

# Zadanie projektu

## Téma:

Projekt je zameraný na analýzu a interpretáciu údajov o portfóliu o spotrebiteľských úveroch.

## Termín:

Nedeľa 24. 05. 2026.

Bonus + 5 bodov pri odovzdaní do 10. 5. 2026.

Možnosť výberu opravného odovzdania alebo bonusu + 5 bodov pri odovzdaní do 3. 5. 2026.

## Požadované výstupy:

- Dokument popisujúci hlavné závery k obom úlohám
- Krátke manažérske zhrnutie hlavných záverov v slovenskom alebo anglickom jazyku v rozsahu pol až jedna strana (high-level non-technical executive summary)
- Štruktúrovaný a komentovaný zdrojový kód.
- Kód musí byť možné samostatne spustiť a po spustení umožní replikáciu všetkých Vašich výsledkov.
- Priložte aj zoznam potrebných knižníc.

## Popis údajov:

- K dispozícii sú údaje o vyše 400 tis. spotrebiteľských úveroch poskytnutých spoločnosťou v USA uvedené v súbore loans.csv. Údaje je možné využiť iba pre účely tohto projektu, nie je povolené ich používať na iné účely ani ďalej šíriť.
- Údaje obsahujú údaj, či úvery zlyhal (t. j. dlžník ho nedokázal v plnej miere splatiť) alebo nie, ako aj viacero charakteristík úveru a dlžníkov. Obsahujú tiež hodnotenie rizika samotnou spoločnosťou (t. j. rating) v čase poskytnutia úveru (grade/subgrade).
- Zoznam parametrov je dostupný v súbore list.csv.
- Súbory so vstupnými údajmi sú dostupné tu:
  - <http://www.iam.fmph.uniba.sk/institute/jurca/qrm/loans.csv>
  - <http://www.iam.fmph.uniba.sk/institute/jurca/qrm/list.csv>

## Úlohy

### Úloha #1: Analýza a kontrola údajov

- Údaje pripravte na spracovanie, skontrolujte ich a v prípade potreby očistite. Uvedte zoznam úkonov, ktoré bolo potrebné pri príprave údajov vykonať.
- Analyzujte a vizualizujte štruktúru portfólia podľa relevantných charakteristík, príp. ich vzájomné vzťahy.

## Úloha #2: Faktory zlyhania úveru

- Identifikujte relevantné premenné, pomocou ktorých možno odhadnúť, ktoré úvery majú väčšiu pravdepodobnosť, že zlyhajú.
- Pri analýze môžete okrem premenných uvedených priamo v dátovej sade využiť rôzne odvodené alebo transformované premenné, dopočítané z dostupných údajov.

## Úloha #3: Odhad modelu pravdepodobnosti zlyhania

- Odhadnite logit model pravdepodobnosti zlyhania, vyberte štatisticky relevantné premenné.
- Ohodnotte predikčnú schopnosť modelu.
- Rady k tvorbe modelu:
  - Model vytvorte na tréningovej vybalancovanej vzorke, otestujte na testovacej vzorke
  - Medzi vysvetľujúcimi premennými by nemali byť premenné, ktoré už môžu byť výsledkom hodnotenia bonity na základe ostatných premenných (napr. úroková sadzba).

## Úloha #4: Aplikácia modelu

- Pokúste sa posúdiť, či stanovená úroková sadzba zodpovedá odhadnutým pravdepodobnostiam zlyhania, resp. či úrokové príjmy dostatočne kompenzujú straty v prípade zlyhania. Predpokladajte, že v prípade zlyhania je strata vo výške polovice nesplatenej časti úveru.