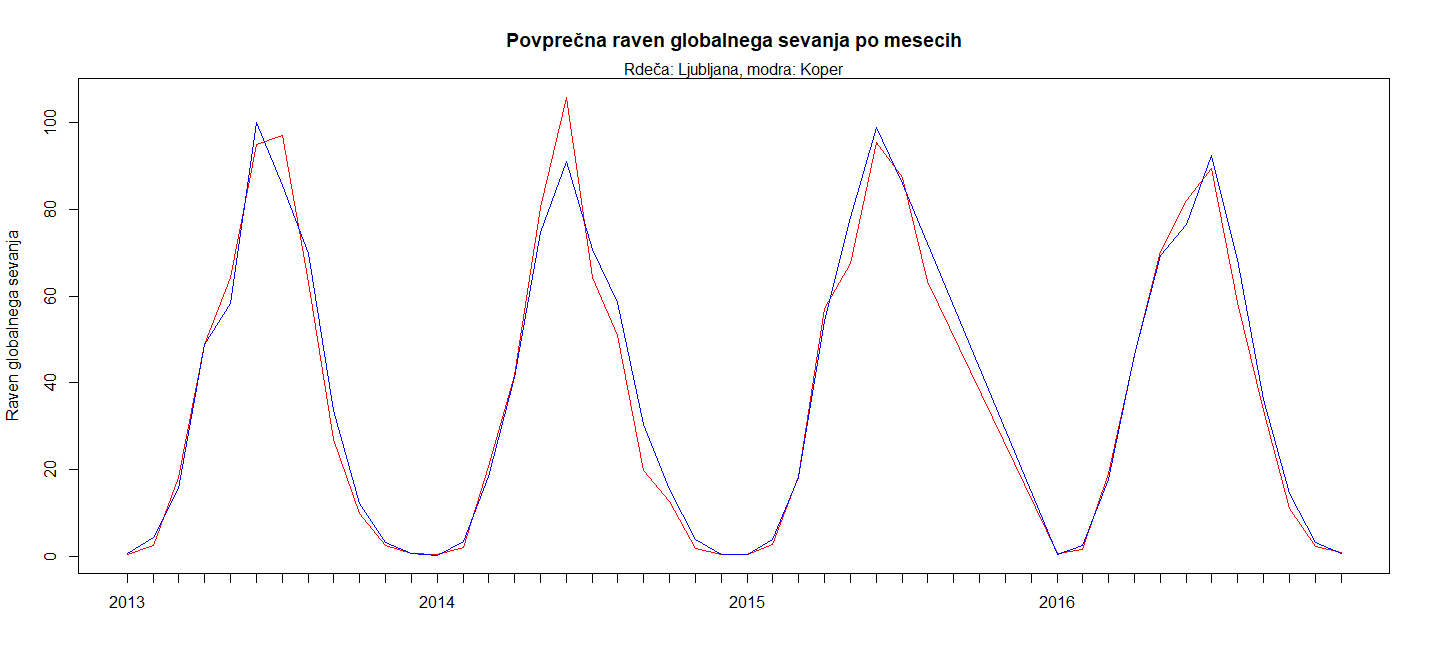
Histograma koncentracij O3 in PM10 na sliki 1 pokažeta, da večina meritev koncentracij O3 spada v srednji razred (med 60 in 120), koncentracij PM10 pa v nizki razred (pod 35).

Med vizualizacijo podatkov sva opazila, da koncentracija O3 in PM10, kot je pokazala najina raziskava, res najbolj niha glede na letni čas (slika 2). Izkazalo se je tudi, da je koncentracija O3 višja v toplejših mesecih, koncentracija PM10 pa v hladnejših.



