

Balíček
Strojové učenie

Karta

Zaujatost'

03



Umelá
inteligencia
vo vzdelávaní



MINISTERSTVO
ŠKOLSTVA, VÝSKUMU,
VÝVOJA A MLÁDEŽE
SLOVENSKEJ REPUBLIKY

Metodický materiál Kurikula umelej inteligencie pre základné a stredné školy
Informatika na 2. stupni ZŠ a SŠ – karty

Zaujatosť

Základné informácie o zaujatosti

Predstavte si, že chcete vytvoriť aplikáciu umelej inteligencie, ktorá rozpoznáva na obrázkoch topánky. Najprv jej musíte ukázať veľa obrázkov topánok, aby na nich aplikácia našla podobnosti (napríklad šnúrky, podrážky a pod.). A teraz si skúste predstaviť čo najviac typov topánok. Všetky tieto topánky by ste aplikácii ukázali. Ale teraz pozor! Predstavili ste si nielen všetky typy topánok, ale aj zo všetkých uhlov, vo všetkých farbách, materiáloch a so všetkými jednotlivými detailmi? Topánky, aké sa nosili v starovekom Egypte alebo v Číne v 5. storočí? AI nevníma veci tak ako my ľudia. Nemôže si topánky ohmatať, nedokáže zovšeobecňovať tak dobre ako my. Preto by aplikácii niektoré príklady, ktoré ste jej neukázali, nemusela rozpoznať ako topánky. A tomu sa hovorí zaujatosť.

Rozšírenie definície pre pokročilých

Systémy umelej inteligencie sa učia na dátach, ktoré pripravujú ľudia. Zle pripravené, nevyvážené dáta alebo ich nedostatok môžu spôsobiť, že umelá inteligencia bude určitým spôsobom zaujatá alebo tiež skreslená (po anglicky „biased“). Zaujatosť môže byť napríklad historická, keď sa dlho opakované ľudské predsudky (rasové, rodové a pod.) premietnu aj do dát – umelá inteligencia bude na vedúce pozície odporúčať skôr mužov, pretože v minulosti ich naozaj zastávali najmä oni. Zaujatosť môže byť aj reprezentatívna – bude rozpoznávanie tváří fungovať správne v Afrike, keď sme systém umelej inteligencie natrénovali na tvárach ľudí zo strednej Európy?

Pokiaľ má umelá inteligencia hrať väčšiu úlohu v rozhodovacích procesoch, je dôležité, aby sme sa dokázali spoľahnúť na jej presnosť a nezaujatosť. Preto je veľmi dôležité, aby sme AI systémy učili na vhodných dátach.

Balíček Strojové učenie

Dataset

Strojové učenie s učiteľom

Zaujatosť

Posilňované strojové učenie



Materiál vytvoril tím neziskovej organizácie AI deťom v rámci projektu AI Kurikulum a podlieha licencií [Creative Commons 4.0 – Medzinárodná](#).

Preklad do slovenčiny zabezpečilo Ministerstvo školstva, výskumu, vývoja a mládeže Slovenskej republiky.



AI deťom

[Formulár na pripomienky.](#)



Pozn.: Rodová rovnosť je pre AI deťom kľúčová, ale na zostručenie využívame v našich metodikách formulácie v mužskom rode.

Aktivita na hodinu

Mimozemská detektívna kancelária – III. diel

45
minút

Opis aktivity

V predchádzajúcom cvičení sme v aplikácii Teachable Machine trénovali model strojového učenia pomocou dát dvoch mimozemských rodín (Fluffovcov a Earlovcov). Od modelu sme očakávali, že bude s veľkou mierou pravdepodobnosti správne klasifikovať obrázky nájdencov (pridelí ich do správnej rodiny). Zistili sme ale, že pre náš model boli zásadným rozlišovacím atribútom uši, preto nájdenca klasifikoval ako člena rodiny Earlovcov. Svoju teóriu sme si mohli overiť tak, že sme model testovali na obrázkoch ďalších nájdencov. Obrázok nájdenca bez uší model priraďoval do skupiny Fluffovcov. Inými slovami: hoci mal náš prvý nájdenc veľmi podobné atribúty s nájdencom bez uší (zdalo by sa, že patrí do rovnakej rodiny), model nakoniec oboch klasifikoval ako členov iných rodín.

Nasledujúca aktivita priamo nadväzuje na predchádzajúcu úlohu z karty Strojové učenie s učiteľom. V praktickom cvičení využijete predpripravený projekt do aplikácie Teachable Machine. Vďaka dobrodružnej expedícii na mimozemskej planéte sa však objavia ďalšie dve rodiny, ktoré do projektu zahrnieme, a tým rozšírite trénovacie dáta. Ani to však nepomôže správne priradiť nájdencov a model bude rovnako zaujatý. Zmenu konečne prinesie až expedícia do zakázaných končín planéty Yemi.

