

## Výsledná známka studentů v závislosti na vybraných faktorech

Úspěšnost studentů je kombinací socio-ekonomických, psychologických, rodinných faktorů a životního prostředí (Hijazi & Naqvi, 2006). Jako další hlavní faktory se ukazují dosavadní známky studentů a výkon během výuky, aktivita a zapojení studentů. Student, který je zvyklý mít dobré známky od zahájení studia po tomto aspiruje zpravidla po zbytek akademického života (Abu Saa et al., 2019).

Ve veřejně dostupné databázi datasetů byl zveřejněn výsledek dotazníku, kde autor, Joakim Arvidsson, zveřejnil údaje z dotazníku o studentech a jejich finální výsledek zkoušky. Jako původ těchto dat uvádí odborný zdroj<sup>1</sup>.

V dotazníku bylo zjišťováno spoustu údajů o daných studentech, od rodinného zázemí, přes pracovní a volnočasový život, až po přístup ke studiu a výsledné známky. V naší zprávě se pokusíme respondenty rozdělit do skupin podle toho:

- kolik hodin týdně se učí,
- jak často chodí na výuku,
- jak se na zkoušku připravovali (o samotě, s přáteli, nepřipravovali, průběžně během semestru, na poslední chvíli),
- zda si při výuce dělají poznámky,
- zda při výuce poslouchají výklad,
- a jakou mají výslednou známku.

Pro rozdělení respondentů byla využita metoda shlukové analýzy. Tento termín značí metodu klasifikace respondentů do skupin podle podobností ve zvolených proměnných. Výhodou této metody je, že není nutné mít předem představu, kolik skupin bude, stačí znát proměnné, podle kterých chceme shlukování provést. V rámci metody je možné použít několik přístupů, nejčastější je hierarchická a průměrová, tzv. k-means (STEM/MARK, 2020).

---

<sup>1</sup> Yılmaz N., Sekeroglu B. (2020) Student Performance Classification Using Artificial Intelligence Techniques. In: Aliev R., Kacprzyk J., Pedrycz W., Jamshidi M., Babanlı M., Sadikoglu F. (eds) 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019. ICSCCW 2019. Advances in Intelligent Systems and Computing, vol 1095. Springer, Cham

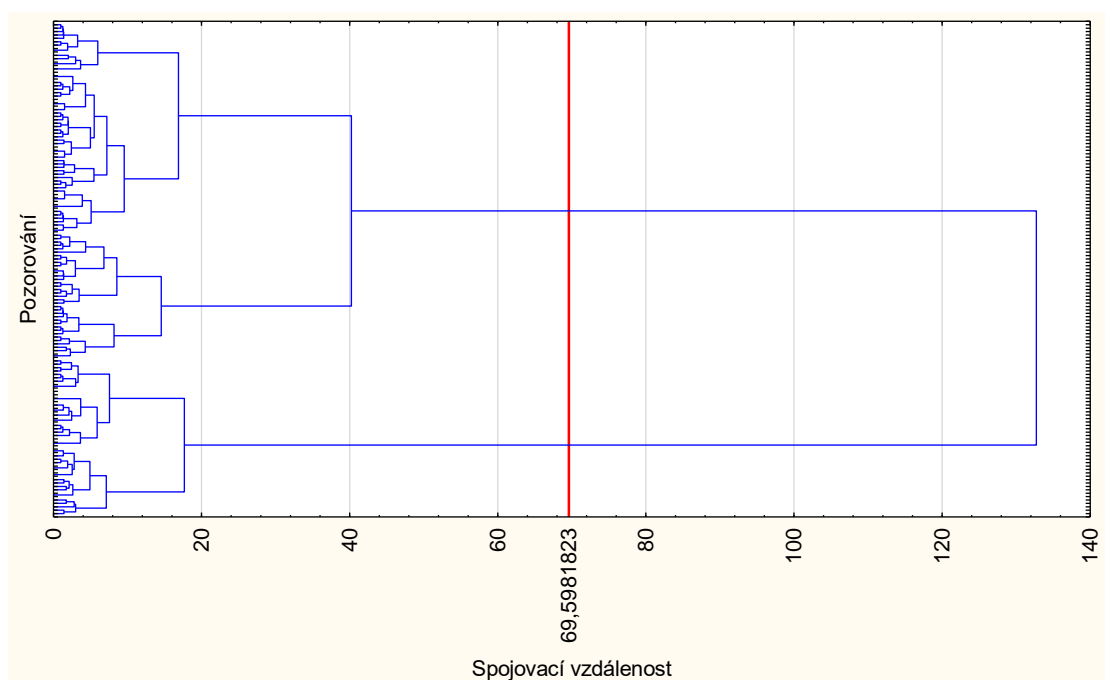
V prvním kroku naší analýzy jsme použili hierarchickou metodu, jejíž výsledkem je dendrogram, který ukazuje hierarchii shluků. Následně byla využita k-means shluková analýza. Ta je založena na principu nalezení takových skupin v mnoharozměrném prostoru, kdy vnitroskupinová podobnost je co největší, a to pomocí minimalizace celkové sumy čtverců vzdáleností uvnitř skupin. Výsledkem je vytvoření skupin, které jsou od sebe co nejvíce odděleny. Tato metoda však nenabízí odpověď na to, do kolika shluků by bylo vhodné data roztrždit, k tomu slouží právě výše zmíněná metoda (Institut biostatistiky a analýz Lékařské fakulty Masarykovy univerzity).

