

# Predikce akademické úspěšnosti<sup>1</sup>

---

Akademická úspěšnost je považována za klíčový ukazatel kvality vysokoškolského vzdělávání, jelikož slouží jako jedno z hlavních kritérií pro hodnocení efektivity vzdělávacích institucí (National Commission for Academic Accreditation & Assessment, 2015). Přestože se v literatuře objevuje více pojetí tohoto pojmu, většina definic se shoduje na tom, že akademická úspěšnost zahrnuje nejen studijní výsledky, ale také spokojenost studentů, získané dovednosti a kompetence, setrvání ve studiu, dosažení vzdělávacích cílů a úspěch v profesní dráze po ukončení studia (Kuh et al., 2006; York et al., 2015).

Navzdory rostoucímu důrazu na komplexní pojetí tohoto konstruktů většina výzkumů z praktických důvodů redukuje akademickou úspěšnost na měřitelné akademické výsledky, obvykle ve formě průměrného studijního prospěchu (GPA/CGPA) nebo dosaženého počtu kreditů (Parker et al., 2004; Choi, 2005). Výzkumy využívají tuto variantu, jelikož operacionalizace umožňuje kvantitativní vyhodnocení výkonu studentů a srovnání napříč institucemi.

## Práce s daty

Pro účely binomiální logistické analýzy byla zvolena data autorů Realinho et al. (2021), kteří publikovali datový soubor na platformě [Kaggle](#). Dataset obsahuje detailní informace o socio-demografických charakteristikách a akademických výsledcích studentů vysoké školy v Portugalsku.

Stanovená závislá proměnná (v datech označena *Target*), měla tři možné hodnoty:

- **Graduate** (student úspěšně absolvoval)
- **Dropout** (student předčasně ukončil studium)
- **Enrolled** (student stále studuje)

Z metodologických i praktických důvodů byla kategorie Enrolled vyřazena z analýzy. Tato skupina studentů dosud neměla finální výsledek studia, což by vedlo ke zkreslení predikce. Testování logistické regrese ukázalo zásadní problém, model dokázal správně klasifikovat

---

<sup>1</sup> Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=580>.

pouze **2,1 %** případů v této kategorii. Její vyřazení proto výrazně zvýšilo predikční schopnost modelu. Po odstranění této kategorie činil finální vzorek **3 630** studentů.

Původní dataset obsahoval 36 proměnných, které mohly sloužit jako prediktory studentské úspěšnosti. Tyto proměnné zahrnovaly socio-demografické údaje, akademické ukazatele a proměnné související s finanční situací studenta. Vyloučili jsme irelevantní proměnné, ty s nízkou variabilitou, a také proměnné s chybějícími a nejasnými údaji. Po selekci jsme do modelu zahrnuli **10 nezávislých proměnných**, které jsou znázorněny níže.

| Proměnná                            | Typ proměnné              | Popis                                              |
|-------------------------------------|---------------------------|----------------------------------------------------|
| Marital status                      | Kategorická (6 kategorií) | Stav (svobodný, ženatý apod.)                      |
| Educational special needs           | Binární                   | Zda má student speciální vzdělávací potřeby        |
| Debtor                              | Binární                   | Zda má student dluhy vůči škole                    |
| Tuition fees up to date             | Binární                   | Zda má student uhrazené školné                     |
| Gender                              | Binární                   | Pohlaví                                            |
| Scholarship holder                  | Binární                   | Zda je student stipendista                         |
| Age at enrollment                   | Spojité                   | Věk při nástupu ke studiu                          |
| Curricular units 1st sem (approved) | Spojité                   | Počet úspěšně absolvovaných předmětů v 1. semestru |
| Curricular units 2nd sem (approved) | Spojité                   | Počet úspěšně absolvovaných předmětů v 2. semestru |
| Unemployment rate                   | Spojité                   | Nezaměstnanost v zemi v daném období               |

Tabulka 1 – Přehled regresorů

## Výsledky analýzy

Výsledky ukázaly, že nejvýznamnějšími faktory ovlivňujícími pravděpodobnost dokončení studia byly zejména **akademické výsledky**. Konkrétně se ukázalo, že čím více předmětů student úspěšně absolvoval v prvním a druhém semestru, tím nižší byla pravděpodobnost, že studium předčasně ukončí. Podobně i průměrná známka z těchto semestrů měla negativní vztah k riziku dropoutu. Lepší známky tedy rovněž predikovaly vyšší pravděpodobnost úspěšného dokončení studia.

Významnou roli hrála rovněž **finanční stabilita** studentů – ti, kteří měli školné řádně uhrazeno, měli výrazně nižší pravděpodobnost předčasného ukončení studia. Tento faktor může souviset s vyšší motivací, zodpovědností či celkovým organizačním zázemím studentů.

Naopak **věk** při nástupu ke studiu měl opačný efekt, kdy vyšší věk studentů byl spojen s mírně nižší pravděpodobností úspěšného dokončení studia. Tento výsledek lze interpretovat tak, že starší studenti mohou čelit více závazkům, například pracovním či rodinným, což může negativně ovlivnit jejich studijní výsledky a motivaci ke studiu.

Z hlediska **pohlaví** se projevilo, že muži mají statisticky významně vyšší pravděpodobnost úspěšného dokončení studia než ženy.

