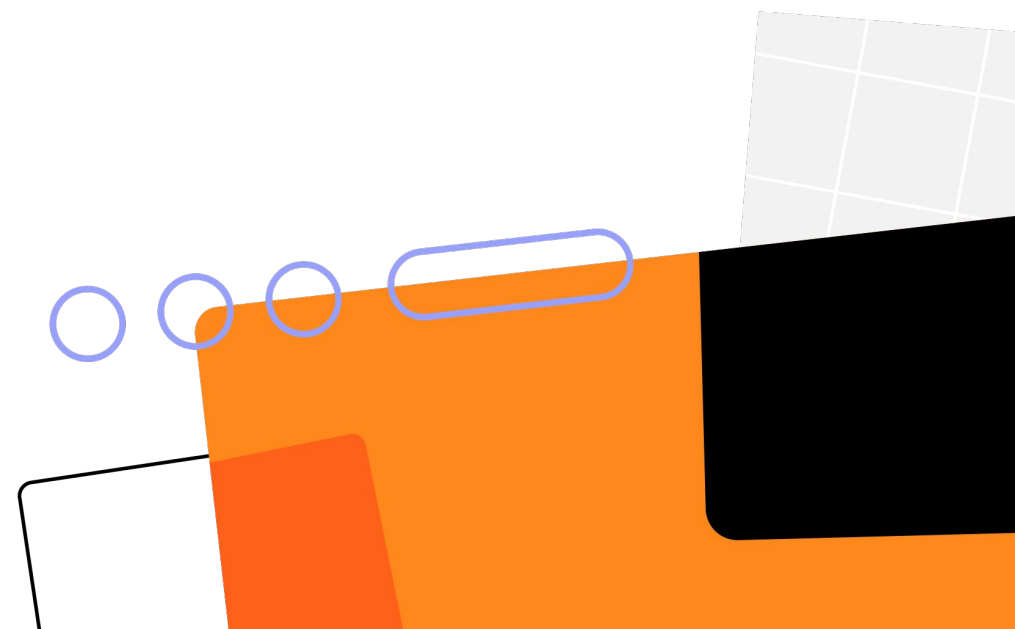




Datový vědec

Finální projekt

Kurz Data Science



Kde se nacházíme?

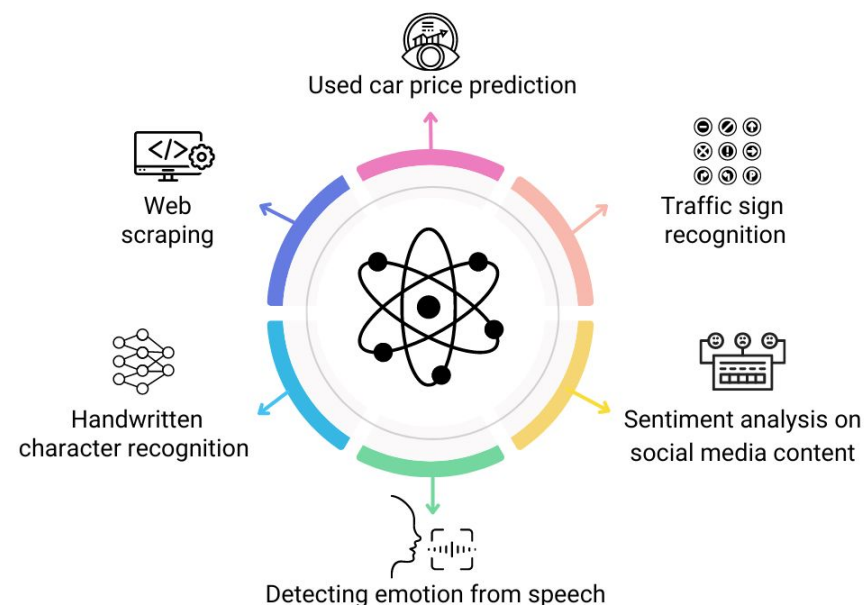
Úvod	Programování v Pythonu	Matematika & Datová analýza	Klasické strojové učení	Umělé neuronové sítě	Finální projekt
1-3	4-6	7-10	11-17	18-21	22
<ul style="list-style-type: none">★ Úvod do data science★ Nástroj GIT (video)★ Systémové nástroje, GIT správa verzí a projektový management	<ul style="list-style-type: none">★ Základy v Pythonu★ Python pro mírně pokročilé★ Python v data science	<ul style="list-style-type: none">★ Datové pipeline (samostudium)★ Matematický základ pro data science★ Zpracování dat & Explorativní datová analýza★ Vizualizace dat	<ul style="list-style-type: none">★ Klasické strojové učení★ Učení s učitelem: Regrese★ Cvičení: Regrese (projekt)★ Učení s učitelem: Klasifikace★ Cvičení: Klasifikace (projekt)★ Učení bez učitele★ Cvičení: Učení bez učitele (projekt)	<ul style="list-style-type: none">★ Úvod do umělých neuronových sítí★ TensorFlow knihovna & Keras API★ Trénování hlubokých neuronových sítí★ Hluboké učení: Přizpůsobené trénování & Aplikace	<ul style="list-style-type: none">★ Finální projekt

ML v praxi:

Učení bez učitele (projekt)

Praktické projekty hrají klíčovou roli ve výuce ML (Machine Learning = strojové učení), protože poskytují praktickou zkušenost a pomáhají upevnit vaše porozumění teoretickým konceptům včetně jejich aplikace v kontextu reálného světa.

