

Balíček
Strojové učenie

Karta

Dataset

01



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, balíček Strojové učenie

Dataset

Základné informácie o datasetoch

Predstavte si dataset (dátovú množinu) ako album plný rôznych obrázkov. Každý obrázok má svoj opis, ktorý hovorí, čo na ňom je – napríklad „mačka na slnku“, „pes v snehu“ alebo „rodina na pláži“. Keď chceme naučiť systém umelej inteligencie rozpoznávať, čo je na obrázkoch, dávame mu takýto album na prehliadnutie (vo forme datasetu). Systém si „prečíta“ naše opisy a na obrázkoch hľadá časté vizuálne vzory (podobnosti), napríklad biely sneh a podobne. Čím viac obrázkov a opisov album obsahuje, tým lepšie sa systém naučí rozpoznávať rôzne veci na obrázkoch.

Definícia pre pokročilých

Dataset je obvykle veľké množstvo dát, ktoré sa používajú na tréning, testovanie a validáciu systémov strojového učenia (tzv. modelov strojového učenia). Môžu to byť napríklad hlasové záznamy, hudba, videá, obrázky či súbory čísel. Výber datasetu z veľkej časti definuje, čo bude model vedieť, ako sa bude správať a aké úlohy bude riešiť. Dataset by mal v ideálnom svete obsahovať kvalitné a pre danú úlohu relevantné dáta, aby sa zabezpečilo, že model bude fungovať dobre. Niekedy ale vývojári uprednostnia kvantitu pred kvalitou.

Balíček Strojové učenie

Dataset

Strojové učenie s učiteľom
Zaujatosť
Posilňované strojové učenie



Pozn.: Na zjednodušenie v tejto metodike hovoríme o datasetoch iba v kontexte obrázkov a strojového učenia s učiteľom.

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.

[Formulár na pripomienky.](#)



AI deťom

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

Aktivita na hodinu

Mimozemská detektivná kancelária

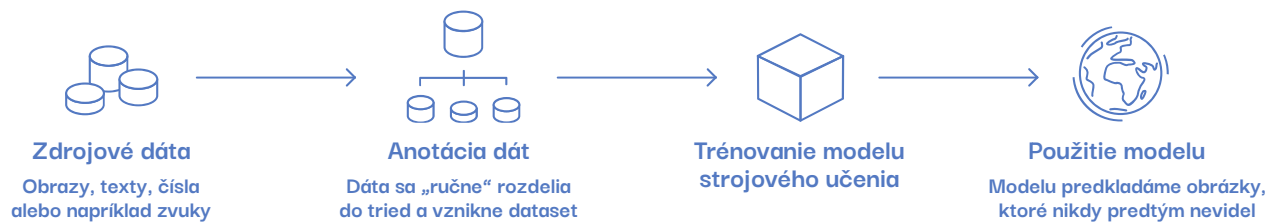
45 minút

Opis aktivity

Žiaci sa v rámci aktivity zahrajú na mimozemskú detektívnu kanceláriu. Ich úlohou je vypátrať, ku ktorej mimozemskej rodine patrí malý nájdenec, ktorý sa náhle objavil v ich dedine. Žiakom rozdáte (alebo premietnete) portréty, na ktorých sú členovia mimozemských rodín Fluffovcov a Earlovcov. Úlohou žiakov bude správne rozdeliť členov rodín a opísať ich vizuálne vlastnosti. Z toho následne odvodia, ku ktorej rodine malý nájdenec patrí. Žiadna odpoveď nie je zlá a existuje dokonca aj možnosť, že nájdenec je vďaka spoločným atribútom (vlastnostiam) potomkom členov oboch rodín.

