

# Návrh projektu: Predikcia a klasifikácia kvality vína pomocou metód strojového učenia

Lucia Lahučká

November 9, 2024

## Popis problému

Cieľom tohto projektu je vyvinúť modely strojového učenia (machine learning, ML) na predikciu a klasifikáciu kvality vína na základe jeho fyzikálno-chemických vlastností. Tento problém bude riešený v dvoch formách:

- **Regresia:** Predikcia hodnoty kvality vína na základe rôznych atribútov, ako sú kyslosť, obsah alkoholu, pH a obsah síry. Kvalita vína je kontinuálna hodnota medzi 0 a 10.
- **Klasifikácia:** Rozdelenie kvality vína do niekoľkých tried (napr. veľmi zlé, priemerné, vynikajúce) na základe tých istých fyzikálno-chemických vlastností.

Dáta použité v tomto projekte pochádzajú z datasetu \*Wine Quality\*, ktorý obsahuje údaje o červených a bielych vínach zo portugalskej oblasti Vinho Verde. Cieľom je aplikovať rôzne metódy strojového učenia na tento problém a porovnať výkonnosť modelov pri predikcii kvality vína.

## Dáta

Dataset \*Wine Quality\* obsahuje dve sady údajov: jednu pre červené vína (1599 vzoriek) a jednu pre biele vína (4898 vzoriek). Každá vzorka obsahuje 11 fyzikálno-chemických vlastností (ako sú kyslosť, obsah alkoholu, pH, obsah síry,...) a kvalitu vína, ktorá bola hodnotená odborníkmi na stupnici od 0 do 10. Kvalita vína bude v prípade regresie predikovaná ako konkrétna hodnota, a v prípade klasifikácie bude kvalita vína upravená do kategórií:

- **Veľmi zlé víno** (0-3)
- **Priemerné víno** (4-6)
- **Vynikajúce víno** (7-10)

## Metódy strojového učenia a rozdelenie dát

Na riešenie problému budeme používať rôzne metódy strojového učenia pre obe úlohy:

- **Lineárna regresia**
- **Support Vector Machines (SVM)**
- **Random Forest**
- **Neurónové siete**

Dátová sada bude rozdelená na tréningovú (70%) validačnú (20%) a testovaciu (10%) časť.

## Očakávané výsledky

Očakávame, že pokročilé modely ako Random Forest a SVM dosiahnu lepšie výsledky v porovnaní s jednoduchšou lineárnou regresiou a klasifikáciou. Ďalej očakávame, že výsledky klasifikácie budú ľahšie interpretovateľné, nakoľko budú vína klasifikované do konkrétnych kvalitatívnych tried.