

Výkon studentů při zkouškách¹

Úvod

Studijní výkon studentů je ovlivňován řadou faktorů, které zahrnují nejen jejich studijní návyky, ale také osobní motivaci, životní styl a dostupnost studijních zdrojů. V současném výzkumu je proto stále větší pozornost věnována analýze faktorů, které mohou ovlivňovat úspěšnost studentů. Porozumění těmto faktorům může přispět k lepšímu pochopení rozdílů ve studijním výkonu a umožnit identifikaci typických studijních profilů.

Studijní výkon studentů je ovlivňován řadou individuálních, sociálních a institucionálních faktorů. Mezi klíčové individuální determinanty patří zejména motivace, studijní návyky, čas věnovaný studiu a předchozí výsledky, které významně predikují akademickou úspěšnost (Suleiman et al., 2024; Tamang & Khanal, 2023). Významnou roli hrají také psychologické faktory, jako je sebedůvěra, stres či testová úzkost, které mohou výkon při zkouškách ovlivňovat pozitivně i negativně (Wang & Chen, 2025).

Vedle individuálních charakteristik ovlivňují studijní výkon i faktory rodinného a školního prostředí, například zapojení rodičů, socioekonomický status, kvalita výuky či dostupnost studijních zdrojů (Wang & Chen, 2025; Wan Abdul Razak et al., 2021). Významné jsou také životní stylové faktory, jako je spánek nebo mimoškolní aktivity, které ovlivňují schopnost efektivní přípravy (Suleiman et al., 2024; Wan Abdul Razak et al., 2021). Studijní výkon tak představuje výsledek komplexní interakce osobnostních, behaviorálních a environmentálních faktorů.

Jedním z přístupů, jak takové rozdíly analyzovat, je využití vícerozměrných statistických metod, které umožňují zkoumat vztahy mezi více proměnnými současně. Mezi tyto metody patří také shluková analýza, jejímž cílem je identifikovat přirozeně se vyskytující skupiny objektů na základě jejich podobnosti v analyzovaných charakteristikách.

Cílem této práce je identifikovat typologii studentů na základě vybraných faktorů souvisejících se studijním chováním a podmínkami studia a následně analyzovat, jak se tyto skupiny studentů liší ve svém výkonu u zkoušky. Výsledky této analýzy mohou přispět k lepšímu pochopení struktury analyzovaného souboru studentů a umožnit identifikaci typických studijních profilů, které mohou být spojeny s rozdílnou úspěšností při zkouškách.

Metodologie

Pro identifikaci jednotlivých skupin v analyzovaném datovém souboru byla použita shluková analýza (cluster analysis), která představuje soubor statistických metod, jejichž cílem je rozdělit respondenty do skupin tzv. shluků, na základě podobnosti ve vybraných proměnných. Základním principem této metody je, že případy uvnitř jednoho shluku jsou si navzájem co nejvíce podobné a od případů v jiných shlucích se maximálně liší. Shluková analýza je explorativní technikou, jelikož nevyžaduje předem definované skupiny a struktura dat je odhalována přímo na základě analyzovaných proměnných.

¹ Data a další informace o této zprávě jsou dostupné na adrese [https://dostal.vyzkum-
psychologie.cz/stat4?i=711](https://dostal.vyzkum-psychologie.cz/stat4?i=711)

Postup: **hierarchická shluková analýza** → **určení počtu shluků** → **metoda K-means** → **výsledné shluky**

Výše je uvedeno znázornění postupu, který dále blíže popíši. V této práci byla shluková analýza provedena ve dvou navazujících krocích. Nejprve jsem aplikovala **hierarchickou shlukovou analýzu**, která slouží k exploraci struktury dat a k určení vhodného počtu shluků. Hierarchické metody pracují tak, že každé pozorování je na počátku považováno za samostatný shluk, dále jsou postupně slučovány ty shluky, které jsou si navzájem nejpodobnější, a to na základě zvolené míry vzdálenosti. Tento proces pokračuje až do okamžiku, kdy jsou všechna pozorování spojena do jednoho shluku. Výsledkem hierarchické shlukové analýzy je **dendrogram**, tedy stromový graf, který znázorňuje postupné slučování jednotlivých pozorování a umožňuje posoudit vhodný počet výsledných shluků.

