

Umetna inteligenca 2021-2022

Seminarska naloga 1

Naloga se izvaja v **parih**. Zagovori bodo potekali v terminu vaj v tednu **6. 12. – 10. 12. 2021.**

Na učilnici sta podani datoteki "ucnaSem1.txt" in "testnaSem1.txt", ki vsebujeta **učne** in **testne** podatke o porabi električne energije v odvisnosti od lastnosti in namembnosti stavbe ter trenutne vremenske slike.

Opis atributov:

datum	datum meritve v formatu YYYY-MM-DD
regija	oznaka regije, v kateri je stavba (možni vrednosti: "vzhodna", "zahodna")
stavba	oznaka stavbe
namembnost	namembnost stavbe (možne vrednosti: "izobrazevalna", "javno_storitvena", "kulturno_razvedrilna", "poslovna" in "stanovanjska")
povrsina	uporabna površina stavbe v m ²
leto_izgradnje	leto izgradnje stavbe v YYYY
temp_zraka	temperatura zraka v C°
temp_rosisca	temperatura rosišča v C°
oblacnost	stopnja pokritosti neba z oblaki (0 – popolnoma jasno, 10 – popolnoma pokrito)
padavine	količina padavin v zadnji uri v mm (vrednost -1 označuje pršenje)
pritisk	zračni pritisk v mbar
smer_vetra	smer pihanja vetra v stopinjah (npr. 90 – z vzhoda, 180 – z juga, 270 – od zahoda, 360 – s severa, 0 - brezvetrje)
hitrost_vetra	hitrost vetra v m/s
poraba	poraba elektrike v kWh

Cilj seminarske naloge je uporabiti metode strojnega učenja za gradnjo modelov za napovedovanje porabe električne energije (**regresijski problem**) in namembnosti stavbe (**klasifikacijski problem**), ustrezno ovrednotiti modele in jasno predstaviti dobljene rezultate.

Konkretne naloge, ki jih je potrebno opraviti:

1. Vizualizacija podatkov
Pripravite nekaj zanimivih grafov, ki ilustrirajo podane podatke (porazdelitve vrednosti, soodvisnosti med atributi, ponavljajoče se vzorce in podobno).
2. Ocenjevanje in konstrukcija 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, vikend...), ki potencialno pomagajo pri napovedovanju porabe električne energije. Atributni prostor lahko razširite tudi s statistikami, ki povzemajo **prejšnje** meritve. Na primer, vsakemu učnemu primeru lahko dodate podatek o porabi elektrike v isti stavbi iz prejšnjega oziroma nekaj zadnjih dni. Poleg

dejanskih meritev lahko dodate povprečno (oz. maksimalno, minimalno, skupno) porabo iz prejšnjega obdobja (npr. tedna ali meseca). Statistike se lahko nanašajo tudi na vremensko sliko (temperatura zraka, padavine, hitrost vetra...). **Pri dodajanju novih atributov je dovoljena uporaba samo tistih podatkov, ki se nanašajo na opazovano stavbo in so v trenutku zajema učnega primera bili na voljo (drugače povedano, so izračunani iz učnih primerov s starejšimi datumi).**

3. Modeliranje

Uporabite vsaj tri učne algoritme za napovedovanje:

- a. porabe električne energije (regresijski problem) – stolpec "poraba" predstavlja ciljno spremenljivko,
- b. namembnosti stavbe (klasifikacijski problem) - stolpec "namembnost" predstavlja ciljno spremenljivko.

4. Evalvacija modelov

Za učenje modelov uporabite samo podatke iz datoteke "ucnaSem1.txt". Za ocenjevanje kvalitete napovedovanja uporabite samo podatke iz datoteke "testnaSem1.txt".

5. 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.

Ocenjevanje

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

Za oceno 6: vizualizacija; 3 klasifikacijski modeli; 3 regresijski modeli; evalvacija kvalitete dobljenih modelov.

+ 1 ocena: kreiranje vsaj treh dodatnih atributov; pri gradnji modelov poskušajte izbrati ustrezno podmnožico atributov in primerjajte tako dobljene modele s tistimi, ki so naučeni na celotni učni množici.

+1 ocena: preizkusite različne načine kombiniranja modelov strojnega učenja (npr. glasovanje, uteženo glasovanje, bagging in podobno) in jih primerjajte z osnovnimi modeli.

+1 ocena: primerjajte uspešnost modelov, ki se učijo samo iz podatkov ene regije, z modeli, ki se učijo iz podatkov obeh regij.

+1 ocena: pri evalvaciji modelov upoštevajte kronologijo podatkov. Najprej združite vse podatke iz učne in testne množice. Celotno zbirko podatkov razdelite na 12 podmnožic, glede na mesec zajema. Najprej uporabite januarske podatke za učenje modelov, ki jih testirate na februarških podatkih. Nato uporabite januarske in februarske podatke kot učne, naučene modele testirate na marčevskih podatkih. Tako nadaljujete po mesecih do decembra, ki ga testirate na modelih, naučenih na vseh podatkih od januarja do novembra.