

Izveštaj

Analiza podataka – PM_{2.5} čestica

Nikola Šljivkov, IN56-2018, sljivkov@gmail.com

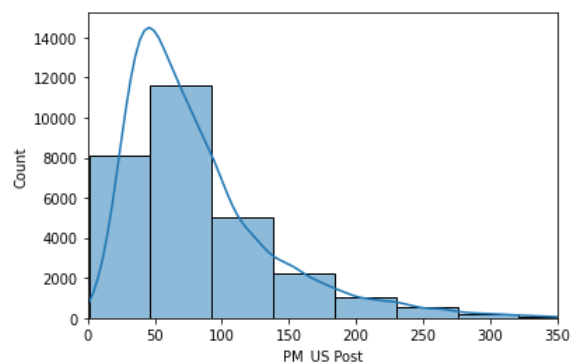
I Opis baze podataka

Ovaj izveštaj se bavi analizom podataka PM_{2.5} čestica (u daljem tekstu PM čestice) u gradu Čeng du. U bazi postoji 52584 (pedeset dve hiljade petsto osamdeset četiri) uzorka. Svaki uzorak sadrži podatke izmerene u toku jednog sata. Obeležja koje sadrži jedan uzorak su: redni broj uzorka, godina, mesec, dan u mesecu, sat u danu, godišnje doba, koncentracija PM čestica u vazduhu, temperatura, temperatura rose, vlažnost vazduha, vazdušni pritisak, kombinovan pravac vetra, kumulativna brzina vetra, količina padavine na sat i kumulativne padavine. Kategorički podaci su: godina, mesec, dan, sat, godišnje doba, pravac vetra dok su ostali podaci numerički.

Analizom ovih podataka probaćemo da nađemo koji od ovih atributa imaju najveći uticaj na porast PM čestica u vazduhu kao i koji atributi mogu da nam pomognu u pokušaju da smanjimo ili usporimo rast tih čestica. PM čestice kod čoveka mogu da izazovu astmu, hronični bronhitis kao i srčana oboljenja.

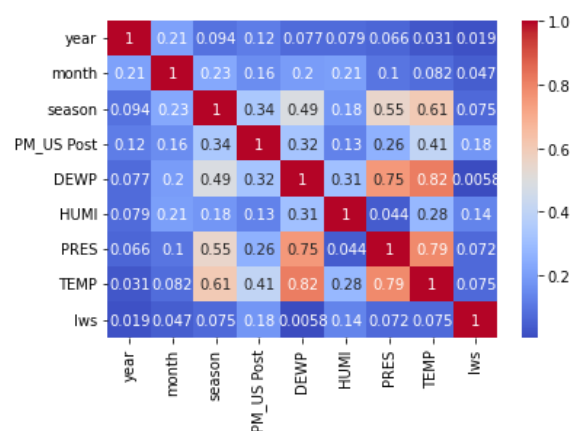
II Analiza podataka

Prilikom analiziranja obeležja izbačene su kolone sa podacima o PM česticama na drugim mestima, kako je zadatak i nalogao, zatim je izbacena i kolona za redni broj kako je to identifikaciono obeležje i ono ne ulazi u dalju analizu. Nedostajući podaci u kolonama za temperaturu, vlažnost vazduha, pritisak i temperaturu rose popunjeni su metodom popunjavanja unapred dok su podaci koji su se odnosili na kumulativnu brzinu vetra popunjeni sa 0 pod pretpostavkom da vetar nije duvao u tim satim i da je to razlog zašto ne postoje podaci. Svi uzorci koji nisu imali podatak u koloni PM čestica izbačeni su.



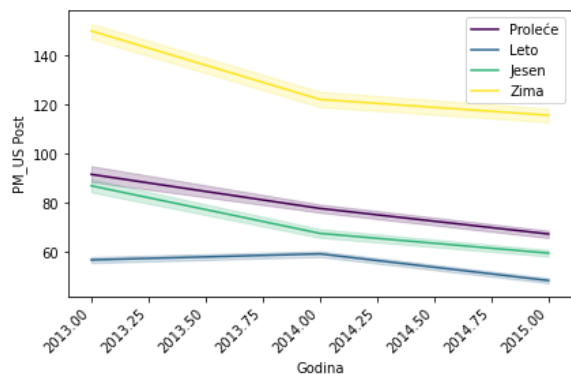
Slika 1. Raspodela PM čestica

Na slici 1 vidimo da se 88.02% uzoraka ima koncentraciju čestica od 0 do 150 ug/m3.