Jako míra podobnosti mezi jednotlivými pozorováními byla použita **Eukleidovská vzdálenost**, která je jednou z nejčastěji používaných. Pro samotné slučování shluků byla využita **Wardova metoda**, jejímž cílem je minimalizovat nárůst vnitroshlukové variability při každém dalším spojování shluků. Tento přístup vede k vytváření relativně homogenních skupin s nízkou vnitřní variabilitou.

Na základě výsledků hierarchické shlukové analýzy byl následně stanoven optimální počet shluků. V druhé fázi analýzy byla proto použita **metoda K-means**, která umožňuje jednoznačné přiřazení jednotlivých pozorování k předem stanovenému počtu shluků. Metoda K-means pracuje tak, že nejprve jsou v prostoru dat určeny počáteční centroidy (středy shluků), ke kterým jsou jednotlivá pozorování přiřazena na základě minimální vzdálenosti. Následně jsou centroidy přepočítány jako průměr všech pozorování přiřazených k danému shluku a proces se opakuje, dokud nedojde ke stabilizaci přiřazení jednotlivých pozorování.

Kombinace hierarchické shlukové analýzy a metody K-means představuje běžný postup při vícerozměrné explorativní analýze dat. Výsledkem tohoto postupu je klasifikace pozorování do několika homogenních skupin, které se mezi sebou liší na základě analyzovaných proměnných. Tyto skupiny lze následně interpretovat a porovnávat z hlediska jejich charakteristik, což umožňuje lépe porozumět struktuře analyzovaného souboru.

Data a proměnné

Analýza vychází z datového souboru obsahujícího informace o Výkonu studentů při zkouškách. Dataset byl získán z www.kaggle.com a původně obsahoval celkem **6607** záznamů (pozorování). Pro účely této analýzy byl použit soubor o velikosti **1000** pozorování, který byl vybrán systematicky. Do výzkumu jsem zařadila každého šestého respondenta, jakmile jsem došla k počtu 1000, tak jsem s výběrem skončila.

Původní datový soubor zahrnuje **20** proměnných popisujících akademické, životní a socioekonomické faktory týkající se studentů. Hlavní cílovou proměnnou je konečné skóre studenta ze zkoušky. Pro účely této zprávy jsem vybrala **7** proměnných, které byly zvoleny na základě jejich relevance k analyzovanému tématu a předpokladu, že mohou přispět k identifikaci smysluplných skupin v rámci sledovaného souboru. Zároveň i dle teoretického úvodu byly tyto proměnné často zmiňovány, a proto jsem je vyhodnotila jako relevantní. Byly to především proměnné popisující studijní styl, podmínky a životní režim studenta. Dále později pro interpretaci shluků jsem použila proměnnou **Výsledek zkoušky**.

Pro samotnou shlukovou analýzu byly jako vstupní proměnné zvoleny následující:

- **Počet hodin studia** – typ proměnné: spojitá, popis: hodiny, které student věnoval studiu, zaznamenaná v procentech,
- **Účast** – typ proměnné: spojitá, popis: míra účasti studenta na výuce, zaznamenaná v procentech,
- **Počet hodin spánku** – typ proměnné: spojitá, popis: průměrný počet hodin spánku studenta,
- **Úroveň motivace** – typ proměnné: ordinální, popis: úroveň studijní motivace studenta, zaznamenaná v kategoriích nízká, střední a vysoká,
- **Přístup ke zdrojům** – typ proměnné: ordinální, popis: přístup k studijním materiálům, zaznamenaná v kategoriích nízká, střední a vysoká,
- **Zapojení rodičů** – typ proměnné: ordinální, popis: míra zapojení rodičů do studijního procesu student, zaznamenaná v kategoriích nízká, střední a vysoká,
- **Výsledek zkoušky** – typ proměnné: spojitá, popis: dosažený výsledek u zkoušky.

