

# NSU DN2 - Učenje iz podatkovnih tokov

## Uvod

Domači nalogi sta priloženi 2 datoteki formata CSV: **dn2a.csv**, ki pripada 1. nalogi, ter **dn2b.csv**, ki pripada 2. nalogi. Pri obeh nalogah najprej pripravite podatkovni nabor v ARFF formatu, tako da dodate in/ali odstranite ustrezne značilke. ARFF datoteke nalog, ki jih boste reševali, pripnite končnemu poročilu. Modele učite v okolju MOA.

Poročilo naj bo v PDF formatu, vsebuje pa opis postopka priprave podatkov, ter odgovore na vprašanja zastavljena pri posamezni nalogi. **Poročilo naj ne presega 400 besed na rešeno nalogo.**

## 1. naloga - regresija

V datoteki dn2a.csv so podane meritve treh zveznih spremenljivk, **x**, **y** in **c**; napovedovati želimo vrednost spremenljivke **c**.

**1.0. (4 točke)** Pripravite ustrezni podatkovni nabor.

**1.1. (3 točke)** Z metodo FIMT-DD na podatkovnem naboru naučite drevo. Izrišite njegovo povprečno absolutno napako, ter končni model.

**1.2. (3 točke)** Izrišite še graf ujemanja pravih ter napovedanih vrednosti spremenljivke **c**.

## 2. naloga - klasifikacija

V datoteki dn2b.csv so podane meritve petih zveznih spremenljivk **f1**, **f2**, **f3**, **f4** in **f5**, ter ene binarne spremenljivke **y**. Napovedovati želimo vrednost spremenljivke **y**.

**2.0. (2 točki)** Pripravite ustrezni podatkovni nabor.

**2.1. (5 točk)** V naboru se pojavljajo spremembe. Ocenite, kdaj se pojavijo, ter opredelite njihov tip, tj., ali so spremembe navidezne/prave, ter ali so nenade/koračne/postopne/itd. Odgovore utemeljite.

**2.2 (3 točke)** Primerjajte modela naučena z metodo za učenje Hoeffdingovih dreves z in brez zaznavanja sprememb (HoeffdingAdaptiveTree in HoeffdingTree v MOA). Izrišite njuni napovedni točnosti, ter ocenite kolikšna je razlika v točnosti po spremembah.