



kurikulum

Metodický materiál Kurikula umělé inteligence
pro základní a střední školy – informatika na 2. stupni

Karta

Dataset

5.–9. ročníky a SŠ
30–45 minut

02



kurikulum.aidetem.cz/informatika2

Vypracovali: Radek Špáta a Eva Nečasová
Pilotáž: Radek Špáta, Eva Nečasová
Odborní garanti: XXX
Výstupy RVP doplnila: XXX
Jazyková korektura: Marcela Wimmerová
Poslední aktualizace: XX/2024



Formulář pro
připomínky

Metodický materiál Kurikula umělé inteligence pro základní a střední školy
Informatika na 2. stupni ZŠ a SŠ – karty

Dataset

Toto by bylo dobré vědět o datasetu pro začátek

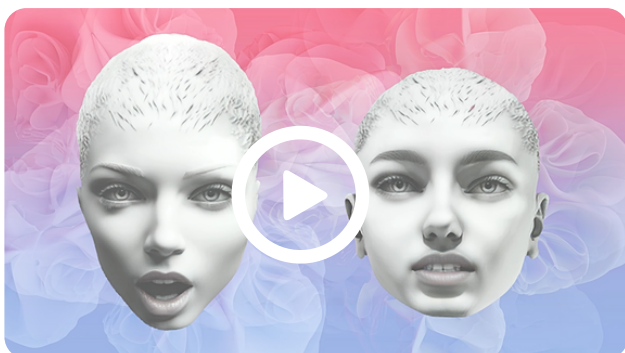
Představte si dataset jako album plné různých obrázků. Každý obrázek má svůj popis, který říká, co na něm je – například „kočka na slunci“, „pes ve sněhu“ nebo „rodina na pláži“. Když chceme naučit systém umělé inteligence rozpoznávat, co je na obrázcích, dáváme mu takové album „prohlédnout“ (ve formě datasetu). Systém si „přečte“ naše popisy a na obrázcích hledá časté vizuální vzory (podobnosti) – například sníh je bílý apod. Čím více obrázků a popisů album obsahuje, tím lépe se systém naučí rozpoznávat různé věci v obrázcích, které předtím neviděl.

Definice pro pokročilé

Dataset je obvykle velké množství dat, která se používají pro trénování, testování a validaci systémů strojového učení (tzv. modelů strojového učení). Mohou to být třeba hlasové záznamy, hudba, videa, obrazy... Volba datasetu z velké části definuje, co bude model umět, jak se bude chovat a jaké úlohy řešit. Dataset by měl obsahovat kvalitní a pro danou úlohu relevantní data, aby se zajistilo, že model bude dobře fungovat.

Více o datasetu a datech obecně naleznete také v naší příručce [Obecný úvod do umělé inteligence](#) →

Video ke kartě Dataset



Další materiály ke kartě Dataset

- 1) Pracovní list 1 ([link](#))
- 2) Prezentace ([link](#))

Související karty

- 03 Strojové učení
- 04 Strojové učení z příkladů
- 05 Strojové učení ze zkušeností
- 06 Předpojatost

[Hledáte podporu?](#)

Nevíte si rady? Připojte se do [FCB skupiny AI dětem](#) a zeptejte se komunity nebo správců.



Aktivita do hodiny

Mimozemská detektivní kancelář

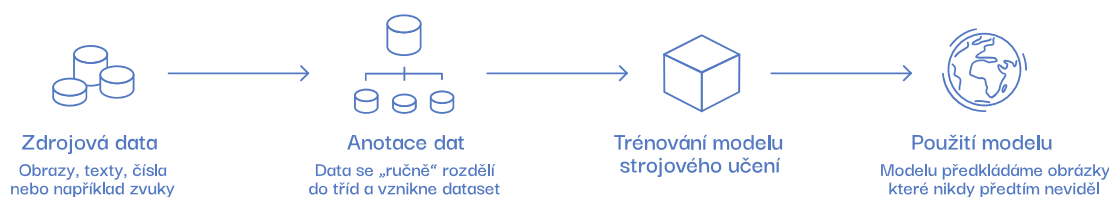
30–45
minut

Popis aktivity

Žáci si v aktivitě zahrají na mimozemskou detektivní kancelář. Jejich úkolem je vypátrat, ke které mimozemské rodině náleží malý nalezenec, který se náhle objevil v jejich vesnici. Žákům rozdáte (nebo promítnete) portréty, na nichž jsou členové mimozemských rodin Flufů a Earlů. Úkolem žáků bude správně rozdělit členy rodin a popsat jejich vizuální atributy. Z toho následně odvodit, ke které rodině malý nalezenec patří. Žádná odpověď není špatná, ale nabízí se, že nalezenec může být – díky společným atributům – potomkem členů obou rodin.