Před samotnou analýzou byla data podrobena kontrole chybějících hodnot, identifikaci extrémních hodnot a transformaci proměnných. Ordinální proměnné byly před analýzou převedeny na numerické hodnoty. Byla provedena standardizace, aby jednotlivé proměnné měly srovnatelný vliv na výpočet vzdáleností mezi pozorováními, konkrétně šlo o z-skór standardizaci.

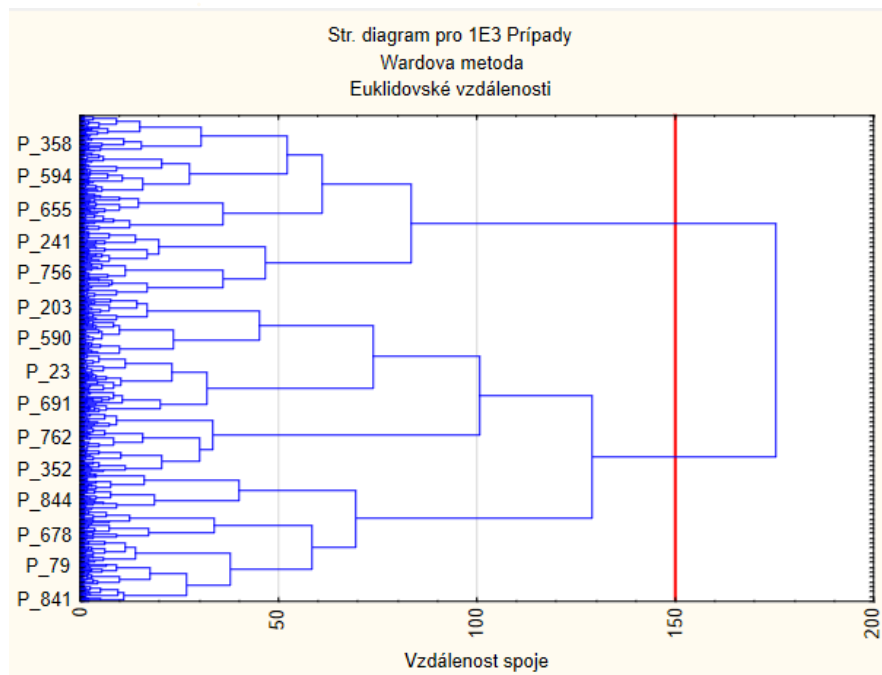
Zbylé proměnné, které nebyly použity v analýze, byly: **Mimoškolní aktivity, Dosavadní výsledky, Přístup k internetu, Doučování, Příjem rodiny, Kvalita učitele, Typ školy, Vliv vrstevníků, Tělesná aktivita, Poruchy učení, Úroveň vzdělání rodičů, Vzdálenost od domova, Pohlaví.**

Výsledky a interpretace

Hierarchická shluková analýza

V první fázi analýzy byla provedena hierarchická shluková analýza s cílem prozkoumat strukturu dat a určit vhodný počet výsledných shluků. Výsledkem této analýzy byl dendrogram znázorňující postupné slučování jednotlivých pozorování do větších skupin. Na základě vizuálního posouzení dendrogramu (viz Obrázek 1) a analýzy vzdáleností mezi jednotlivými kroky slučování bylo možné identifikovat několik potenciálních řešení. Nejvhodnější se jevílo rozdělení dat do 2 shluků, které představovalo kompromis mezi interpretovatelností výsledků a dostatečným rozlišením jednotlivých skupin. Zároveň zde je viditelný velký skok ve vzdálenosti. Toto řešení bylo následně využito v další fázi analýzy pomocí metody K-means.

