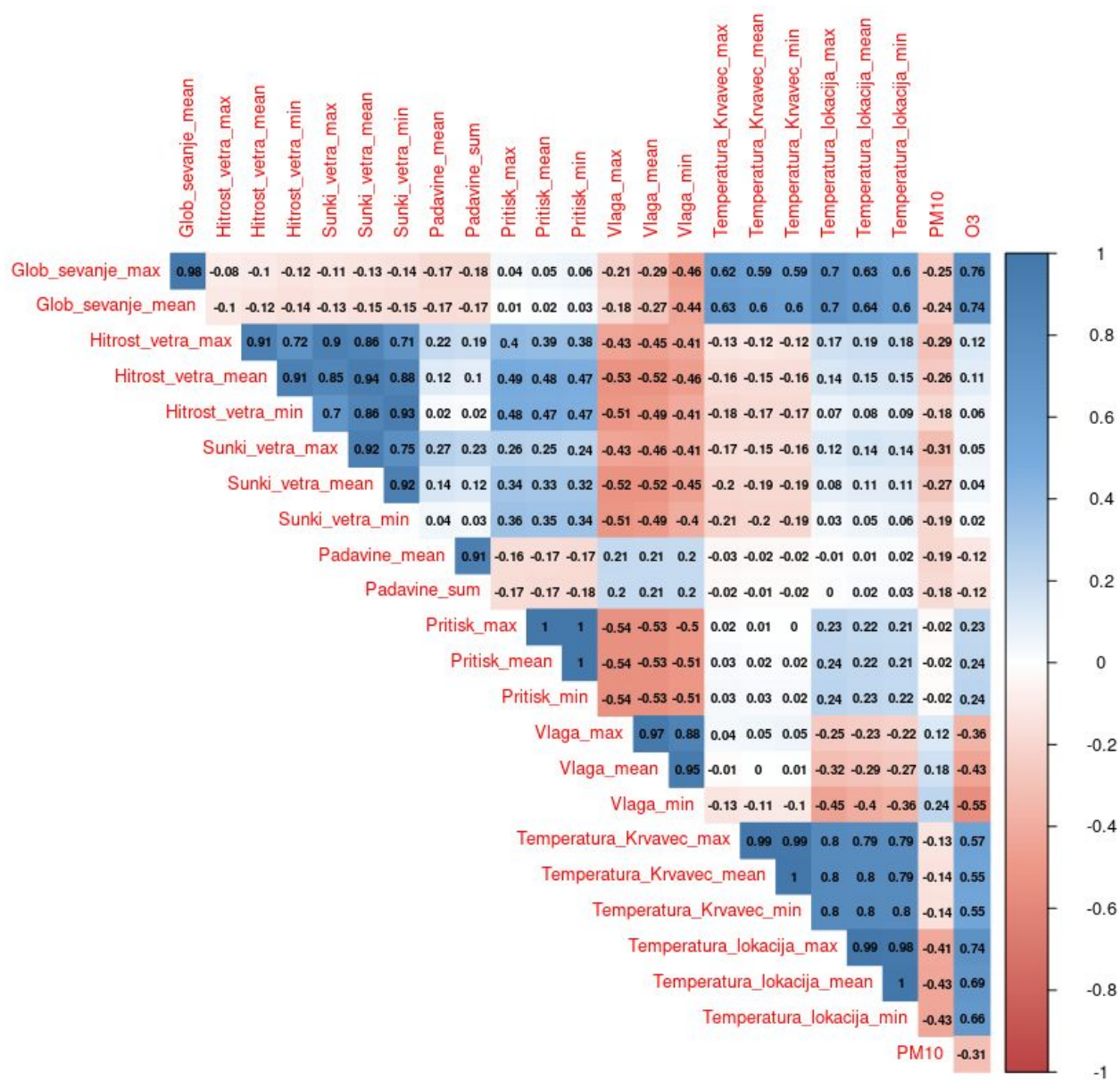


Inteligentni sistemi

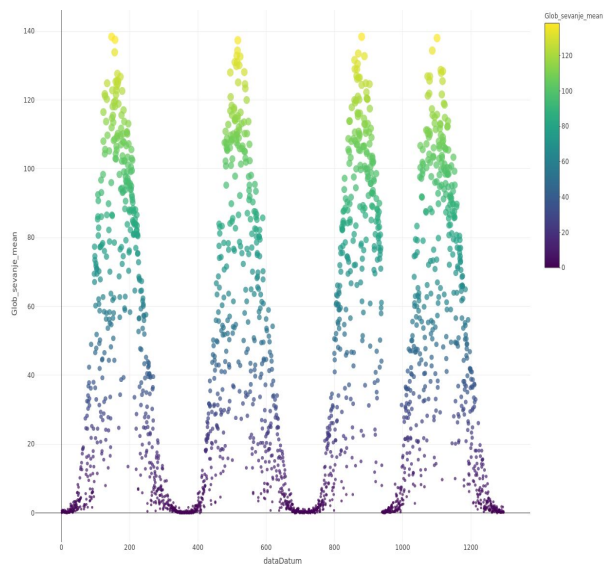
Prva seminarska naloga

1. Vizualizacija podatkov

V tem poglavju sem narisal nekaj grafov, ki prikažejo korelacijo in odvisnosti atributov. Iz tega sem lahko ugotovil kateri atributi so pomembni pri napovedovanju.



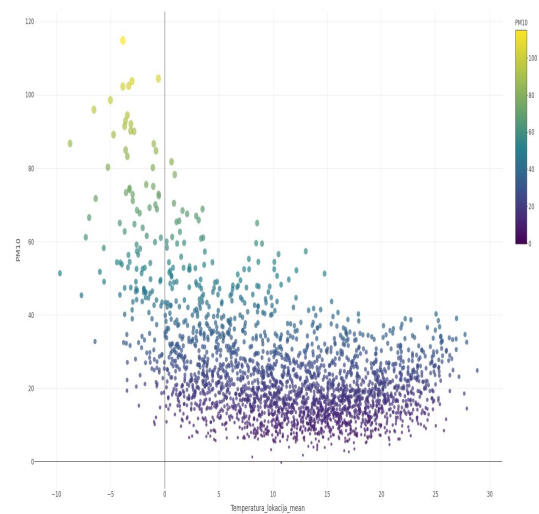
Graf 1 (Korelacija atributov)



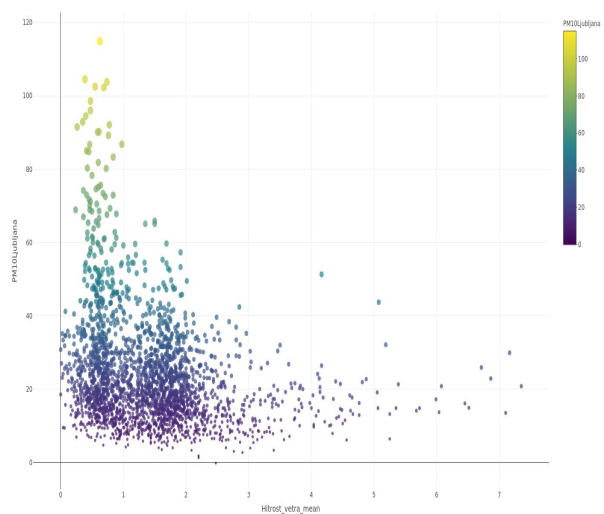