Ako detektívna kancelária súvisí s datasetmi

Dataset (známy aj ako dátová množina či dátový súbor) sú dáta, ktoré sa používajú pri tréovaní, validácii a testovaní systémov umelej inteligencie. Môžu to byť obrázky (ale aj audio, video či tabuľky), ktoré sú v našom prípade reprezentované vygenerovanými obrázkami mimozemšťanov. Na natrénovanie takéhoto systému (modelu strojového učenia) je najskôr potrebné dáta opísať (anotovať). To urobíme tak, že ich rozdelíme do tzv. tried (tu sú to rodiny mimozemšťanov) a tieto triedy pomenujeme (tu sú to rodiny Fluffovcov a Earlovcov). Model strojového učenia sa trénuje na hľadanie podobností (sú to vzory, ktorým my ľudia hovoríme napríklad uši, fúzy, chlpy, farby...). Na základe takto identifikovaných vzorov model následne aj pri obrazoch, ktoré pri učení „nevidel“, rozhodne, do ktorej triedy patria (rodina Fluffovcov a Earlovcov).



Informácie o lekcii

Vstupné znalosti/odporúčané ročníky, dĺžka lekcie

6. – 9. ročník ZŠ, 45 minút.

Stavebné kamene

Dataset

Čo sa žiaci učia?

Dataset sú dáta pripravené na tréning systémov umelej inteligencie.

Prečo sa to učia?

Cieľ pre balíček kariet Strojové učenie:

Kriticky posúdiť rozhodovanie systémov umelej inteligencie.

Ako spoznáme, že sa to naučili?

Vysvetlí pojem „dataset“ a opíše postup jeho vytvárania.

Pomôcky

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

Digitálne kompetencie

Informácie a komunikácia.

Bloomova taxonómia

Zapamätanie: Žiaci sa zoznámia s pojmom dataset a jeho významom.

Analýza: Rozlišujú významné atribúty u jednotlivých tried mimozemšťanov a určujú, ktoré kľúčové znaky mali zásadný vplyv na klasifikáciu nájdených.

Súlad so štátnym vzdelávacím programom

ŠVP ZV 2023: Informatika

3. cyklus:

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

3. cieľ: Skúmať a posudzovať správnosť riešenia, riešiteľnosť úlohy a opravovať chyby.

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

Komunikácia a spolupráca – vyhľadávanie na webe

Informačná spoločnosť - digitálne technológie v spoločnosti

Evokácia

10 minút

Diskutuj

Prezentácia strana 03

Ako sa orientuješ v supermarkete, keď ideš na nákup?

Žiaci budú pravdepodobne odpovedať, že sa ľahšie orientujú vďaka jednotlivým oddeleniam, v ktorých nakupujú výrobky (jedlo, elektroniku, drogeriu, oblečenie atď.). Diskusia by mala žiakov viesť k poznatku, že ľudia triedia veci do kategórií, napríklad podľa vlastností daných vecí, spôsobu ich použitia (jedlo, nástroje, oblečenie, kozmetika a pod.) a mnohých ďalších vecí. Žiaci môžu samostatne nachádzať ďalšie princípy usporiadania vecí, a to nielen v supermarkete, ale napríklad aj v online obchode s aplikáciami a podobne.

Zásadný poznatok by mal byť aj ten, že veci opisujeme a vďaka tomu ich dokážeme triediť, filtrovať a hľadať medzi nimi podobnosti či súvislosti.

Uvedomenie

25 minút

Aktivita 1

25 minút

V tejto aktivite pracujte s pracovnými listami. Odporúčame ich pred hodinou rozstrihať a žiakom ich rozdávať postupne. Ak ich nechcete tlačiť, stačí využiť prezentáciu a vytlačené tabuľky zo strany 08 tejto metodiky.

Pracovný list 1
horná časť

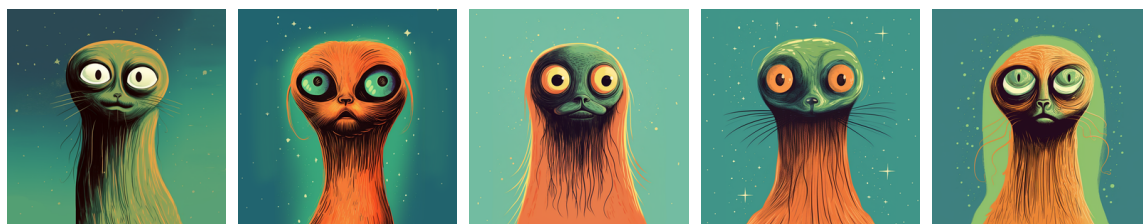
5 minút

Prezentácia strana 05 a 06

Rozdajte do skupín hornú časť pracovného listu 1 s kartami mimozemšťanov: rodín Fluffovcov a Earlovcov. Ak karty nechcete tlačiť, premietnite ich v prezentácii (strana 05). Žiaci v rámci skupín diskutujú o tom, ako by členov zatriedili. Hľadajú odlišnosti a spoločné vizuálne atribúty (vlastnosti). Členov rozdeľia podľa svojho uváženia.