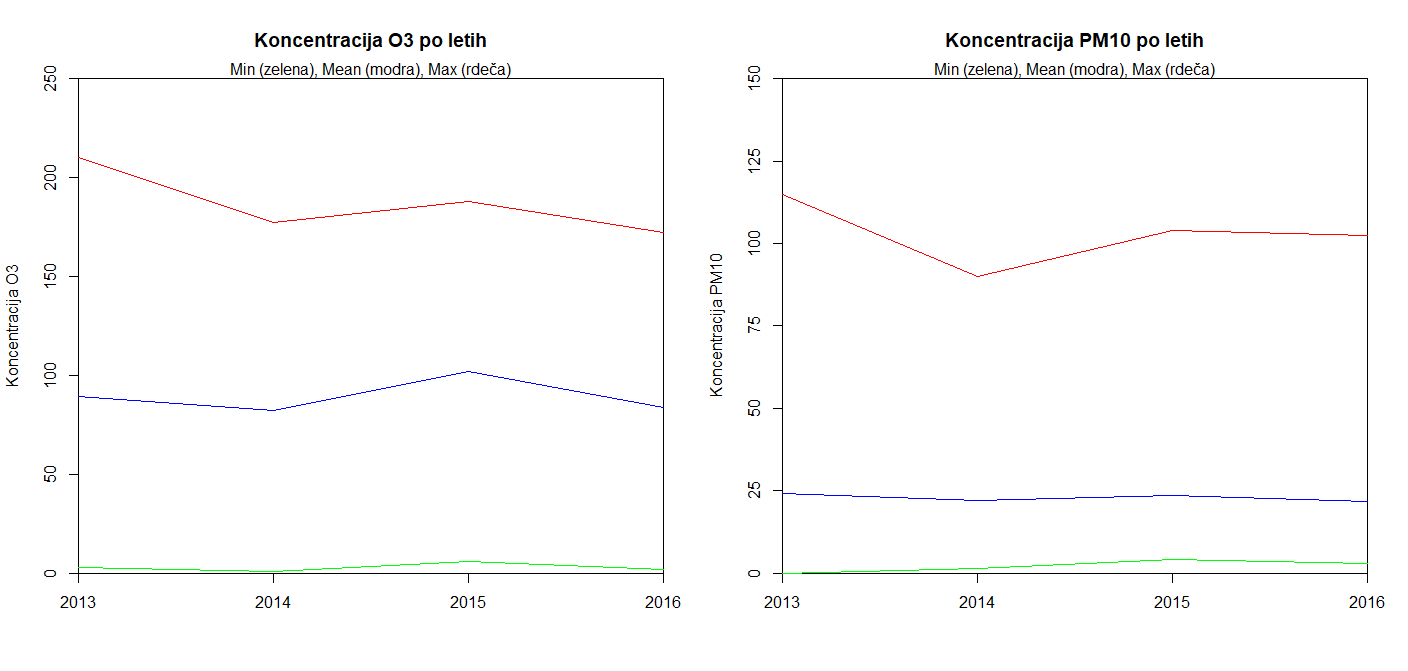
Slika - Povprečni ravni koncentracije O3 in PM10 v letu 2016

Najin sum, da je globalno sevanje najbolj odgovorno za koncentracijo ozona je potrdil graf na sliki 3, iz katerega je lepo razviden vzorec nihanja sevanja, ki je zelo podoben tistemu na sliki 2. Opazna je tudi rahlo višja raven globalnega sevanja na merilni postaji v Kopru. Edina izjema je junij 2014, ko je bila na postaji v Ljubljani izmerjena višja vrednost.



Slika - Raven globalnega sevanja skozi leta (Ljubljana in Koper)

Iz rezultatov zadnjega grafa je razvidno, da koncentracija O3 in PM10 skozi leta ne niha preveč, z izjemo rahlega upada najvišje izmerjene letne koncentracije tako O3 kot PM10 v letu 2014 (slika 4).



Slika - Najmanjša, najvišja in povprečna koncentracija O3 ter PM10 v letih 2013-2016

# Klasifikacija

Za klasifikacijo sva izbrala modele Odločitveno drevo, naivni Bayes in K-najbližjih sosedov, za izbiro atributov pa sva uporabila funkcijo wrapper z učilnice. Množico podatkov sva razdelila v učno, z meritvami v letih 2013 in 2014, ter testno, z meritvami v letih 2015 in 2016. Tako je imela učna množica 1340 vrstic, testna pa 1138.

