

Univerzitet u Kragujevcu

Master 4.0



Predmet: Sistemi za podršku odlučivanju

Domaci zadatak 2

Student: Kristina Satarić

20/2019

Profesor: red. prof. dr Nenad Filipović

Asistent: Tijana Šušteršič

Kragujevac, april 2020.

Baza koja je korišćena za izradu zadatka je prikazana na *Slici 1*.

Pacijent broj	Dijagnoza	sensor 1	sensor 2	sensor 3	sensor 4
Pacijent 1	OA	6,3819	10,0855	15,2717	10,5492
Pacijent 2	OA	12,5242	14,1316	12,0653	17,245
Pacijent 3	HY	10,5453	9,8261	13,8631	12,5302
Pacijent 4	MA	3,1812	13,9793	6,6729	18,9704
Pacijent 5	HY	15,5398	13,4774	15,299	19,2242
Pacijent 6	HY	14,9515	13,8512	16,3812	20,09
Pacijent 7	HY	15,505	13,8639	18,4492	17,0094
Pacijent 8	HY	15,24	11,0525	16,917	13,4258
Pacijent 9	HY	16,6769	10,4063	16,7752	18,0579
Pacijent 10	OA	15,3732	16,6454	15,0656	16,3916
Pacijent 11	MA	1,8109	13,693	16,1077	19,5674
Pacijent 12	CA	17,7293	16,9379	18,7762	12,5159
Pacijent 13	OA	14,9301	16,5402	12,101	17,6034
Pacijent 14	OA	6,7443	8,5979	0,1133	4,8688
Pacijent 15	CA	9,3205	8,1404	15,1233	0,683
Pacijent 16	MA	1,8109	10,693	16,1077	19,5674
Pacijent 17	CA	9,799	7,8564	1,0609	1,0278
Pacijent 18	CA	6,3205	6,1404	15,1233	0,61
Pacijent 19	HY	11,0736	8,5256	0,0877	0,04
Pacijent 20	HY	14,0344	3,9035	11,5287	11,2915
Pacijent 21	HY	15,1183	11,4564	15,5569	16,3629
Pacijent 22	CA	9,889	7,5564	1,0609	1,0278
Pacijent 23	MA	6,2567	10,7803	15,5786	14,8403
Pacijent 24	OA	12,5919	13,5176	6,2939	16,161
Pacijent 25	MA	3,2443	9,3636	6,7614	5,9694
Pacijent 26	HY	14,0344	3,9035	11,5287	11,2915
Pacijent 27	HY	15,0083	13,2563	11,3368	16,1529
Pacijent 28	HY	13,4449	5,5620	16,0809	11,2763
Pacijent 29	MA	6,8013	12,6644	8,6707	16,8033
Pacijent 30	MA	6,1797	9,7803	15,4686	14,8403
Pacijent 31	OA	12,2408	13,3810	5,1431	18,3535
Pacijent 32	OA	12,1771	17,0372	6,9298	16,4749
Pacijent 33	OA	11,6434	11,5554	9,7108	15,7851
Zdrav 1	-	16,2013	16,3692	16,9985	18,2890
Zdrav 2	-	14,1295	14,1990	14,8216	14,5003
Zdrav 3	-	13,5479	13,3415	16,2864	16,2359
Zdrav 4	-	15,9077	15,4755	17,3645	18,3335
Zdrav 5	-	16,7221	16,2755	16,1302	17,5298
Zdrav 6	-	15,2184	15,2990	14,9216	14,6003
Zdrav 7	-	14,5479	14,3415	16,2864	16,2359
Zdrav 8	-	16,3822	16,7341	18,6174	17,2454
Zdrav 9	-	16,2013	16,4444	17,0985	18,3158

Slika 1. Baza podataka

Tabela nam služi kao input za primenu algoritma. Cilj je da analiziramo očitavanja sa senzora kako bismo postavili tačnu dijagnozu.

Dijagnoze su:

- OA – obstructive sleep apnea
- CA – central sleep apnea
- MA – mixed sleep apnea
- HY – hypopnea

Zdravi su označeni sa -.

Očitane vrednosti sa četiri senzora ne prelaze vrednost 21 (max).

Za bolesti:

- OA – trebalo bi da su vrednosti sa senzora 3 manje
- CA – trebalo bi da su vrednosti sa senzora 4 manje
- MA – trebalo bi da su vrednosti sa senzora 1 manje
- HY – trebalo bi da su vrednosti sa senzora 2 manje

Zdravi su ljudi koji imaju očitavanja vrednosti sa senzora 1 i 2 skoro ista i da vrednosti sa senzora 3 i 4 takođe budu slična.

