

Umetna inteligenca 2017-2018

Seminarska naloga 1

Naloga se izvaja **v parih**. Zagovori bodo potekali v terminu vaj v tednu **11. 12. – 15. 12. 2017**.

Na učilnici je podana datoteka "podatkiSem1.txt", ki vsebuje podatke o vremenu in onesnaženju zraka v obdobju od 2013 do 2016. Podatkovna množica vsebuje 2478 meritev.

Neodvisne spremenljivke (atributi):

| | |
|---------------------------|--|
| Datum | Čas meritve v formatu YYYY-MM-DD |
| Postaja | Diskretni atribut, identifikacijska oznaka merilnega mesta |
| Glob_sevanje_max | Zv. atribut, najvišja vrednost globalnega sevanja med 0:00 in 7:00 |
| Glob_sevanje_mean | Zv. atribut, povprečna vrednost globalnega sevanja med 0:00 in 7:00 |
| Glob_sevanje_min | Zv. atribut, najnižja vrednost globalnega sevanja med 0:00 in 7:00 |
| Hitrost_vetra_max | Zv. atribut, najvišja hitrost vetra med 0:00 in 7:00 |
| Hitrost_vetra_mean | Zv. atribut, povprečna hitrost vetra med 0:00 in 7:00 |
| Hitrost_vetra_min | Zv. atribut, najnižja hitrost vetra med 0:00 in 7:00 |
| Sunki_vetra_max | Zv. atribut, najvišja hitrost sunkov vetra med 0:00 in 7:00 |
| Sunki_vetra_mean | Zv. atribut, povprečna hitrost sunkov vetra med 0:00 in 7:00 |
| Sunki_vetra_min | Zv. atribut, najnižja hitrost sunkov vetra med 0:00 in 7:00 |
| Padavine_mean | Zv. atribut, povprečna količina padavin (na uro) med 0:00 in 7:00 |
| Padavine_sum | Zv. atribut, skupna količina padavin med 0:00 in 7:00 |
| Pritisk_max | Zv. atribut, najvišja vrednost zračnega pritiska med 0:00 in 7:00 |
| Pritisk_mean | Zv. atribut, povprečna vrednost zračnega pritiska med 0:00 in 7:00 |
| Pritisk_min | Zv. atribut, najnižja vrednost zračnega pritiska med 0:00 in 7:00 |
| Vlaga_max | Zv. atribut, najvišja vrednost vlažnosti zraka med 0:00 in 7:00 |
| Vlaga_mean | Zv. atribut, povprečna vrednost vlažnosti zraka med 0:00 in 7:00 |
| Vlaga_min | Zv. atribut, najnižja vrednost vlažnosti zraka med 0:00 in 7:00 |
| Temperatura_Krvavec_max | Zv. atribut, najvišja temperatura zraka na Krvavcu med 0:00 in 7:00 |
| Temperatura_Krvavec_mean | Zv. atribut, povprečna temperatura zraka na Krvavcu med 0:00 in 7:00 |
| Temperatura_Krvavec_min | Zv. atribut, najnižja temperatura zraka na Krvavcu med 0:00 in 7:00 |
| Temperatura_lokacija_max | Zv. atribut, najvišja temp. zraka na merilnem mestu med 0:00 in 7:00 |
| Temperatura_lokacija_mean | Zv. atribut, povpr. temp. zraka na merilnem mestu med 0:00 in 7:00 |
| Temperatura_lokacija_min | Zv. atribut, najnižja temp. zraka na merilnem mestu med 0:00 in 7:00 |

Odvisni (ciljni) spremenljivki:

| | |
|------|---|
| O3 | Zvezni atribut, maksimalna dnevna koncentracija ozona |
| PM10 | Zvezni atribut, dnevna koncentracija prašnih delcev premera 10 µm |

Cilj seminarske naloge je uporabiti metode strojnega učenja za gradnjo modelov za napovedovanje onesnaženja zraka, ustrezno ovrednotiti modele in jasno predstaviti dobljene rezultate.

Konkretne naloge, ki jih je potrebno opraviti:

1. Vizualizacija podatkov

Prpravite nekaj zanimivih grafov, ki ilustrirajo podane podatke (distribucije vrednosti, soodvisnosti med atributi, ponavljajoče se vzorce in podobno).

2. Ocenjevanje atributov

Ocenite kvaliteto podanih atributov in konstruirajte nove attribute, ki lahko izboljšajo kvaliteto zgrajenih modelov. Namig: datum je v obstoječi obliki relativno neuporaben, iz njega pa lahko izpeljemo nove attribute (npr. letni čas, dan v tednu...), ki potencialno pomagajo pri napovedovanju onesnaženja zraka.

3. Klasifikacija

Zgradite vsaj tri klasifikacijske modele za napovedovanje:

- a. maksimalne dnevne koncentracije ozona – možni razredi so: NIZKA (pod 60.0), SREDNJA (med 60.0 in 120.0), VISOKA (med 120.0 in 180.0) in EKSTREMNA (nad 180.0),
- b. dnevne koncentracije prašnih delcev PM10 – možni razredi so: NIZKA (do 35.0), VISOKA (nad 35.0).

4. Regresija

Zgradite vsaj tri regresijske modele za napovedovanje:

- a. maksimalne dnevne koncentracije ozona,
- b. dnevne koncentracije prašnih delcev PM10.

5. Evalvacija modelov

Zgrajene modele ustrezno ovrednotite in predstavite dobljene rezultate. Primerjajte različne modele na podlagi dosežene točnosti napovedovanja in razumljivosti delovanja.

6. Poročilo (dokument v formatu .doc ali .pdf)

V poročilu opišite vaš pristop, uporabljene modele in attribute, predstavite dosežene rezultate ter strnite zaključke na podlagi eksperimentalne evalvacije.

Za višjo oceno: primerjajte uspešnost modelov, ki se učijo samo iz meritev tekočega dne (originalna podatkovna množica), z modeli, ki se učijo iz meritev nekaj zadnjih dni (nova podatkovna množica, pri kateri so posamezne vrstice dobljene z združevanjem več zaporednih vrstic iz originalne podatkovne množice). Koliko zadnjih dni je potrebno upoštevati, da dobimo najbolj točne napovedi?

Ocenjevanje

Na končno oceno seminarske raziskovalne naloge vplivajo kvaliteta zgrajenih modelov, inovativnost in elegantnost rešitve, ambicioznost pri raziskovanju problema, argumentacija izbranih postopkov, vizualizacija in razlaga dobljenih rezultatov.