Ako súvisí detektívna kancelária s zaujatosťou

Ak natrénujeme model strojového učenia na dátach, ktoré nie sú jednoznačné a je ich príliš málo (tak ako je to v prípade oboch rodín), je preň veľmi ťažké rozpoznať (klasifikovať) nové dáta (nájdencov) správne. Uvidíme, že ani rozšírenie datasetu o dve ďalšie rodiny nepomôže. Atribút uší bude stále rozhodujúci, vďaka čomu bude model priraďovať nájdencov k nesprávnym rodinám. V „skutočnosti“ ale uši nemusia byť tým správnym atribútom. Môže to byť srst alebo iný detail, ktorý sme ani nezaznamenali. Model sa však napriek tomu podľa tohto detailu rozhoduje – je teda zaujatý.

Informácie o lekcii

Vstupné znalosti/ročníky, dĺžka lekcie

6. – 9. ročník ZŠ, 45 minút.

Stavebné kamene

Zaujatosť, dataset, strojové učenie.

Čo sa žiaci učia?

Systémy strojového učenia sú zaujaté.

Zaujatosť sa dá predchádzať testovaním modelu a úpravou datasetu.

Prečo sa to učia?

Cieľ pre balíček kariet Strojové učenie:

Kriticky posúdiť rozhodovanie systémov umelej inteligencie.

Ako spoznáme, že sa to naučili?

Vysvetlia pojem „zaujatosť“.

Pomôcky

Pedagóg: Súbory kartičiek s rodinami mimozemšťanov, projektor, prezentácie.

Žiaci: Počítač, notebook alebo tablet (ale nie s OS Android) pre každého žiaka alebo do skupiny.

Digitálne kompetencie

Informácie a komunikácia.

Bloomova taxonómia

Porozumenie: Rozumejú, ako nedostatočné alebo zle reprezentatívne dáta vedú k zaujatosti v AI modeloch.

Aplikácia: Používajú nástroj Teachable Machine na tréning a testovanie modelu, identifikujú prípady zaujatosti v dátach.

Súlad so štátnym vzdelávacím programom

ŠVP ZV 2023: Informatika

3. cyklus:

1. cieľ: Analyzovať problémy a rozvíjať abstrakciu a logiku.

5. cieľ: Zoznámiť sa s digitálnou reprezentáciou informácií rôznych typov (výkonový štandard: dekodovať informáciu z jednoduchých reprezentácií, vyhľadať a získať informácie v informačnom systéme a databáze).

6. cieľ: Spoznať princípy a koncepty informatiky, vrátane reprezentácie a spracovania údajov v tabuľkách, stromoch a grafoch.

8. cieľ: Používať sieťové zariadenia a nástroje na prácu v počítačovej sieti (výkonový štandard: diskutovať o správnosti a kvalite vyhľadaných informácií).

9. cieľ: Analyticky posudzovať digitálne technológie (výkonový štandard: diskutovať o digitálnych technológiách a ich vplyve na ostatných).