### Hierarchická shluková analýza

Jelikož každá číselná hodnota reprezentuje jiný význam a ne konkrétní bodovou hodnotu, a také jsou proměnné ve stejných jednotkách a vykazují shodnou variabilitu, nebyla data před zahájením analýzy standardizována (Meloun & Militký, 2004, citováno v Poláčková, 2013). Data, na nichž byla analýza provedena, obsahovala údaje od 145 respondentů.

Jako kritérium pro výběr dvou nejbližších shluků byla na základě obecných doporučení zvolena Wardova metoda, která minimalizuje součet kvadratických odchylek a vytváří kompaktní, poměrně malé množiny zhruba stejné velikosti (Institut biostatistiky a analýz Lékařské fakulty Masarykovy univerzity).

Obrázek 1 Dendrogram hierarchické shlukové analýzy



Na obrázku 1 můžeme na ose Y vidět jednotlivá pozorování jako samotné shluky, které jsou postupně slučovány do větších skupin, a na ose X vidíme spojovací vzdálenost mezi nimi. Červená vertikální čára ukazuje doporučené rozdělení respondentů do dvou skupin, tohoto doporučení se budeme pro další analýzu držet.

#### Metoda k-mean (nehierarchická)

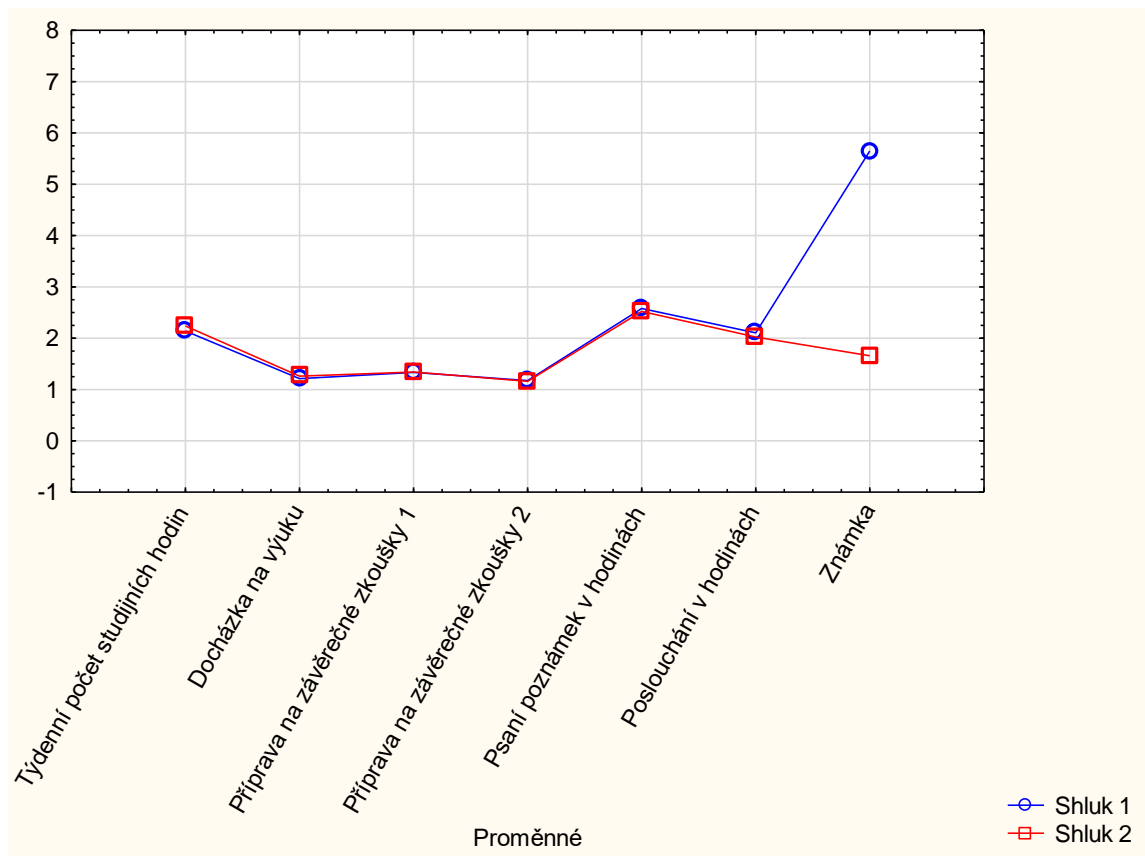
Při hierarchické metodě shlukové analýzy byl jako vhodný počet shluků zvolen počet dva, toto zachováme i při metodě k-means, a vzniklé shluky budeme označovat jako Shluk 1 (n=57) a Shluk 2 (n=88).

V tabulce 1 můžeme vidět výsledky analýzy rozptylu shluků. Z tabulky můžeme vidět, že odpovědi ve skupinách se signifikantně vůbec neliší, kromě položky výsledná známka. Stejně tak na obrázku 2 v grafu průměrných odpovědí respondentů všech shluků vidíme, že se liší pouze výslednou známkou.