Graf 2 (Sprememba globalnega sevanja v odvisnosti od letnega časa)



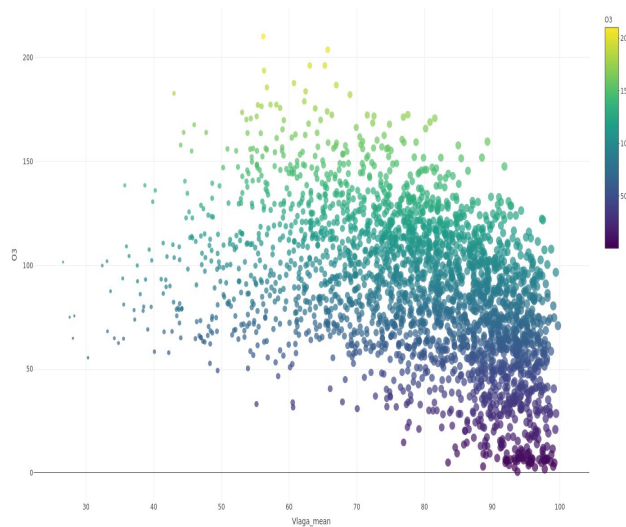
Graf 3 (Sprememba O3 v odvisnosti od temperature)



Graf 4 (Sprememba PM10 v odvisnosti od temperature)



Graf 5 (Sprememba PM10 v odvisnosti od hitrosti vetra)



Graf 6 (Sprememba O3 v odvisnosti od vlage)

2. Ocenjevanje atributov

V tem poglavju sem ocenil attribute glede na vizualizacije iz prejšnjega poglavja in s pomočjo algoritmov za ocenjevanje atributov.

