Adaptívne modely pre klasifikáciu nevybalansovaných dátových prúdov

IDDCW — Imbalance-aware Dynamic Diversified Class-Weighted ensemble

Autor: Bc. Matúš Konopinský
Vedúci práce: doc. Ing. Martin Sarnovský, PhD.
Technická univerzita v Košiciach · Fakulta elektrotechniky a informatiky
Ústav umelej inteligencie · 2026

1 Problém

Dáta zo senzorov, sociálnych sietí či sieťovej prevádzky prichádzajú nepretržite ako **dátové prúdy**, ktorých charakter sa v čase mení (**konceptový drift**). To najdôležitejšie – útok, podvod, toxický komentár – je pritom len **malý zlomok** dát. Bežné modely majú tendenciu takéto zriedkavé prípady prehliadať, hoci ich prehliadnutie býva najdrahšou chybou.

2 Príklady z praxe

- Sieťové útoky v záplave bežnej prevádzky
- Toxické komentáre medzi miliónmi príspevkov
- Podvodné transakcie medzi legitímnymi
- Dezinformácie počas rýchlo sa meniacich tém

3 Výskumná medzera

Drift aj nevyváženosť tried väčšina metód rieši **oddelene**. Keď nastanú súčasne, model musí sledovať meniace sa dáta a zároveň si „nezabudnúť všímať“ vzácne prípady – presne to model IDDCW rieši integrovane.

4 Riešenie: model IDDCW

Rozširuje existujúci model DDCW o mechanizmy cielené priamo na nevyváženosť tried, spracované vzorka po vzorke v reálnom čase.

Naïve Bayes · Hoeffding Tree · Hoeffding Adaptive Tree · EFDT · Perceptron

Heterogénny ensemble piatich rôznych klasifikátorov hlasuje formou **soft votingu** (váži istotu, nielen áno/nie).

▼

Adaptívne per-class váhovanie – odmena za správnu predikciu menšinovej triedy rastie s mierou nevyváženosti.

▼

Replay a augmentácia menšinových vzoriek z pamäte, regulované spätnou väzbou, aby model nestratil ani väčšinovú triedu.

▼

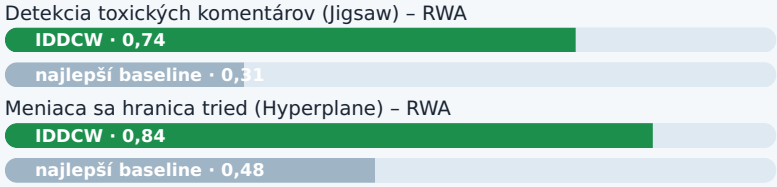
Page–Hinkley detektor driftu – pri zmene v dátach model dočasne zintenzívni učenie sa na nových príkladoch.

▼

Výsledok: spoľahlivejšie rozpoznanie zriedkavých, no kritických prípadov – bez straty presnosti na väčšinovej triede.

5 Experimentálne overenie

Model bol testovaný na **15 dátových prúdoch** – syntetických, reálnych tabuľkových (sieťová prevádzka, elektrina, letecké meškania) aj textových (toxické komentáre, spravodajské články) – a porovnaný s piatimi bežne používanými metódami (ARF, Leveraging Bagging, Online Bagging/Boosting, Hoeffding Tree).



RWA (Rarity-Weighted Accuracy) odmeňuje rozpoznanie zriedkavých tried, nie len celkovú presnosť. Model bol vyhodnocovaný priebežne, vzorka po vzorke – presne tak, ako by pracoval v ostrej prevádzke.

6 Výsledky a využitie

15 testovaných dátových prúdov	**5×** lepšie rozpoznanie menšiny pri silnej nevyváženosti
5 typov klasifikátorov v jednom ensemble	**2–3×** vyššia výpočtová náročnosť oproti bežným metódam

Najväčší prínos priniesol model pri výraznej nevyváženosti (nad 5:1) – v scenároch ako **detekcia škodlivého obsahu, útokov či podvodov**, kde je prehliadnutie vzácneho prípadu drahšie než falošné podozrenie.

Perspektíva do budúcna: nasadenie pri monitoringu siete, moderácii obsahu či odhaľovaní podvodov.