Pozn.: Podľa našej skúsenosti žiaci zaraďujú členov rodín najprv chronologicky, čo je v poriadku. Časom je ale dobré, keď prejdú k deleniu do dvoch rodín (strana 06).

Rodina Fluffovcov



Rodina Earlovcov





V prostrednej časti pracovného listu 1 majú žiaci pripravenú tabuľku. Po tom, čo roztriedili členov do dvoch rodín, je ich úlohou zaznamenať do tabuľky pri každej rodine aspoň tri spoločné znaky, podobnosti, vlastnosti rodín mimozemšťanov. Môžu to byť napríklad uši, chlpy či farba.

Rodina Fluffovcov – spoločné atribúty Nájdeneč

Celkový počet bodov _____

Rodina Earlcoov – spoločné atribúty Nájdeneč

Celkový počet bodov _____



Prezentácia strana 07

Potom ukážte portrét malého nájdeneča. Ak žiaci nepracujú s pracovnými listami, premietnite prezentáciu s portrétom malého nájdeneča na strane 07. Podľa zvolených atribútov z predchádzajúcej časti aktivity žiaci obkresľujú v tabuľke obrázok nájdeneča. Ak sa niektorý z atribútov rodiny a nájdeneča zhoduje, zapíšu v stĺpci „Nájdeneč“ jeden bod. V prípade, že sa atribút rodiny a nájdeneča odlišuje, zaznačia nulu.

Potom žiaci sčítajú všetky získané body a podľa vyššieho počtu bodov rozhodnú o pridelení nájdeneča do konkrétnej rodiny.



Žiaci prezentujú svoje výsledky ostatným skupinám. Aké spoločné vlastnosti vybrali v jednotlivých rodinách? Vnímali ich ako jednoznačné (rovnaké) pre všetkých členov rodiny? Zvolili by spätne iné atribúty? Bol malý nájdeneč odovzdaný do správnej rodiny?



Prezentácia strana 08 a 09

Nechajte žiakov, nech si prečítajú prezentáciu na stranách 08 a 09, a potom diskutujte. Žiaci môžu zhrnúť, čo na hodine robili a ako to súvisí s datasetmi. Nie je potrebné používať správnu terminológiu, ale je dôležité pochopiť tento jednoduchý princíp: Keď sa program umelej inteligencie trénuje, treba mu ukázať zdrojové dáta a tie opísať, teda rozdeliť do kategórií. Takto vznikne dataset. Dataset sú dáta pripravené na tréning programu umelej inteligencie.



[Tesla autopilot](https://www.youtube.com/watch?v=fKXztwtXaGo)

Pustite žiakom video (0:31).

Adresa: www.youtube.com/watch?v=fKXztwtXaGo

Vo videu môžu žiaci vidieť, ako program samojazdiaceho auta vníma svoje okolie. Okolo rozpoznávaných objektov vykreslí ohraničovacie rámčeky. Na to, aby mohol program rozpoznávať objekty, bolo nutné natrénovať ho na veľkom počte príkladov objektov, ako sú napríklad chodci, cyklisti či dopravné značky. Ľudia teda vytvorili dataset, v ktorom tieto objekty rozdelili do kategórií, a program v nich pri tréningu hľadal podobnosti.

Reflexia

10
minútOpíš
a vyhodnoť

Vráťte sa k téme lekcie a podel'te sa o odpovede:

Čo ste sa naučili o datasetoch a ich triedení?

Prečo je dôležité správne definovať kategórie a atribúty?

Programy umelej inteligencie môžu automaticky rozlišovať, čo je na obrázkoch. Musia ale od ľudí dostať správne dáta. Opíš, ako a čo musia ľudia pre systém umelej inteligencie pripraviť.

