

Predikce počasí – Souhrn projektu

1. Cíl projektu

Cílem projektu bylo vytvořit model, který na základě vstupních meteorologických dat (teplota, pocitová teplota, rychlost větru a část dne) dokáže předpovědět popis počasí, například *"overcast clouds"*, *"mist"*, *"light snow"*. Projekt zahrnoval i analýzu a vizualizaci dat pro lepší pochopení modelu a jeho výstupů.

2. Použité metody a poznatky

Příprava dat

- Načetl/a jsem meteorologická data a provedl/a jejich zpracování.
- Použil/a jsem LabelEncoder pro převod kategorií (části dne a popisy počasí) na číselné hodnoty.
- Rozdělil/a jsem data na vstupní proměnné (teplota, pocitová teplota, rychlost větru, část dne) a cílovou proměnnou (popis počasí).

Strojové učení

- Využil/a jsem RandomForestClassifier z knihovny scikit-learn jako klasifikační algoritmus.
- Rozdělil/a jsem data na trénovací a testovací sadu v poměru 80:20 (train_test_split).
- Vyhodnotil/a jsem model pomocí metrik jako přesnost (accuracy), matice záměn a klasifikační report (precision, recall, F1-score).

Vizualizace dat

- Vytvořil/a jsem pokročilé vizualizace:
 - **Sloupcový graf** porovnávající skutečné a predikované hodnoty.
 - **Heatmapu (matice záměn)** zobrazující chyby modelu.
 - **Interaktivní graf** zobrazující rozdělení počasí podle části dne pomocí knihovny Plotly.

Interpretace výsledků

- Pochopil/a jsem, že diagonála matice záměn reprezentuje správné predikce, zatímco hodnoty mimo ni ukazují chyby.
 - Model měl největší úspěšnost u predikce *"overcast clouds"* a *"light snow"*, ale například u *"light rain"* selhával kvůli nedostatku dat.
-

3. Odpovědi na klíčové otázky

- **Jak přesný je model?**
 - Model dosáhl přesnosti **93,22 %**.
 - **Které kategorie počasí model predikuje dobře a které špatně?**
 - Nejlepší výsledky byly u "*overcast clouds*" a "*light snow*".
 - Největší problémy měl model s "*light rain*", kvůli malému vzorku dat.
 - **Kde model nejčastěji chybí?**
 - Nejčastěji docházelo k záměně mezi "*mist*" a "*light snow*".
 - **Jaké jsou rozdíly v počasí během dne?**
 - Nejčastější popis počasí odpoledne byl "*overcast clouds*", zatímco ráno se častěji vyskytovalo "*light snow*".
-

4. Interpretace grafů

Sloupcový graf (skutečné vs. predikované hodnoty)

- Ukazuje, jak model predikoval jednotlivé kategorie počasí.
- Barvy reprezentují predikované hodnoty a jejich výška počet případů.
- Například u "*overcast clouds*" byla většina predikcí správná, ale objevily se i chyby (např. "*mist*").

Heatmapa (matice záměn)

- Zobrazuje chyby modelu.
- Diagonála ukazuje správné predikce (například **62 správných predikcí** pro "*overcast clouds*").
- Hodnoty mimo diagonálu ukazují chyby, například **3 případy**, kdy "*light snow*" bylo predikováno jako "*mist*".

Interaktivní graf (část dne a počasí)

- Ukazuje, jaké počasí je typické pro různé části dne.
 - Například "*overcast clouds*" převládá odpoledne, zatímco "*light snow*" je častější ráno.
-

5. Časový odhad

Celý projekt zabral **dva dny**:

- **Den 1:** Příprava dat a trénink modelu.
 - **Den 2:** Vizualizace a analýza výsledků.
-

6. Závěr

Projekt ukázal, jak lze pomocí strojového učení predikovat popis počasí na základě historických dat. Pracoval/a jsem s čištěním dat, trénoval/a model, analyzoval/a jeho výstupy a vytvořil/a vizualizace pro lepší pochopení. Výsledky ukázaly, že model funguje dobře pro některé kategorie, ale selhává tam, kde je málo dat. To poskytuje dobrý základ pro další vylepšení modelu.

