

Umetna inteligenca 2020-2021

Seminarska naloga 1

Naloga se izvaja v **parih**. Zagovori bodo potekali v terminu vaj v tednu **7. 12. – 11. 12. 2020**.

Na učilnici je podana datoteka "dataSem1.txt", ki vsebuje podatke o porabi električne energije v odvisnosti od lastnosti in namembnosti stavbe ter trenutne vremenske slike. Podatki obsegajo celotno leto 2016 in so zajeti štirikrat na dan (ob 5:00, 11:00, 17:00 in 23:00).

Neodvisne spremenljivke (atributi):

datum	datum meritve v formatu YYYY-MM-DD
ura	ura meritve (možne vrednosti: 5, 11, 17, 23)
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

Odvisni (ciljni) spremenljivki:

poraba	poraba elektrike v kWh (regresijski problem)
norm_poraba	poraba elektrike na enoto površine (klasifikacijski problem , možni razredi: "ZELO NIZKA", "NIZKA", "SREDNJA", "VISOKA" in "ZELO VISOKA")

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

Konkretna naloga, 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 ob istem času iz prejšnjega dne. Lahko dodate povprečno (oz. maksimalno, minimalno, skupno) porabo iz prejšnjega dne. Učne primere lahko razširite s statistikami za dva, tri ali teden dni nazaj. 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.** Na primer, učnemu primeru, ki se nanaša na porabo elektrike ob 5 uri zjutraj, ne smete dodati podatka o povprečni porabi elektrike ta dan, saj le-ta v tem trenutku ni bila znana.

3. Modeliranje

Uporabite vsaj tri učne algoritme za napovedovanje:

- a. porabe električne energije (regresijski problem) – stolpec "poraba" predstavlja ciljno spremenljivko (pri tej nalogi **ni dovoljena** uporaba stolpca "norm_poraba"),
- b. normalizirane porabe električne energije (klasifikacijski problem) - stolpec "norm_poraba" predstavlja ciljno spremenljivko (pri tej nalogi **ni dovoljena** uporaba stolpca "poraba").

4. Evalvacija modelov

Zgrajene modele ovrednotite na naslednji način. Vse podatke razdelite na 12 podmnožic, glede na mesec zajema. Najprej uporabite januarske podatke za učenje modelov, ki jih testirate na februarjskih podatkih. Nato uporabite januarske in februarjske 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.

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.

Za oceno 7 vse od prej plus: 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.

Za oceno 8 vse od prej plus: preizkusite različne načine kombiniranja modelov strojnega učenja (npr. glasovanje, uteženo glasovanje, bagging in podobno) in jih primerjajte z osnovnimi modeli.

Za ocene 9 in 10 vse od prej plus: primerjajte uspešnost modelov, ki se učijo samo iz podatkov ene regije (stavbe), z modeli, ki se učijo iz podatkov obeh regij (vseh stavb).