Na to, aby systémy umelej inteligencie dokázali rozlišovať, čo kam patrí, ľudia pre ne najprv musia pripraviť dáta (obrázky, zvuky a pod.) a tie obrázky správne opísať, rozdeliť do kategórií. Takto pripraveným dátam sa hovorí dataset.

Napadajú ti príklady situácií, kde umelá inteligencia slúži podobne ako v samojazdiacom aute?

Aké ťažké by pre ňu bolo vytvoriť dataset?

Možné odpovede: Rozpoznávanie tvárí z dôvodu bezpečnosti na letiskách alebo napríklad na odomknutie smartfónu, automatizovaná kontrola kvality na výrobných linkách (keď je systém natrénovaný na defektných výrobkoch)...

Niekedy je svet veľmi zložitý. Na to, aby programy umelej inteligencie dokázali niektoré objekty rozpoznať, musia byť trénované na naozaj veľkom množstve dát. Čo si myslíš, koľko bolo treba obrázkov na to, aby samojazdiace auto správne rozpoznalo chodcov či napríklad priechod pre chodcov?

Boli to milióny obrázkov.

Nadväzujúce lekcie

V tejto lekcii žiaci pracovali s pracovnými listami, aby porozumeli teoretickému konceptu datasetov.

V nadväzujúcej lekcii budú ďalej využívať tému mimozemskej detektívnej kancelárie, ale rozhodnutie, kam malý nájdenec patrí, nechajú na program umelej inteligencie.

Diskutujte v skupinách o tom, čím sa jednotliví mimozemšťania líšia a čo majú spoločné. Podľa svojho uváženia potom rozdeľte členov do dvoch rodín - Fluffovcov a Earlovcov.



