

# Loess regrese a mnohonásobná regrese

Sergii Babichev

Univerzita Jana Evangelisty Purkyně v Ústí nad Labem

*sergii.babichev@ujep.cz*

# Úkol 1: Loess regrese

**Cíl:** Porovnat výkonnost LOESS regrese a polynomiální regrese při modelování vztahů mezi atributy v datové sadě **manufacturing.csv**, přičemž **Hodnocení kvality** je cílová proměnná.

## Postup:

- Rozdělení dat na trénovací a testovací množinu (70/30).
- Aplikace LOESS regrese na všechny atributy vzhledem k **Hodnocení kvality**.
- Aplikace polynomiální regrese na stejné atributy vzhledem k **Hodnocení kvality**.
- Hodnocení efektivity modelů pomocí RMSE a R-squared.
- Vizuální porovnání predikcí obou metod.

# Úkol 2: Mnohonásobná regrese s hodnocením statistické významnosti

1. **Načtení dat:** Stáhněte si datovou sadu *House\_price.csv* a načtěte ji do svého prostředí pro analýzu.
2. **Průzkum dat:** Prozkoumejte načtená data, identifikujte a odstraňte sloupce, které nejsou informativní pro vytvoření modelu.
3. **Analýza dat:**
  - Analyzujte data na přítomnost multikolinearity a nutnost normalizace.
  - Použijte statistické metody k identifikaci potřeby transformace dat.
4. **Vytvoření regresních modelů:**
  - Rozdělení dat na trénovací a testovací množinu (70/30).
  - Vytvořte model mnohonásobné regrese se všemi relevantními atributy.
  - Proveďte testování statistické významnosti koeficientů.
  - Analyzujte celkovou kvalitu modelu pomocí.
  - Vyhodnoťte výkon modelu pomocí metrik (RMSE,  $R^2$ ) a interpretujte výsledky.
5. **Porovnání výsledků.**