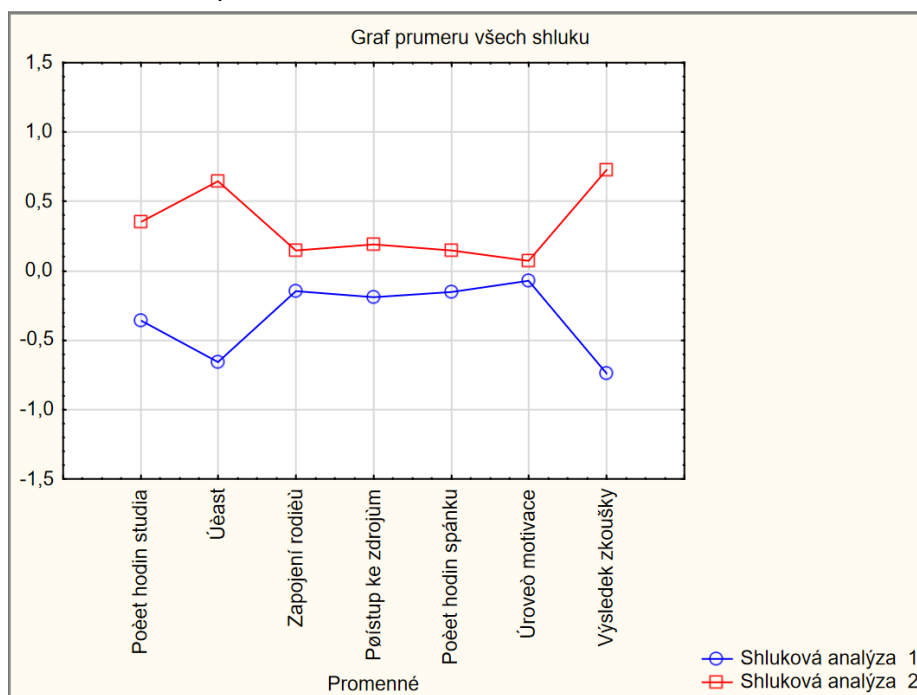
Obrázek 1: Dendrogram (Wardova metoda)



Metoda K-means

Na základě výsledků hierarchické shlukové analýzy byla aplikována metoda K-means, přičemž počet shluků byl stanoven na hodnotu $K = 2$. Tato metoda umožnila jednoznačné přiřazení jednotlivých pozorování do konkrétních shluků na základě minimální vzdálenosti k centroidu daného shluku. Po provedení klasifikace bylo do jednotlivých shluků rozděleno celkem 1000 pozorování. Na obrázku 2 jsou graficky znázorněny průměry všech shluků dle jednotlivých proměnných, které jsou uvedeny pod grafem. Na tomto grafu lze názorně sledovat, ve kterých oblastech se výrazně shluky liší, a naopak ve kterých jsou na podobné úrovni i přes příslušenství k rozdílným shlukům. Rozdělení případů mezi jednotlivé klastry je detailněji uvedeno v tabulce 1.

Obrázek 2: Graf průměrů všech shluků



Tabulka 1: Souhrn výsledků klasifikace (K-Means)

Klastr	Počet případů	Podíl [%]	Počet hodin studia	Účast	Zapojení rodičů	Přístup ke zdrojům	Počet hodin spánku	Úroveň motivace	Výsledek zkoušky
1	496	49,6 %	-0,36	-0,66	-0,15	-0,19	-0,15	-0,07	-0,74
2	504	50,4 %	0,35	0,65	0,14	0,19	0,14	0,07	0,73

V tabulce 1 je uvedeno příslušenství jednotlivých pozorování k jednotlivým klastrům, je zde uveden i procentuální podíl a souhrn výsledků po metodě K-Means.

Na základě provedené shlukové analýzy byly identifikovány 2 skupiny studentů, které se mezi sebou liší především ve studijních návycích, míře motivace a životním stylu. **Klastr 2** lze charakterizovat jako skupinu studentů s vyšší studijní aktivitou. Tito studenti vykazují vyšší hodnoty proměnných Počet hodin studia, Účast a zároveň dosahují lepších výsledků u zkoušky. Lze předpokládat, že se jedná o studenty, kteří věnují studiu dostatečné množství času a vykazují vyšší míru disciplinovanosti. **Klastr 1** zahrnuje studenty s jinými charakteristikami, zejména nižšími hodnotami proměnných Počtu hodin studia a Účasti ve výuce. Tato skupina může představuje studenty s nižší motivací nebo méně efektivními studijními návyky.

