

Typy proměnných: Numerické (Continuální, Diskrétní), Kategorické (Nominální, Ordinální)

- Numerické – číselné
 - Continuální – může nabývat i necelých hodnot (věk, hmotnost)
 - Diskrétní – celá čísla (děti, zaměstnanci)
- Kategorické – kategorie nebo skupiny
 - Nominální – nemají žádné přirozené pořadí nebo hierarchie
 - Ordinální – mají přirozené pořadí nebo hierarchie

Vzorek vs. Populace

- Populace – soubor všech možných jednotek, o kterých máme zájem ve statistickém výzkumu
- Vzorek – podmnožina populace, kterou vybíráme nebo pozorujeme pro účely výzkumu

Charakteristika vzorku (Rozdělení Centralita, Rozptyl)

- Centralizace – informace o tom, kde se data ve vzorku nacházejí
 - Průměr – aritmetický průměr hodnot ve vzorku
 - Medián – dělí data na dvě stejně velké části, nachází se v prostředku uspořádaného vzorku
 - Modus – hodnota nejčastěji se vyskytující ve vzorku
- Rozptyl – informace o tom, jak jsou data ve vzorku rozprostřena kolem míry centralizace

Příprava dat v knihovně Pandas (DataFrame, read_csv, import csv z internetu)

- DataFrame – tabulky

```
import pandas as pd

# Vytvoření DataFrame ručně
data = {'Jméno': ['Alice', 'Bob', 'Charlie', 'David'],
        'Věk': [25, 30, 35, 40],
        'Město': ['Praha', 'Brno', 'Olomouc', 'Ostrava']}
df = pd.DataFrame(data)
```

- Read_csv – načtení dat (df = pd.read_csv('soubor.csv'))

- Z internetu

```
url = 'https://example.com/data.csv'
df = pd.read_csv(url)
```

Deskriptivní statistika v Pandas (počet, střední hodnota, medián, směrodatná odchylka, minimum, maximum)

```
# Počet řádků
count = df.count()
print("Počet řádků:\n", count)

# Průměr
mean = df.mean()
print("Průměr:\n", mean)

# Medián
median = df.median()
print("Medián:\n", median)

# Směrodatná odchylka
std_dev = df.std()
print("Směrodatná odchylka:\n", std_dev)

# Minimum
minimum = df.min()
print("Minimum:\n", minimum)

# Maximum
maximum = df.max()
print("Maximum:\n", maximum)
```

Vizualizace dat, typy grafů Vykreslení grafů v Pandas

- `df.plot(x='Jméno', y='Věk', kind='line', title='Věk podle jména')`
- podle toho jaký graf chceme měníme „kind“ element(pie, line, bar, ...)

Strojové učení Typy algoritmů strojového učení

- UČENÍ POD DOHLEDEM – Jsou nejčastěji používanými algoritmy strojového učení pro prediktivní analýzu, opírají se o soubory dat, které byly zpracovány lidskými odborníky
 - Regresní algoritmy – predikují hodnotu nového datového bodu na základě historických dat
 - Klasifikační algoritmy – využívají prediktivních výpočtů k zařazení dat do přednastavených kategorií
- UČENÍ BEZ DOHLEDU – Algoritmus sám uspořádá data a popíše jejich strukturu, algoritmus autonomně objevuje vzory v datech

Proces strojového učení

- Krok 1 - proces přípravy dat, převodu dat na strukturovaná, řešení chybějících dat, odstranění odlehlých pozorování
- Krok 2 - Vytvoření tréninkové sady dat, která bude použita k trénování modelu
- Krok 3 - Vytvoření testovací sady dat
- Krok 4 – Smyčka, je vybrán vhodný algoritmus na základě daného problému
- Krok 5 - Otestování modelu na testovacích datech se nazývá vyhodnocení modelu
- Krok 6 - Když model dosáhne uspokojivého výkonu na testovacích datech, model může být následně implementován

Regresní analýza

- Regresní analýza je statistická metoda pro nalezení vztahu mezi proměnnými. Tento vztah se vyjadřuje regresní funkcí, což umožňuje predikci hodnoty jedné proměnné na základě hodnot druhé. Využívá se k predikci budoucích hodnot, přičemž graficky je zobrazena vztahem mezi dvěma proměnnými.