Rodina Fluffovcov – spoločné atribúty

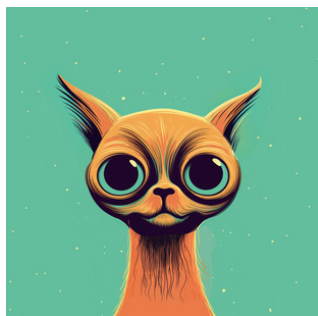
Nájdeneč

Celkový počet bodov _____

Rodina Earlovcov – spoločné atribúty

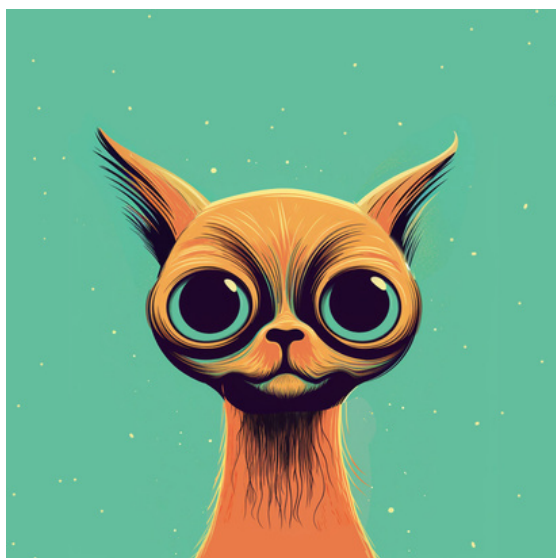
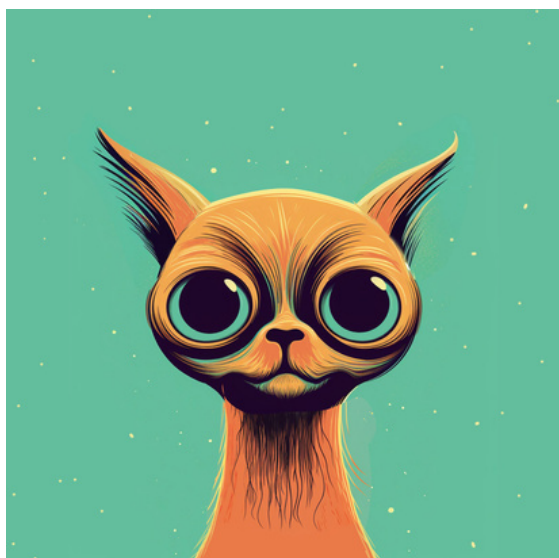
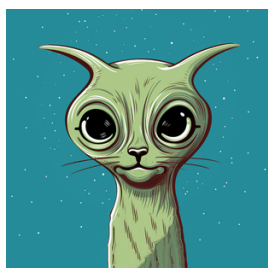
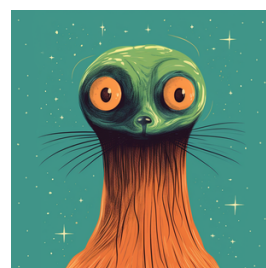
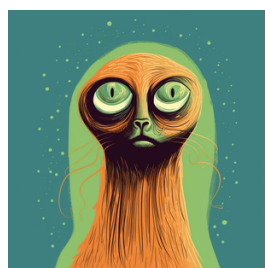
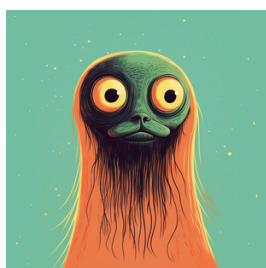
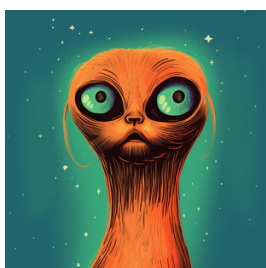
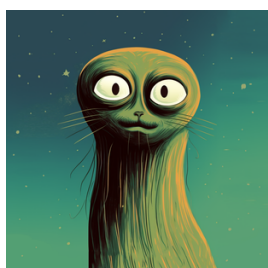
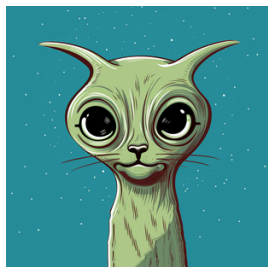
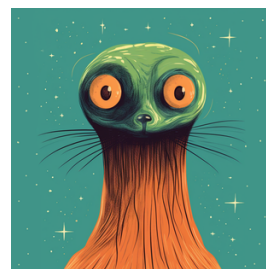
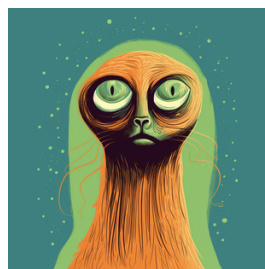
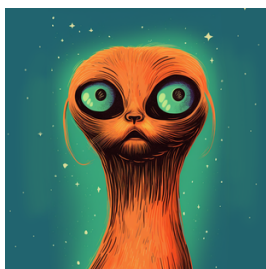
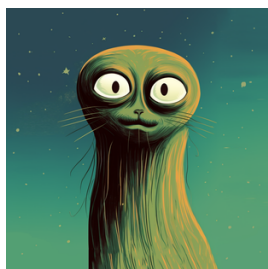
Nájdeneč

Celkový počet bodov _____



Pozrite sa na portrét malého nájdeneča.

Podľa zvolených vlastností z predchádzajúcej časti aktivity obodujte v tabuľke obrázok nájdeneča. Ak majú nájdeneč a rodina niečo spoločné, zapíšte v stĺpci „Nájdeneč“ jeden bod. V prípade, že sa vlastnosť rodiny s nájdencom nezhoduje, zaznamenajte nulu. Následne sčítajte všetky získané body. Najvyšší počet bodov rozhodne, do ktorej rodiny nájdeneč patrí. Čo však v prípade, že to vyjde nerozhodne? 😊



Rodina Fluffovcov – spoločné atribúty

Nájdeneč

Celkový počet bodov _____

Rodina Earlovcov – spoločné atribúty

Nájdeneč

Celkový počet bodov _____



Rodina Fluffovcov – spoločné atribúty

Nájdeneč

Celkový počet bodov _____

Rodina Earlovcov – spoločné atribúty

Nájdeneč

Celkový počet bodov _____



Rodina Fluffovcov – spoločné atribúty

Nájdeneč

Celkový počet bodov _____

Rodina Earlovcov – spoločné atribúty

Nájdeneč

Celkový počet bodov _____