Jak souvisí detektivní kancelář s datasetem

Dataset jsou data, které se používají při trénování, validaci a testování systémů umělé inteligence. Mohou to být obrázky (ale také audio, video, tabulky...), které jsou v našem případě reprezentovány vygenerovanými obrázky mimozemšťanů. Abychom takový systém (model strojového učení) natrénovali, je třeba data nejprve popsat (anotovat). To uděláme tak, že je rozdělíme do tzv. tříd (rodiny mimozemšťanů) a tyto třídy pojmenujeme (rodina Flufů a Earlů). Když se model strojového učení trénuje, hledá podobnosti (jsou to vzory, které my lidé můžeme nazývat například uši, fousy, chlupy, barvy...). Na základě takto identifikovaných vzorů model následně rozhodne u obrázků, které při učení „neviděl“, do které třídy patří (rodina Flufů a Earlů).



Informace o aktivitě

Prekoncepty/ročníky, délka lekce

5.–9. ročníky a SŠ, 30–45 minut

Co se žáci učí?

Dataset mohou být lidmi popsané obrázky, z nichž se systémy umělé inteligence učí rozpoznávat jiné, které předtím „neviděly“. Aby takové rozpoznávání fungovalo správně, dataset musí být dobře připravený.

Proč se to učí?

Porozumí, že vzhledem k nejednoznačnosti a komplexitě světa je často velmi složité připravit data pro systémy umělé inteligence správně.

Jak poznáme, že se to naučili?

Žáci vysvětlí pojem dataset a popíší problém nejednoznačnosti dat v datasetu.

Pomůcky

- Vytištěné karty s mimozemšťany pro každou skupinu: sadu rodiny Flufů, sadu rodina Earlů a malého nalezence
- Bodovací tabulky
- Prezentaci

Výstupy RVP – Informatika

I-9-1-04 zhodnotí, zda jsou v modelu všechna data potřebná k řešení problému; vyhledá chybu v modelu a opraví ji

Digitální kompetence

Informace a komunikace – získává, vyhledává, kriticky posuzuje, spravuje a sdílí data, informace a digitální obsah, k tomu volí postupy, způsoby a prostředky, které odpovídají konkrétní situaci a účelu.

Pět velkých myšlenek

- 1-B-II Zpracování (extrakce příznaků)
- 2-A-II Symbolická reprezentace
- 3-A-II Podstata učení (hledání vzorů v datech)

Bloomova taxonomie

XXX

Evokace

Diskutuj

Promítněte obrázek supermarketu z prezentace.

Jakým způsobem se orientujete v supermarketu, když jdete nakoupit?

Žáci budou pravděpodobně odpovídat, že se snáze orientují díky jednotlivým oddělením, ve kterých nakupují produkty (jídlo, elektroniku, drogerii, oblečení...). Diskuze by měla žáky vést k poznatku, že lidé třídí věci do kategorií např. na principu vlastností daných věcí, způsobu jejich užití (jídlo, nástroje, oblečení, drogerie, apod.) a řady dalších. Žáci mohou samostatně nalézat další principy uspořádávání věcí. A to nejen v supermarketu, ale třeba také v online obchodě s aplikacemi apod.

Zásadní poznatek by měl být také ten, že věci popisujeme a díky tomu je dokážeme třídit, filtrovat, hledat podobnosti a souvislosti.

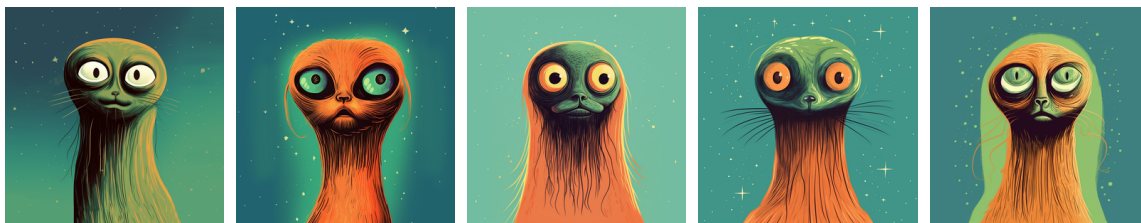
Uvědomění

Aktivita 1

Rozdejte do skupin Pracovní list 1 s kartami mimozemšťanů – rodin Flufů a Earlů. Případně promítněte karty v prezentaci, nechce-li je tisknout. Žáci diskutují uvnitř skupin, čím se jednotliví členové rodin liší a které vizuální atributy mají společné. Dle svého uvážení rozdělí členy do dvou rodin.

