

I-SUNS: Zadanie č.2

ANALÝZA DÁT A REGRESORY

Vo vybranom programovacom jazyku implementujte program, ktorý bude predpovedať hlasnosť (loudness) piesne. Trénovať budete rôzne regresné modely. K dispozícii v AIS sú csv súbory s testovacou a trénovacou množinou. Čas odovzdania je určený časom vloženia do AIS. Deadline pre získanie 10 bodov je **11.11.2021 o 13:00/15:00**. Každý týžden omeškania je penalizovaný stratou dvoch bodov.

Dáta

Dáta boli zozbierané pomocou [Spotify API](#), kde sa dá nájsť aj popis jednotlivých stĺpcov. Pridaný bol stĺpec query, kde je zadané, podľa akého kľúčového boli piesne (playlisty) vyhľadávané.

Úlohy

- Spoznajete, analyzujete a pripravte dáta na ďalšie spracovanie:
 - *EDA* - analyzujte stĺpce, hľadajte vzťahy, odpovedajte na vzniknuté otázky. Pomôžte si grafickými nástrojmi. Snažte sa analyzovať a opísať, ktoré dáta sú vhodné na ďalšie spracovanie. **2.5b**
 - Vzhľadom na analýzu vyhľadajte a odstráňte chybné dáta, hraničné hodnoty, duplikáty. **1b**
 - Spracujte textové dáta. Zakódujte stĺpce žánér a rok do číselných hodnôt. **1b**
 - Nezabudnite obe množiny správne normalizovať. **0.5b**
- Natrénujte SVM regresor na predpoveď hlasitosti piesne z dostupných informácií o skladbe.
 - Natrénujte SVM na regresnú úlohu predikcie hlasitosti skladby. Na vyhodnotenie použite metriku MSE a R^2 pre testovaciu aj trénovaciu množinu/ validačné množiny. Analyzujte aj reziduály v testovacej množine. **1.5b**
 - Počas tréovania používajte krížovú validáciu. **0.5b**
 - Nájdite najvhodnejšie parametre tréovania pomocou algoritmu mriežkového vyhľadávania (*GridSearch*). Zlepšenie ukážte na vyššie spomenutých metrikách. **1b**

- Pomocou súborového učenia natrénujte ďalšie regresory a opäť ich vyhodnoťte pomocou MSE a R^2 .
 - Použite súborové učenie a aplikujte metódu učenia bagging. **1b**
 - Použite súborové učenie a aplikujte metódu učenia boosting. **1b**
- Nezabudnite zhodnotiť, ktorý regresor bol váš najlepší.

Nepovinné úlohy

- Dáta obohaťte o ďalšie voľne dostupné metadáta. (Aj my sa chceme naučiť nejaký nový nástroj). **max. 2b**
- Ak používate forest klasifikátor, vizuálne ukážte jeden strom z lesa, ktorý vznikne po ensemble learning. **1b**
- Ak používate forest klasifikátor, zhodnoťte silu vstupných príznakov. **1b**
- Zopakujte si poznatky z minulého zadania. Natrénujte a vyhodnoťte ANN regresor (s výsledkami aspoň porovnateľnými s metódami z hlavnej časti zadania). **1b**