Atributi z močno korelacijo:

- Temperatura, sevanje, vlaga in O3
- Hitrost vetra, sunki vetra, temperatura in PM10

Atribut Globalno_sevanje_min je bil neuporaben, ker je vseboval same ničle.

Atribut Datum je bil neuporaben in zato sem iz njega ustvaril dva nova atributa:

- letni čas in
- dan v tednu

3. Klasifikacija

		po dodajanju novih atributov	
Algorithm	O3	O3 ver2	PM10
Majority class	0.6089588	0.6089588	0.8688458
DT	CA : 0.7177419	0.7217742	0.8870968
	Brier: 0.4381122	0.4106455	0.1728015
DT with cross val.	1-err: 0.7619048	0.7601	0.876
NB	CA: 0.5994624	0.6169355	0.7768817
	Brier: 0.6427464	0.6231271	0.3946745
NB with cross val.	1-err: 0.5960452	0.6067	0.782
KNN k=5	CA: 0.7473118	0.7392473	0.8978495
	Brier: 0.3674194	0.3810753	0.1777419
	1-err:	0.7941	0.8841
KNN k = 10	CA: 0.7701613	0.7715054	0.9153226
	Brier: 0.3329301	0.3352957	0.1494892
		0.7849	0.887
KNN k = 15	CA: 0.780914	0.7876344	0.9112903
	Brier: 0.3265352	0.3260573	0.1425209
KNN cross val.	1-err: 0.7841001	0.7924	0.887
RF	CA: 0.7903226	0.7889785	0.9233871
	Brier: 0.3127936	0.3166823	0.1311955
Rf with cross val.	1-err: 0.81477	0.8206	0.9008
SVM	CA: 0.7876344	0.7903226	0.9327957
	Brier: 0.4996265	0.4743393	0.1239688
	1-err: 0.7964	0.8005	0.8968
ANN	1-err: 0.801	0.7733564	
Voting SVM, RF, KNN15	CA: 0.796	0.8024194	0.9274194
Weighted voting SVM, RF, KNN15	CA: 0.7405914	0.7419355	0.8763441
Boosting	CA: 0.781	0.7688172	0.9005376

S tabele lahko razberemo rezultate posameznega algoritma in zmagovalca (označene z rdečo barvo pisave).

4. Regresija

	O3				PM10			
Algorithm	MSE	RMSE	MAE	RMAE	MSE	RMSE	MAE	RMAE
LM	345.0908	0.2686063	14.4805	0.4958697	124.8822	0.6174796	8.580496	0.8531391
Regression tree with modelTypeReg = 6 (kNN)	347.4947	0.2704775	14.5212	0.4972636	91.72279	0.453523	7.507755	0.746479
Regression tree RReliefF and with modelTypeReg = 6 (kNN)	343.1279	0.2670785	14.37433	0.492234	74.16908	0.3667288	6.794317	0.6755434
RF and MSEofMean	301.5771	0.2347369	13.65303	0.4675339	73.62446	0.3640359	6.786866	0.6748026
KNN5	432.513	0.3366527	16.40366	0.5617262	113.3946	0.5606794	7.899606	0.7854397
KNN10	378.1599	0.2943462	15.27485	0.5230714	97.29482	0.481074	7.52039	0.7477352
KNN15	354.4875	0.2759204	14.79146	0.5065182	93.14344	0.4605475	7.39454	0.7352223
SVM	276.9921	0.2156008	13.17925	0.4513098	70.74282	0.3497876	6.242065	0.6206343

5. Evalvacija modelov

Pri klasifikaciji in regresiji se je najboljši pokazal algoritem SVM, razen pri klasifikaciji O3, kjer ga je voting med SVM, RF in KNN15 premagal za 1%. Najboljši algoritem sem izbral s pomočjo klasifikacijske točnosti, Brijeve metode in 10-kratnega prečnega preverjanja za klasifikacijo ter MAE, RMAE, MSE, RMSE za regresijo.

6. Bonus

Pri dodatni nalogi sem ustvaril novo podatkovno množico sestavljeno iz meritev iz več dneh (temperatura za tekoči dan, temperatura za en dan nazaj itn). Pri testiranju sem uporabil algoritem SVM. Novi modeli so imeli nižje ocene za razliko od starih modelov, ki so se učili na originalni podatkovni množici.

	klasifikacija		regresija O3			
	O3	PM10	MAE	RMAE	MSE	RMSE
2 dni	0.7725437	CA: 0.9286676	343.1548	0.2218056	14.63781	0.4641346
3 dni	0.7264151	CA: 0.9231806				
5 dni	0.7256757	CA: 0.9094595				

Miloš Kostadinovski, 63160172, VSŠ