Nyní je před vámi poslední a největší projekt tohoto kurzu! V rámci projektu můžete prohloubit a upevnit získané znalosti a pochlubit se svými nově nabytými dovednostmi. Pojd'te vytvořit další skvělý projekt pro své portfolio!



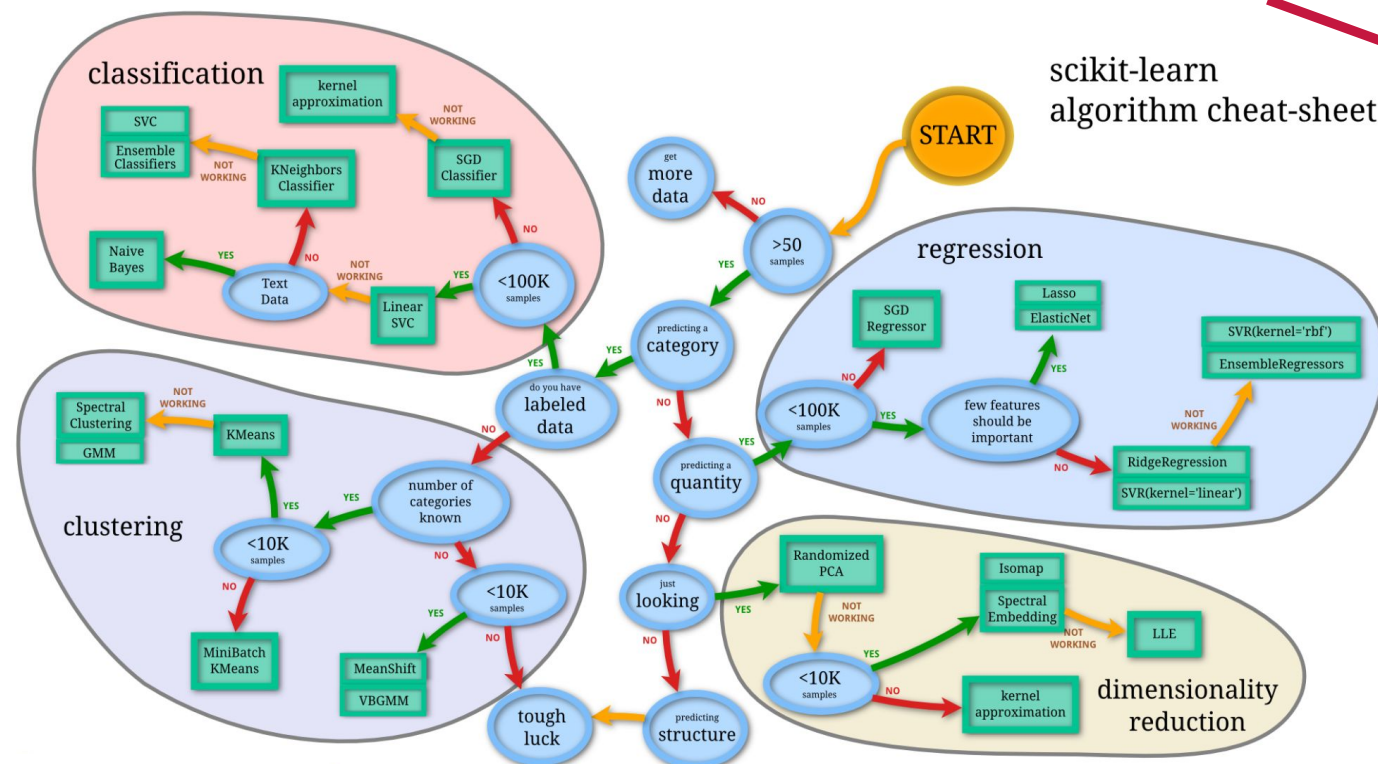
ML v praxi:

Učení bez učitele (projekt)

K vyřešení problému, který vás zajímá, můžete použít jakýkoli algoritmus ze scikit learn knihovny a/nebo architektury založené na neuronových sítích!