Závěr

Cílem této práce bylo identifikovat typologii studentů na základě vybraných charakteristik souvisejících se studijním chováním a podmínkami studia a následně porovnat tyto skupiny z hlediska jejich výkonu u zkoušky.

K dosažení tohoto cíle byla využita shluková analýza. Na základě výsledného dendrogramu zvoleno rozdělení dat do dvou shluků, a to studenti s lepšími výsledky a studenti s horšími výsledky. Následně byla použita metoda K-means, která umožnila jednoznačné přiřazení jednotlivých studentů k výsledným skupinám, které se od sebe liší zejména v proměnných **Počet hodin studia, Účast**, a tedy ve **Výsledcích zkoušek**. Naopak v proměnných **Zapojení rodičů, Přístup ke zdrojům, Počet hodin spánku a Úroveň motivace** se výsledky nelišily tolik a jejich hodnoty byly na podobných úrovních. Jednotlivé klastry zahrnovaly 496 a 504 studentů a vykazovaly rozdílné průměrné hodnoty sledovaných charakteristik.

Z hlediska studijního výkonu se jednotlivé skupiny studentů lišily také ve **Výsledku zkoušky**. Nejvyšší průměrné hodnoty proměnné Výsledek zkoušky **0,73** byly zaznamenány v **klastru 2**, zatímco nejnižší hodnoty **-0,74** vykazoval **klastr 1**. Tyto rozdíly naznačují, že identifikované skupiny studentů se mohou lišit nejen ve studijních návycích a podmínkách studia, ale také v dosaženém akademickém výkonu. Bylo zjištěno, že studenti větším časem věnovaným studiu dosahují lepších výsledků.

Výsledky této analýzy přispívají k lepšímu porozumění struktury analyzovaného souboru a ukazují, že studenti mohou vytvářet odlišné typologické skupiny na základě jejich studijního chování a dalších faktorů. Tyto poznatky mohou být dále využity při interpretaci faktorů ovlivňujících studijní úspěšnost nebo při navrhování podpůrných opatření zaměřených na specifické skupiny studentů.

V budoucím výzkumu by bylo vhodné analýzu rozšířit o další proměnné, které jsem do výzkumu nezahrnula, ale mohly by přispět k hlubšímu porozumění rozdílům mezi jednotlivými skupinami studentů.

Použité zdroje:

Shoaib, M. (2025). *Student Exam Performance Dataset Analysis*. Kaggle.

<https://www.kaggle.com/datasets/grandmaster07/student-exam-performance-dataset-analysis/data>

Suleiman, I.B., Okunade, O.A., Dada, E.G. et al. (2024). Key factors influencing students' academic performance. *Journal of Electrical Systems and Inf Technol* 11, 41.

<https://doi.org/10.1186/s43067-024-00166-w>

Tamang, S., & Khanal, S. P. (2023). Factors Affecting Academic Performance of M. Sc. Level Students at Tribhuvan University. *Nepalese Journal of Statistics*, 7(1), 19–32.

<https://doi.org/10.3126/njs.v7i1.61055>

Wan Abdul Razak, W. A. I., Khairul Akmal, S. N. A., Azizan, N. F., Wani, S., Hamal, A. Q., & Muhammad, A. H. (2021). Factors Affecting Student's Academic Performance. *International Journal on Perceptive and Cognitive Computing*, 7(1), 99–107. Získáno z

<https://journals.iium.edu.my/kict/index.php/IJPC/article/view/228>

Wang, L. & Chen, C.J. (2025). Factors affecting student academic performance: A systematic review. *International Journal on Studies in Education (IJonSE)*, 7(1), 1-

47. <https://doi.org/10.46328/ijonse.276>