Celkově lze říci, že **akademické ukazatele** (počty úspěšně zvládnutých předmětů a známky), finanční stabilita, věk a pohlaví patřily mezi klíčové faktory ovlivňující úspěšnost studia.

| Efekt                     | Odhad šancí                      | Waldova statistika               | P-Hodnota                        |
|---------------------------|----------------------------------|----------------------------------|----------------------------------|
| Marital status            | Pro všechny kategorie nevýznamné | Pro všechny kategorie nevýznamné | Pro všechny kategorie nevýznamné |
| Educational special needs | 1,325408                         | 0,3063                           | 0,5799                           |
| Debtor                    | 2,717308                         | 26,9031                          | <0,00                            |
| Tuition fees up to date   | 0,066019                         | 120,3256                         | <0,00                            |

|                                     |          |          |        |
|-------------------------------------|----------|----------|--------|
| Gender                              | 1,454908 | 11,3534  | <0,00  |
| Scholarship holder                  | 0,296767 | 73,5000  | <0,00  |
| Age at enrollment                   | 0,945085 | 35,9585  | <0,00  |
| Curricular units 1st sem (approved) | 0,863284 | 12,0492  | <0,00  |
| Curricular units 2nd sem (approved) | 2,124822 | 271,8270 | <0,00  |
| Unemployment rate                   | 1,000337 | 0,2328   | 0,6294 |

Tabulka 2 – Výsledky logistické regrese

Pro zhodnocení kvality prediktivního modelu byly použity dva koeficienty determinace: **Cox-Snell  $R^2$**  a **Nagelkerke  $R^2$** . Hodnota Cox-Snellova  $R^2 = 0,50$  ukazuje, že model vysvětluje přibližně 50 % variability ve výstupní proměnné. Hodnota Nagelkerkeho  $R^2 = 0,68$  pak představuje upravenou verzi tohoto koeficientu, která zohledňuje maximální možnou míru vysvětlené variability. Výsledek tedy naznačuje, že model vysvětluje zhruba 68 % variability.

| Cox-Snell $R^2$ | Nagelkerke $R^2$ |
|-----------------|------------------|
| 0,500992        | 0,678999         |

Tabulka 3 – Koeficienty determinace

Výsledky **klasifikace případů** dále potvrzují dobrou predikční schopnost modelu. Model správně identifikoval:

- **93,4 %** případů, kdy student úspěšně absolvoval studium (*Graduate*),
- **81,1 %** případů, kdy student studium nedokončil (*Dropout*).

Celkově lze říci, že model má vysokou míru přesnosti predikce a zároveň dobrou schopnost rozlišovat mezi oběma kategoriemi cílové proměnné. Hodnota **Odds ratio = 60,79** a **log odds ratio = 4,11** ukazují na výrazný rozdíl mezi skupinami v pravděpodobnosti úspěšného absolvování, což potvrzuje, že prediktory použité v modelu poskytují relevantní informace pro odhad výsledku studia.

| Skutečná hodnota | Předpovězeno:<br>Absolvent | Předpovězeno:<br>Odstoupení | Procento<br>správných |
|------------------|----------------------------|-----------------------------|-----------------------|
| Absolvent        | 2063                       | 146                         | 93,39 %               |
| Odstoupení       | 268                        | 1153                        | 81,14 %               |

*Tabulka 4 – Klasifikace případů*

## **Závěr**

Cílem této analýzy bylo identifikovat faktory, které mohou předpovídat úspěšné dokončení vysokoškolského studia. Po důkladném očištění dat a odstranění případů studentů s neukončeným studiem (enrolled) byl vytvořen model logistické regrese, který ukázal velmi dobrou prediktivní schopnost.

Model dokázal s vysokou přesností klasifikovat studenty do kategorií „absolvoval“ a „nedokončil“, přičemž vysvětlil značnou část variability v závislé proměnné (Nagelkerke  $R^2 = 0,68$ ). Mezi nejvýznamnější prediktory patřily akademické výsledky, finanční stabilita, věk a pohlaví.

## Literatura

- “National Commission for Academic Accreditation & Assessment Standards for Quality Assurance and Accreditation of Higher Education Institutions,” 2015.
- Choi, N. (2005). Self-efficacy and self-concept as predictors of college students’ academic performance. *Psychology in the Schools*, 42(2), 197–205. <https://doi.org/10.1002/pits.20048>
- Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). What matters to student success: A review of the literature commissioned report for the National Symposium on postsecondary student success: Spearheading a dialog on student success. [https://nces.ed.gov/npec/pdf/kuh\\_team\\_report.pdf](https://nces.ed.gov/npec/pdf/kuh_team_report.pdf)
- Parker, J. D. A., Summerfeldt, L. J., Hogan, M. J., & Majeski, S. A. (2004). Emotional intelligence and academic success: Examining the transition from high school to university. *Personality and individual differences*, 36(1), 163–172. [https://doi.org/10.1016/S0191-8869\(03\)00076-X](https://doi.org/10.1016/S0191-8869(03)00076-X)
- Realinho, V., Vieira Martins, M., Machado, J., & Baptista, L. (2021). Predict Students' Dropout and Academic Success [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5MC89>.
- York, T. T., Gibson, C., & Rankin, S. (2015). Defining and Measuring Academic Success. *Practical Assessment, Research & Evaluation*, 20, 5. <https://doi.org/10.7275/hz5x-tx03>