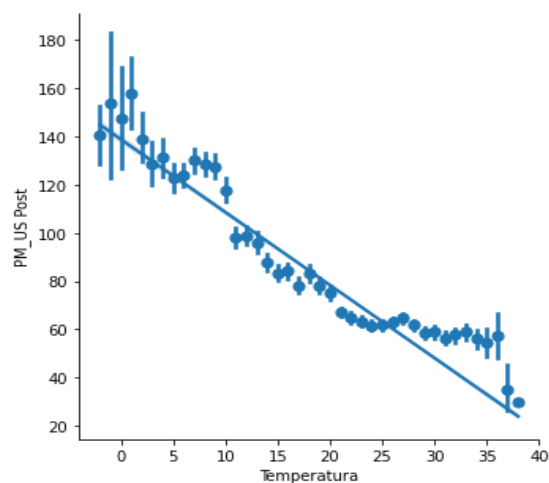
Slika 2. Koeficijenti korelacije između svih obeležja

Vidimo korelacije između obeležja koje smo izabrali i obeležja PM_US Post nad kojim treba da se odradi predviđanje. Kriterijum za izbor obeležja koje ćemo koristiti za predikciju je da apsolutna vrednost koeficijenta bude veća od 0.1. Obeležja koja su izbačena jer nisu imala uticaj tj. koeficijent korelacije sa obeležjem PM_US Post manji od 0.1 su dani u mesecu, sati u danu, kombinovan pravac vetra, količina padavine na sat i kumulativne padavine.



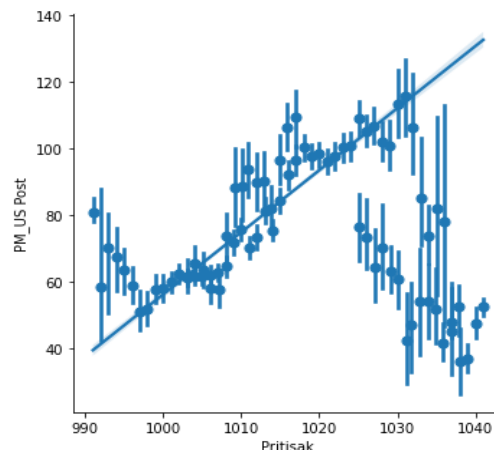
Slika 3. Zavisnost koncentracije PM čestica od godišnjeg doba tokom godina

Vidimo kako se prosečna količina čestica kretala od 2013. do 2015. godine. Takođe vidimo da je došlo do pada posle 2013. godine i to u svim godišnjim dobima.



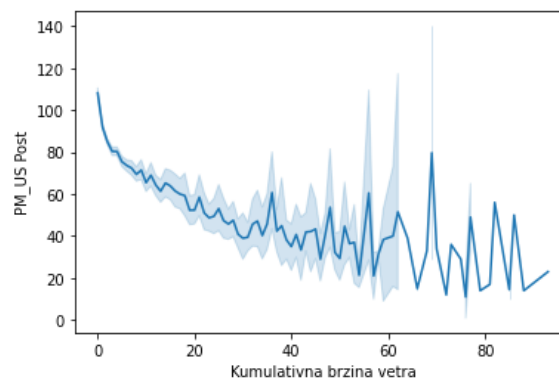
Slika 4. Uticaj temperature na koncentraciju PM čestica

Negativna korelacija između temperature i PM čestica je velika, pa tako sa porastom temperature vidimo znatno manju koncentraciju PM čestica u vazduhu. To smo takođe dokazali na slici 3 gde vidimo da najmanju vrednost PM čestica vidimo na leto a najveću na zimu.