ŠVP 2015: Informatika - nižšie stredné vzdelávanie

Reprezentácie a nástroje – informácie

Reprezentácie a nástroje – štruktúry

Komunikácia a spolupráca – práca s webovou stránkou

Informačná spoločnosť - digitálne technológie v spoločnosti

ŠVP ZV 2023: AI gramotnosť

Súčasť balíčka	Strojové učenie
Vzdelávací cyklus	3. cyklus
Stručný obsah metodiky	Metodika priamo nadväzuje na kartu Strojové učenie s učiteľom. Žiaci pracujú s predpripraveným projektom v Teachable Machine a rozšíria tréningové dáta o dve nové mimozemské rodiny z expedície na planétu Yemi. Napriek rozšíreniu datasetu model naďalej nesprávne klasifikuje nájdenčov – stále sa zameriava na atribút uší. Žiaci tak priamo zažijú, čo znamená zaujatost' (bias) v modeloch strojového učenia a ako vzniká z nevyváženého alebo nedostatočného datasetu.
Komponenty AI gramotnosti	Komponent 1: Princípy fungovania AI Komponent 3: Kritické myslenie
Ciele AI gramotnosti	3. cyklus <ul style="list-style-type: none"> Porozumieť princípom fungovania AI (LLM, asistenti, agenti) a vplyvu kvality dát na zaujatost' výstupov. Kriticky overovať výstupy AI, rozpoznávať manipuláciu a dezinformácie.
Kľúčové obsahy a očakávané spôsobilosti	3. cyklus Dáta ako základ AI <ul style="list-style-type: none"> Analyzovať úlohu dát pri tréningu modelov a vysvetliť vzťah medzi kvalitou dát, zaujatostou a spoľahlivosťou výstupov. Posúdiť dôležitosť ľudského dohľadu nad dátami a výstupmi AI modelov. Procesný model <ul style="list-style-type: none"> Prakticky realizovať proces tvorby jednoduchého modelu strojového učenia – od prípravy tréningovej a testovacej sady údajov cez samotný tréning až po overenie úspešnosti modelu. Overovanie výstupov AI <ul style="list-style-type: none"> Identifikovať rôzne typy chýb a skreslení vo výstupoch AI (napr. halucinácie, stereotypy, etickú predpojatost' alebo nadmernú snahu AI vyhovieť používateľovi). Systematicky overovať správnosť a relevantnosť výstupov AI pomocou viacerých overených zdrojov.
Zdôvodnenie prepojenia	Metodika je dôležitou pre pochopenie zaujatosti AI modelov. Žiaci si na vlastnej skúsenosti overia, že ani rozšírenie datasetu nemusí stačiť, ak sú tréningové dáta nevyvážené alebo ak model zachytí nevhodný atribút ako rozhodujúci. Záverečná diskusia o reálnych prípadoch zaujatosti (historická, rodová, reprezentačná) prepája laboratórnu skúsenosť so spoločensky relevantnými dopadmi AI systémov. Týmto metodika priamo napĺňa spôsobilosť analyzovať vzťah medzi kvalitou dát, zaujatostou a spoľahlivosťou výstupov.
Poznámky a odporúčania	Metodika nadväzuje na kartu Strojové učenie s učiteľom. Odporúčame ju realizovať ako jej priame pokračovanie. Metodika obsahovo korešponduje s metodikami Ju a Pí 07: Zaujatost' v počítačovom videní a Ju a Pí 05: Učenie sa z príkladov, ktoré zaujatost' riešia na úrovni 2. cyklu. Pre úplné pokrytie spôsobilostí 3. cyklu odporúčame diskusiu rozšíriť o príklady historickej a rodovej zaujatosti z reálnych AI systémov (napr. bodovací systém, rozpoznávanie tváří).

Evokácia

10 minút

Vybav si

V predchádzajúcej lekcii sme v aplikácii Teachable Machine trénovali model strojového učenia pomocou dát dvoch mimozemských rodín (Fluffovcov a Earlovcov). Od modelu sme očakávali, že prideli nájdencia do správnej rodiny. Zistili sme ale, že pre náš model boli zásadným rozlišovacím atribútom uši. Ako sme si to overili?

Svoju teóriu sme preverili tak, že sme model testovali na obrázkoch ďalších nájdencov. Obrázok nájdenca bez uší model ako jediný priradil do skupiny Fluffovcov. Inými slovami: hoci mal náš prvý nájdenc veľmi podobné atribúty s nájdencom bez uší (zdalo by sa, že patrí do rovnakej rodiny), model ich nakoniec klasifikoval ako členov iných rodín.

Vysvetli

Model v aplikácii Teachable Machine sme trénovali pomocou tzv. strojového učenia s učiteľom. Opíš, ako tento typ strojového učenia funguje.

Stroje sa môžu, podobne ako ľudia, učiť z príkladov. Využívajú na to okrem iného metódu, ktorá sa nazýva strojové učenie s učiteľom. Zjednodušene to znamená, že nemusíme strojom dávať presné zadanie či postup (algoritmus). Namiesto toho im ukážeme príklady, z ktorých sa samy naučia vykonávať rôzne úlohy. Ľudia sú učiteľia, ktorí modelu hovoria, čo je čo – napríklad pomocou štítkov, triedenia do tried/kategórií a pod.

Zamysli sa

Stále nevieme, do akej rodiny nájdenc patrí. Čo by nám mohlo pomôcť?

Rozšíriť dataset o viac dát s obrázkami mimozemšťanov. Mať viac informácií o nájdencovi a rodinách. Snažiť sa nájdencia klasifikovať pomocou iných dát – okrem obrázkov napríklad pomocou DNA a pod.

Uvedomenie

25 minút

5 minút

Aktivita 1

Prezentácia strana 02

Žiaci si otvoria webovú stránku Teachable Machine: teachablemachine.withgoogle.com

- + Vyberú možnosť Get Started + Open an existing project from a file.
- + Poskytnite im predpripravený projekt „2-rodiny.tm“ – lnk.sk/ejzxw.
- + Žiaci tiež môžu načítať QR kód v prezentácii na strane 02.
- + Po nahratívidia projekt s dvoma triedami – Fluffovci a Earlovci.



Prezentácia strana 03

Príbeh pokračuje... (v prezentácii na strane 03).

Miestne médiá publikujú množstvo informácií z expedície do nepreskúmanej oblasti planéty Yemi. Zdá sa, že ľudia dokonca objavili nové rodiny mimozemšťanov! Jeden z našich kolegov v detektívnej kancelárii je zároveň aj spolupracovníkom expedície. Vďaka nemu získavame exkluzívne prístup k novým dôkazom!