Algoritmi koji su implementirani su:

- Logistička regresija (Logistic Regression)
- Stablo odlučivanja (Decision Tree Learning)
- Algoritam slučajnih šuma (Random Forest)
- Klasifikator Naivni Bajes (Naive Bayes Classifier)
- Algoritam potpornih vektora (Support Vector Machine)
- Algoritam K najbližih suseda (K Nearest Neighbour (KNN))

Algoritmi su primenjeni na originalnim podacima ali i na različitim metodama za normalizaciju. Metode koje su korišćene su:

- MinMaxScaler
- MaxAbsScaler
- Normalizer
- StandardScaler

Za trening i test, izvršićemo podelu 41:1 – uzimamo sve kombinacije 1 test i preostalo trening. Takođe vršimo izračunavanje prosečne tačnosti. Na *Slikama* od 2 do 7 prikazani su rezultati tačnosti algoritama sa različitim normalizatorima.

BezNormalizacije

```
LR: 0.8095238095238095
DT: 0.6904761904761905
RF: 0.7142857142857143
SVM: 0.5476190476190477
NB: 0.7619047619047619
KNN: 0.6190476190476191
```

Slika 2. Prva podela, bez normalizacije

MinMax

```
LR: 0.6428571428571429
DT: 0.6904761904761905
RF: 0.7142857142857143
SVM: 0.35714285714285715
NB: 0.5238095238095238
KNN: 0.6428571428571429
```

Slika 3. Prva podela, MinMax

MaxAbs

LR: 0.5396825396825397
DT: 0.6904761904761905
RF: 0.7142857142857143
SVM: 0.3253968253968254
NB: 0.4444444444444444
KNN: 0.6428571428571429

Slika 4. Prva podela, MaxAbs

Normalizacija

LR: 0.47619047619047616
DT: 0.6904761904761905
RF: 0.7142857142857143
SVM: 0.31547619047619047
NB: 0.40476190476190477
KNN: 0.6428571428571429

Slika 5. Prva podela, normalizacija

Standardizacija

LR: 0.5380952380952381
DT: 0.6952380952380952
RF: 0.7142857142857143
SVM: 0.38095238095238093
NB: /
KNN: 0.6476190476190476

Slika 6. Prva podela, standardizacija

Algoritmi Tehnike	LR	DT	RF	NB	SVM	KNN
BezNormalizacije	0.810	0.690	0.714	0.762	0.548	0.619
MinMax	0.643	0.690	0.714	0.524	0.357	0.643
MaxAbs	0.540	0.690	0.714	0.444	0.325	0.643
Normalizacija	0.476	0.690	0.714	0.405	0.315	0.643
Standardizacija	0.538	0.695	0.714	/	0.381	0.648

Slika 7. Prva podela, Tačnosti svih algoritama

Sa Slike 7. možemo zaključiti da najveću tačnost daje algoritam Logističke regresije na originalnim podacima.

Za drugu podelu smo koristili dva za test, sve ostalo za trening. Rezultati tačnosti algoritama prikazani su na *Slikama* od 8 do 13.

BezNormalizacije

LR: 0.8089430894308943
DT: 0.662020905923345
RF: 0.6887340301974448
SVM: 0.532520325203252
NB: 0.7415795586527294
KNN: 0.6155632984901278

Slika 8. Druga podela, bez normalizacije

MinMax

LR: 0.6408246225319396
DT: 0.662020905923345
RF: 0.6887340301974448
SVM: 0.3588850174216028
NB: 0.5136469221835076
KNN: 0.6385017421602788

Slika 9. Druga podela, MinMax

MaxAbs

LR: 0.5480061943476577
DT: 0.6658923732094464
RF: 0.6887340301974448
SVM: 0.3221060782036392
NB: 0.43747580332946184
KNN: 0.6385985288424313

Slika 10. Druga podela, MaxAbs

Normalizacija

LR: 0.48243321718931476
DT: 0.6675377468060395
RF: 0.6887340301974448
SVM: 0.3130081300813008
NB: 0.3995354239256678
KNN: 0.6389372822299652

Slika 11. Druga podela, normalizacija

Standardizacija

LR: 0.5407665505226481
DT: 0.6694541231126597
RF: 0.6887340301974448
SVM: 0.378397212543554
NB: /
KNN: 0.6434378629500581

Slika 12. Druga podela, standardizacija

Algoritmi Tehnike	LR	DT	RF	NB	SVM	KNN
BezNormalizacije	0.809	0.662	0.689	0.742	0.533	0.616
MinMax	0.641	0.662	0.689	0.514	0.359	0.639
MaxAbs	0.548	0.666	0.689	0.437	0.322	0.639
Normalizacija	0.482	0.668	0.689	0.400	0.313	0.639
Standardizacija	0.541	0.669	0.689	/	0.378	0.643

Slika 13. Druga podela, Tačnosti svih algoritama

Sa *Slike 13.* možemo zaključiti da najveću tačnost daje algoritam Logističke regresije na originalnim podacima.