Tabulka 1 Analýza rozptylu shluků

Proměnná	Between SS	df	Within	df	F	signif. P
Týdenní počet studijních hodin	0,334	1	120,866	143	0,395	0,530
Docházka na výuku	0,089	1	26,462	143	0,483	0,488
Příprava na závěrečné zkoušky 1	0,002	1	54,439	143	0,005	0,943
Příprava na závěrečné zkoušky 2	0,009	1	24,018	143	0,055	0,815
Psaní poznámek v hodinách	0,109	1	45,849	143	0,341	0,560
Poslouchání v hodinách	0,236	1	65,323	143	0,516	0,474
Známka	550,735	1	144,755	143	544,057	0,000

Obrázek 2 Graf průměrných odpovědí respondentů



## Literatura

Abu Saa, A., Al-Emran, M. & Shaalan, K. (2019). Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques. *Technology, Knowledge and Learning*. Vol. 24, 567–598.

<https://doi.org/10.1007/s10758-019-09408-7>

Hijazi, S. T. & Naqvi, S.M.M.R. (2006). Factors Affecting Students' Performance. A Case Of Private Collages. *Bangladesh e-Journal of Sociology*. 3(1), 1-10

Institut biostatistiky a analýz Lékařské fakulty Masarykovy univerzity (n.d.). *Metoda k-průměru*. Získáno 17.2.2024 z

<https://portal.matematickabiologie.cz/index.php?pg=analiza-a-hodnoceni-biologickych-dat--vicerozmerne-metody-pro-analyzu-dat--shlukova-analyza--shlukova-nehierarchicka-analyza--metoda-k-prumeru>

Poláčková, J. (2013). *Dataminingové techniky analýz vícerozměrných datových souborů* [Disertační práce, Česká zemědělská univerzita v Praze]. Databáze ČZU

<https://www.pef.czu.cz/dl/46020>

STEM/MARK. (2020). *Shluková analýza; Cluster analýza*. Získáno 17.2.2024 z

<https://stemmark.cz/encyklopedie-shlukova-analyza-cluster-analyza/>