Pracovní list 1
(horní část)

Rodina Flufů



Rodina Earlů

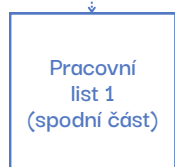




Ve spodní části PL1 mají žáci připravenou tabulku. Poté, co roztřídili členy do dvou rodin mají za úkol zaznamenat do tabulky u každé rodiny alespoň tři společné atributy (znaky) rodin mimozemšťanů.

Pokud žáci nepracují s pracovními listy, promítněte stránku prezentace s portrétem malého nalezence. Podle zvolených atributů z předchozí části aktivity žáci obodují v tabulce obrázek nalezence. Jestliže se shoduje některý z atributů rodiny a nalezence, zapíše ve sloupci „Nalezenec“ jeden bod. V případě, že atribut rodina s nalezcem nesdílí, zaznamenají nulu.

Poté sečtou všechny získané body a podle vyššího počtu bodů rozhodnou o přidělení nalezence do konkrétní rodiny.



Rodina Flufů – společné atributy Nalezenec

Rodina Flufů – společné atributy	Nalezenec

Rodina Earlů – společné atributy Nalezenec

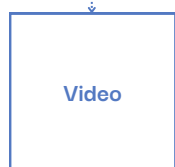
Rodina Earlů – společné atributy	Nalezenec

Celkový počet bodů _____

Celkový počet bodů _____



Žáci prezentují své výsledky ostatním skupinám. Jaké společné vlastnosti vybrali v jednotlivých rodinách? Vnímali je jako jednoznačné (stejně) pro všechny členy rodiny? Zvolili by zpětně jiné atributy? Byl malý nalezenec předán do správné rodiny?



Pusťte žákům vysvětlující video ke kartě Dataset.

Reflexe

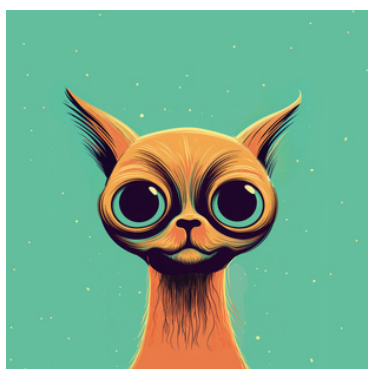


Jak jsi viděl/a ve vysvětlujícím videu, systémy umělé inteligence také mohou dělat podobná rozhodnutí (tzv. klasifikovat). Popiš jak a co musí lidé pro systém umělé inteligence připravit. Aby systémy umělé inteligence dokázaly rozhodovat co kam patří, lidé pro ně musí připravit data (obrázky, zvuky apod.) a ty obrázky správně popsat. Takto připraveným datům se říká dataset.

Někdy je ale svět hodně složitý a některá rozhodnutí je těžké udělat nejen pro systémy umělé inteligence, ale také pro nás lidi. Napadají tě některé příklady, kde by umělá inteligence mohla podobným způsobem sloužit. Jak těžké by bylo pro ni vytvořit dataset?

Možné odpovědi: Rozpoznávání tváří z důvodu bezpečnosti na letištích, automatizovaná kontrola kvality na výrobních linkách (když je systém natrénován na defektních výrobcích), rozpoznávání dopravních značek v samoříditelných autech...

Diskutujte ve skupinách, čím se jednotliví mimozemšťané liší a co mají společného. Dle svého uvážení poté rozdělte členy do dvou rodin – Flufové a Earlové.



Podívejte se na portrét malého nalezence.

Podle zvolených vlastností z předchozí části aktivity obodujte v tabulce obrázek nalezence. Jestliže má nalezenec a rodina něco společného, zapíše ve sloupci „Nalezenec“ jeden bod. V případě, že nic rodina s nalezencem nesdílí, zaznamenajte nulu. Následně sečtete všechny získané body. Nejvyšší počet bodů rozhodne, do které rodiny nalezenec patří. V případě, že to vyjde nerozhodně, tak...? :)

Rodina Flufů – společné atributy

Nalezenec

Rodina Earlů – společné atributy

Nalezenec

Celkový počet bodů _____

Celkový počet bodů _____