Možné zlepšení do budoucna:

- Získání více dat pro méně zastoupené kategorie počasí.
 - Vyzkoušení jiných modelů, například neuronových sítí.
 - Optimalizace hyperparametrů modelu pro lepší přesnost.
-

Predikce počasí – Souhrn projektu

1. Cíl projektu

Cílem bylo naučit se předpovídat počasí bez nutnosti otvírat okno. Model měl na základě vstupních meteorologických dat (teplota, pocitová teplota, rychlost větru a část dne) hádat, jestli venku sněží, prší, je mlha, nebo jestli je ideální den na piknik.

Součástí bylo i zkoumání dat, protože když už se v nich člověk hrabe, chce z toho taky něco pochopit, že jo?

2. Co všechno jsem musel/a zvládnout a přežít

Příprava dat – aneb data potřebují řád, jinak je zmatek

- Nejprve jsem si stáhl/a meteorologická data a zjistil/a, že syrová data jsou jako bordel na stole – musí se uklidit.
- Kategorie jako "*Morning*", "*Afternoon*", "*Evening*" jsem převedl/a na čísla, protože modelu bohužel slovní popisy moc neříkají.
- Rozdělil/a jsem data na vstupní proměnné (teplota, pocitová teplota, vítr, část dne) a cílovou proměnnou (popis počasí).

Strojové učení – aneb nechť les rozhodne

- Na scénu nastoupil **RandomForestClassifier** – protože jeden strom je fajn, ale celý les už má co říct.
- Rozdělil/a jsem data na trénovací a testovací část (80:20), aby model neměl pocit, že mu servíruju jen to, co chce slyšet.
- Po pár testech jsem zjistil/a, že model si vede překvapivě dobře, ale občas si plete "*mist*" a "*light snow*". Možná by pomohlo, kdyby si taky někdy otevřel okno...

Vizualizace – protože bez grafů by to nebyla ta pravá datová magie

- **Sloupcový graf:** Jak moc si model věří vs. jak moc se pletl.
 - **Heatmapa:** Přehled chyb modelu – jinými slovy, kam ho poslat na doučování.
 - **Interaktivní graf:** Ukazuje, kdy je nejlepší jít ven a kdy se radši schovat pod deku s čajem.
-

3. Odpovědi na zásadní otázky

- **Jak přesný je model?** 93,22 % – není to věštecká koule, ale lepší než podívat se na oblohu a tipovat.
 - **Co predikuje dobře a kde tápá?**
 - "*Overcast clouds*" a "*light snow*" zvládá skvěle.
 - "*Light rain*" mu dělá problémy – asi proto, že se mu nechce moknout.
 - **Kde model nejvíc selhává?**
 - Plete si "*mist*" a "*light snow*" – což vlastně není tak špatné, když si uvědomím, že v zimě je to stejně všechno jen mokrá břečka.
 - **Jak se počasí mění během dne?**
 - Odpoledne vede "*overcast clouds*", zatímco ráno má převahu "*light snow*".
-

4. Grafy, aneb obrázky mluví za tisíc řádků kódu

Sloupcový graf (skutečné vs. predikované hodnoty)

- Barevná vizualizace toho, jak moc model ví, co dělá – nebo neví.
- Například u "*overcast clouds*" trefil skoro všechno správně, ale u "*mist*" si občas vymýšlel.

Heatmapa (matice záměn)

- Ukazuje, kde model exceluje a kde by si měl dát pár doučovacích lekcí.
- Hodnoty na diagonále = správné predikce. Hodnoty mimo = "ehm, no... skoro".

Interaktivní graf (část dne a počasí)

- Ideální na to, abych si naplánoval/a, kdy vzít pláštěnku a kdy sluneční brýle.
 - Ukazuje, že "*light snow*" dominuje ráno, zatímco odpoledne je větší šance na "*overcast clouds*".
-

5. Jak dlouho to celé trvalo?

Celkem **2 dny** (a pár káv):

- **Den 1:** Příprava dat a trénink modelu.
 - **Den 2:** Vizualizace, analýza a hluboké zamyšlení nad tím, proč model zaměňuje mlhu za sníh.
-

6. Závěr

Projekt mi ukázal, že předpovídat počasí není jen tak. Model byl docela přesný, ale občas prostě vařil z vody (doslova). Vizualizace pomohly pochopit, kde se model daří a kde selhává.

Co s tím dál?

- Sehnat víc dat pro ty kategorie, kde model tápe.
- Vyzkoušet hlubší neuronové sítě – třeba by to byla lepší věštecká koule.
- Optimalizovat model, protože co si budeme povídat, vždycky je co vylepšovat.