## Odločitveno drevo

Za izgradnjo modela s pomočjo odločitvenega drevesa sva uporabila knjižnico rpart in atribute Glob\_sevanje\_max, Temperatura\_lokacija\_max, Temperatura\_Krvavec\_mean, Pritisk\_min, Sunki\_vetra\_max, Pritisk\_max, Hitrost\_vetra\_max, Vlaga\_max, Padavine\_mean, Temperatura\_Krvavec\_min, Padavine\_sum ter Pritisk\_mean za koncentracijo O3 in Temperatura\_lokacija\_max, Hitrost\_vetra\_min, O3Class, Month, Year, Padavine\_sum, Temperatura\_lokacija\_mean, Postaja, Sunki\_vetra\_min, Pritisk\_max, Temperatura\_Krvavec\_mean, Pritisk\_min, Sunki\_vetra\_mean ter Padavine\_mean za koncentracijo PM10, katere je predlagala funkcija wrapper. Model je pravilen razred koncentracije O3 napovedal v dobrih 75% (0.7513181), razred koncentracije PM10 pa v dobrih 91% (0.9156415) primerov.

Za primerjavo sva zgradila tudi model s knjižnico CORElearn, ki pa se je izkazal za malo slabšega s pravilno napovedjo koncentracije O3 v dobrih 73% (0.7311072), koncentracije PM10 pa dobrih 90% (0.9077329) primerov.

## Naivni Bayes

Model z algoritmom naivni Bayes sva prav tako zgradila s knjižnico CORElearn, atribute pa ponovno izbrala s funkcijo wrapper, ki je predlagala Glob\_sevanje\_max, Pritisk\_max, Month, Postaja, PM10Class, Year, Sunki\_vetra\_mean, Temperatura\_Krvavec\_min ter Vlaga\_min za koncentracijo O3 in Month, Postaja, Sunki\_vetra\_max, Padavine\_sum, Vlaga\_max, Padavine\_mean, Year ter Pritisk\_max za koncentracijo PM10. V primeru koncentracije O3 se je model izkazal za rahlo boljšega od prvega z dobrih 76% (0.7618629) pravilnih napovedi, v primeru koncentracije PM10 pa za malo slabšega z slabih 89% (0.8892794) zadetkov.

## K-najbližjih sosedov

Za izgradnjo modela z algoritmom KNN sva uporabila knjižnico CORElearn in atribute Month + Temperatura\_lokacija\_max, Vlaga\_max, Sunki\_vetra\_max, Pritisk\_max, Sunki\_vetra\_min, Padavine\_mean, Glob\_sevanje\_mean, Weekday, Temperatura\_Krvavec\_min, Sunki\_vetra\_mean + Temperatura\_Krvavec\_mean, Hitrost\_vetra\_max, Vlaga\_min, Padavine\_sum, PM10Class, Temperatura\_Krvavec\_max, Glob\_sevanje\_max ter Season za koncentracijo O3 ter Temperatura\_lokacija\_max, Temperatura\_Krvavec\_max, Sunki\_vetra\_min, Hitrost\_vetra\_min, Temperatura\_lokacija\_min, Temperatura\_Krvavec\_mean, Glob\_sevanje\_mean, Postaja, Padavine\_mean, Glob\_sevanje\_max, Padavine\_sum, Temperatura\_lokacija\_mean, Temperatura\_Krvavec\_min, Hitrost\_vetra\_mean, Hitrost\_vetra\_max, O3Class, Pritisk\_mean, Year, Month, Vlaga\_min, Vlaga\_max, Pritisk\_min, Sunki\_vetra\_max ter Sunki\_vetra\_mean za koncentracijo PM10. Te sva ponovno izbrala s pomočjo funkcije wrapper. Nato sva za zanko for(k in 1:100) poiskala najbolj ustrezen k, v primeru koncentracije O3 19, koncentracije PM10 pa 30. Modela s tema k-jema sta se izkazala za najboljša z natančnostjo dobrih 78% (0.7838313) pri klasifikaciji O3 in dobrih 91% (0.9147627) pri klasifikaciji PM10.

# Regresija

Regresija

## Model 1

Model 1

## Model 2

Model 2

## Model 3

Model 3

# Zaključek