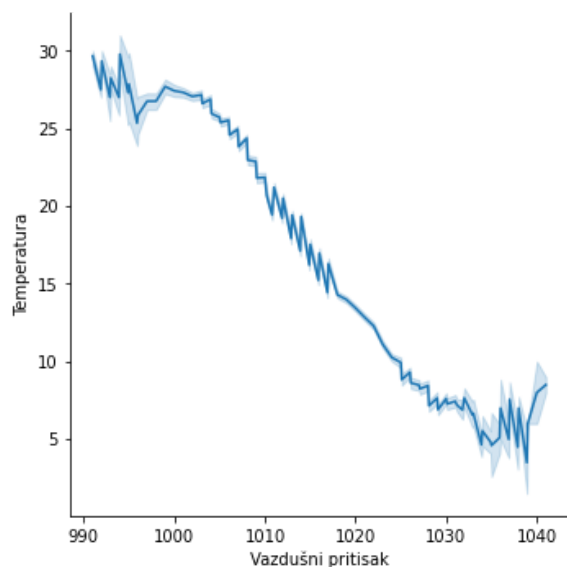
Slika 5. Uticaj vazdušnog pritiska na koncentraciju PM čestica

Uviđamo da sa povećanje pritiska dolazi do mnogo veće varijacije u merenjima koncentracije PM čestica u vazduhu, kao i do povećane koncentracije PM čestica u vazduhu.



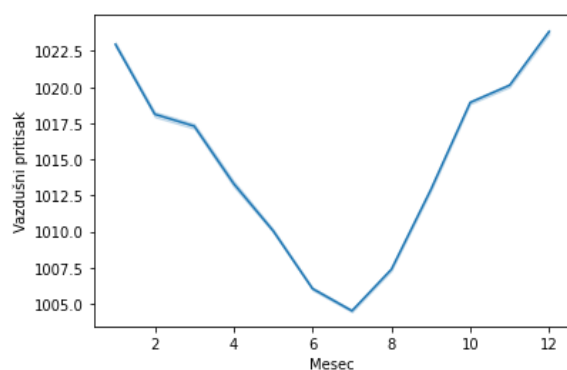
Slika 6. Uticaj kumulativne brzine vetra na koncentraciju čestica

Možemo videti blagu negativnu korelaciju između kumulativne brzine vetra i koncentracije PM čestica. Kao i to da sa porastom kumulativne brzine vetra dolazi do velike varijacije u merenjima koncentracije PM čestica.



Slika 7. Povezanost temperature i vazdušnog pritiska

Vidimo postojanje negativne korelacije između temperature i vazdušnog pritiska što se slaže sa ranijom analizom gde smo videli postojanje negativne korelacije između temperature i koncentracije PM čestica i pozitivne korelacije između vazdušnog pritiska i koncentracije PM čestica.



Slika 8. Prosečna izmerena vrednost vazdušnog pritiska tokom meseci

Kao što smo mogli da vidimo na slici 7. kako topao vazduh utiče na pritisak tako i tokom najtoplijih meseci vidimo da je vazdušni pritisak najmanji.

III Odabir modela

Odabrao sam model *Linear Regression* iz python-ove biblioteke *sklearn* zato što je dao dobre rezultate za mali broj koeficijenata kao i verodostojnije koeficijente za razliku od *Lasso* modela koji je davao koeficijente koji su bili veliki gde je postojala mala verovatnoća da model bude preučen. U tabeli ispod možemo videti rezultate koje je postigao naš model

Test	Rezultat
Mean squared error	2340,849
Mean absolute error	35,397
Root mean squared error	48,382
R2 score	0,281
R2 adjusted score	0,280

Tabela 1. Rezultati modela

IV Zaključak

Svako povećanje PM čestica od 10 ug/m3 dovodi do povećanja stope smrtnosti 3- 26 % takođe povećava šansu oboljenja od bolesti kao što su astma (oko 16%), kancera pluća (oko 36%), srčanih bolesti (oko 44%) kao i mnogih drugih bolesti. U našoj analizi videli smo da postoji trend pada tokom godina koje su dolazile što je definitivno dobar pokazatelj da je vazduh čistiji i zdraviji za udisanje. Takođe smo primetili da sa povećanjem temperatura dolazi do naglog pada koncentracije PM čestica u vazduhu što nas dovodi do zaključka da je vazduh najzdraviji leti. Pomoću korelacije između vazdušnog pritiska i koncentracije PM čestica u vazduhu smo došli do zaključka da se PM čestice grupišu na nižim nadmorskim visinama, pošto vazdušni pritisak pada kako je nadmorska visina veća.

V Reference

- <https://www.airveda.com/blog/what-is-pm2-5-and-why-is-it-important>