10 minút



Prezentácia strany 04 a 05

Poskytnite žiakom nové dáta (odkaz a QR kód v prezentácii na strane 04).

Povedzte žiakom, nech sa oboznámia s novými dátami, ktoré poskytla expedícia. Je možné využiť niektoré dáta na vylepšenie nášho modelu a pokúsiť sa vyriešiť hádanku s priradovaním nájdenca k správnej rodine? Dáta nájdete na lnk.sk/opqgtz.

Žiaci doplnia dataset v nástroji Teachable Machine. Vytvoria nové triedy, pomenujú ich a nahrávajú do nich obrázky. Sledujte, či žiaci pri nahrávaní obrázkov volia vhodné obrázky (t. j. mimozemšťanov), alebo či pri tvorbe novej triedy vkladajú aj obrázky, ktoré do triedy logicky nepatria (rastliny, iné živočichy atď.). Nechajte žiakov vytvoriť dataset hocijako. Po vytrénovaní modelu a vyskúšaní klasifikácie nájdenca môžete s chybou pracovať a analyzovať, prečo neurónová sieť nájdenca stále priraduje do nesprávnej triedy (novú triedu môže byť potrebné vylepšiť, nevhodné obrázky z triedy odstrániť a pod).

5 minút



Prezentácia strana 06

Žiaci otestujú natrénovaný model pomocou pôvodných obrázkov nájdenca.

Pokiaľ nemáte uložené obrázky nájdenca z minulej lekcie, nájdete ich tu: lnk.sk/akt57. Teraz model väčšinou zaraďuje nájdenca do rodiny Kittinovcov. A čo tváre žiakov?

Prezentácia strana 07

Ukážte žiakom vysvetlenie zaujatosti, ktoré je v prezentácii na strane 07.

Rozprávajte sa o tom, kedy sú modely strojového učenia zaujaté. Napadajú žiakom podobné príklady ako s topánkami? Podobné napríklad môžu byť: ovocie a zelenina, tváre ľudí v Európe vs. tváre ľudí v Ázii a pod.

Reflexia

10
minútZdieľaj
a argumentuj**Do akej triedy tvoj model priradil nájdenca a prečo?**

Odpoveď pravdepodobne bude, že do rodiny Kittinovcov. Nájdenca bez uší zase stále určoval ako Fluffa.

Myslíš si, že v tom znova zohral úlohu výrazný atribút uší?

Podľa nášho testovania bol aj s novými dátami atribút uší stále rozhodujúci.

Model fungoval technicky správne (t. j. snažil sa rozpoznávať obrázky), ale napriek tomu nie sme schopní povedať, či priradil nájdenca do správnej rodiny. Ako je to možné?

Nemáme dostatok dát a nevieme, ako to v skutočnosti bolo. Preto ani nemôžeme potvrdiť, či model nájdenca priradil správne alebo nesprávne.

Keď sa pozrieme na naše skúsenosti s tréňovaním a testovaním modelu strojového učenia, môžeme podľa teba (je podľa teba správne) dôverovať AI systémom na 100 %? Prečo?

AI si vytvára vlastnú predstavu (reprezentáciu) na základe dát, ktoré poskytneme. Je však možné, že ľudia takéto dáta pripraví nesprávne, prípadne málo otestujú model, ktorý potom nemusí fungovať správne. Preto je veľmi dôležité kriticky uvažovať nad rozhodnutiami, ktoré sú automatizované pomocou AI.

Ako spoznáme, že je systém (model) umelej inteligencie zaujatý?

Na základe dôsledného testovania.

Ako by to nakoniec s nájdencom mohlo byť? Do ktorej rodiny podľa teba „skutočne“ patril?

Vyústenie príbehu: Počas expedície sa rodinu malého nájdenca vôbec nepodarilo objaviť, a preto tieto dáta v dataseťe úplne chýbali. Preto model nefungoval správne a priradil nájdenca do nesprávnych rodín. Rodina malého nájdenca na seba neskôr upozornila sama, pretože zachytila signál so správou, aby sa všetky rodiny na planéte Yemi prihlásili. Nájdenec teda nakoniec svoju rodinu našiel!

Prezentácia strana 08

Ukážte žiakom stranu 08 v prezentácii.

Nadväzujúce lekcie

V tejto lekcii žiaci zistili, čo je to zaujatost.

Príbeh mimozemskej detektívnej kancelárie sa končí. Nasleduje posledná lekcia z balíčka Strojové učenie, ktorá žiakom ukáže ďalší typ strojového učenia – posilňovaného. To je taký typ strojového učenia, vďaka ktorému sa stroje neučia z príkladov, ale zo skúseností.