Stejně jako v minulosti prosím nezapomeňte

- poskytnout zdrojový kód v souboru s koncovkou .ipynb nebo .py
- nahrát svůj projekt na GitHub
- předvést projekt ve svém portfoliu nebo při pracovních pohovorech



ML v praxi:

Učení bez učitele(projekt)

1. týden (rozvrh)

- Rozdělení do skupin (maximálně 2-3 studenti)
- 1. den:
 - vybrat si projekt a stáhnout si data
 - ujasnit si požadavky
 - porozumění datům
 - předběžná analýza dat
- SOUHRN 1. DNE:
 - podrobný popis a prezentace hypotéz, které mají být testovány (jaké jsou sociální a obchodní aspekty?)
- 2. den:
 - exploratorní analýza dat
 - vizualizace dat
 - zpracování dat
 - úvodní testování algoritmů
 - první výsledky a hodnocení

SOUHRN 2. DNE



ML v praxi:

Učení bez učitele(projekt)

2. týden (rozvrh)

- Rozdělení do skupin (maximálně 2-3 studenti)
- 3. den:
 - testování dalších modelů
 - ladění hyperparametrů modelu
 - diskuse o tom, co by se mělo zlepšit
- 4. den:
 - poslední úpravy
 - příprava prezentace (slide deck)
- SHRUTÍ 2:
 - finální výsledky a závěrečná prezentace (pitch-deck)



ML v praxi:

Učení bez učitele(projekt)

Požadavky

Projekt lze dodat ve formě

- PowerPoint nebo Google Slides prezentace
- alternativně je možné projekt zpracovat také v Jupyter Notebooku nebo v Google Colab
- Python kód ve skriptech .ipynb nebo .py

➤ Projekt může být řešen

- na lokálním zařízení (počítač/notebook), nejprve je nutné nainstalovat všechny potřebné balíčky pomocí příkazu pip
 - pomocí Google Colab
- V určitém okamžiku by měl být kód refaktorován a vyčištěn, např. pomocí PyCharm nebo jiného IDE.
- Kód by měl být sdílen se všemi členy týmu prostřednictvím vzdáleného úložiště git, například na platformě GitHub.



ML v praxi:

Učení bez učitele(projekt)

Detailní popis

1. Rozdělení do pracovních skupin :
 - O rozdělení můžete rozhodnout vy sami nebo vás do skupin náhodně rozdělí trenér
2. Výsledkem této části by měla být krátká 1-2 stránková **úvodní analýza problému** z obchodního/obsahového hlediska:
 - seznámení se s vybraným souborem dat a popisem úkolu, který má být proveden
 - stažení požadovaných dat do počítače nebo na Google disk
 - definice toho, co o daném tématu víme
 - načtení dat, seznámení se s jejich strukturou a základními informacemi

Dále uveďte popis toho, co se vaše skupina chystá s datasetem dělat a jaká jsou omezení, tzn. co nelze udělat kvůli nedostatku informací nebo příliš velkému počtu odlehlých nebo chybějících hodnot.



ML v praxi:

Učení bez učitele(projekt)

Detailní popis

3. Příprava dat, která zahrnuje
 - extrakci číselných a kategorických příznaků
 - čištění dat
 - škálování dat
4. Seznam algoritmů strojového učení, které se mají použít k řešení vašeho problému. Vaším hlavním úkolem může být učení s učitelem (klasifikace nebo regrese) nebo učení bez učitele (clustering nebo redukce dimenzionality).



ML v praxi:

Učení bez učitele(projekt)

Detailní popis

5. Vyvození závěrů z předchozích kroků a na 1–2 slidech vysvětlit, proč jste použili tento konkrétní model, přičemž mějte na paměti:
 - rozlišení lineárních a nelineárních problémů
 - vliv statistik jako je korelace
 - odlehlé hodnoty
 - hodnocení výkonu
 - prediktivní schopnost



ML v praxi:

Učení bez učitele (projekt)

Shrnutí

1. Identifikace obtíží při práci v týmu.
2. Identifikace části, která způsobila největší problémy.
3. Identifikace témat, která vyžadují opakování nebo obnovení.
4. Rozvíjí způsob realizace projektu účastníky a pomáhá jim lépe porozumět problémům, které se dříve naučili?
5. Může vám způsob realizace projektu pomoci s nějakými budoucími problémy?



ML v praxi:

Učení bez učitele(projekt)

➤ Kaggle projekty!

- <https://www.kaggle.com/datasets/fedesoriano/air-quality-data-set>
- <https://www.kaggle.com/datasets/hasibalmuzdadid/global-air-pollution-dataset>
- <https://www.kaggle.com/datasets/usdot/flight-delays>
- <https://www.kaggle.com/datasets/saurabhshahane/twitter-sentiment-dataset>
- <https://www.kaggle.com/datasets/sanidhyak/human-face-emotions>
- <https://www.kaggle.com/datasets/trainingdatapro/cars-video-object-tracking>

➤ datasety

<https://github.com/matzim95/ML-datasets>