

## Stanovení cut-off skóre pro predikci spokojenosti studentů na základě akademických výsledků na VŠ

Spokojenost studentů je dlouhodobě zkoumaným faktorem ve vztahu k jejich akademickým úspěchům. Výzkumy naznačují, že spokojenější studenti mají tendenci dosahovat lepších studijních výsledků a častěji dokončují své studium. Tato souvislost se vysvětluje několika mechanismy – spokojenost studentů ovlivňuje jejich motivaci, zapojení do výuky i celkový vztah ke studijním povinnostem (Oja, 2011; Schreiner, 2009).

Dle studie Weerasinghe & Fernando (2017) je spokojenost studentů výsledkem jejich vzdělávacích zkušeností, kvality výuky, vztahů s vyučujícími a celkového prostředí univerzity. Pokud studenti vnímají svou univerzitu jako podporující a vstřícnou, mají větší motivaci se aktivně účastnit výuky, což se následně odráží na jejich akademických výsledcích.

V této práci je proto naším **cílem** za pomoci **ROC analýzy** stanovit cut-off skóre akademických výsledků, které by mohlo predikovat úroveň spokojenosti studentů se studiem. Cut-off skóre určuje hranici výsledků, která určuje, zda je daná charakteristika přítomná nebo chybí. V našem případě se jedná o **hranici akademických výsledků** neboli *college mark* (=známka), která pomáhá **rozlišit mezi spokojenými a nespokojenými studenty**. Pokud student dosáhne této nebo vyšší hodnoty, můžeme říct, že jeho studijní výsledky pravděpodobně souvisí s větší spokojeností se studiem.

### Popis datasetu

V této práci čerpáme data z datasetu **Student Attitude and Behavior**, který pochází z platformy Kaggle a obsahuje informace o univerzitních studentech z IIIT-Delhi shromážděné prostřednictvím Google formuláře. Celkově dataset zahrnuje **235 respondentů**, z toho 156 mužů a 79 žen. Data obsahují informace o akademických výsledcích studentů, jejich demografických charakteristikách, preferencích studijního prostředí, každodenních studijních návycích a spokojenosti se studiem.

### Pro stanovení cut off skóru jsme zvolili následující proměnné:

- **College mark** (akademické výsledky), dosahuje hodnot od 1 do 100
- **Spokojenost se studiem** (dichotomická proměnná)
  - Kódování 0: Nespokojen
  - Kódování 1: Spokojen

- **Prevalence** byla v tomto případě stanovena na základě odvozených hodnot a dosahuje 91 % (tedy kolik procent studentů uvedlo spokojenost)

Následně jsme vytvořili Tabulku 1, která ukazuje tyto ukazatele:

- **True positive (TP)** označuje počet studentů, kteří jsou spokojeni se studiem (1) a jejich akademické výsledky (college mark) správně predikují spokojenost (vysoké skóre).
- **False positive (FP)** představuje studenty, u nichž akademické výsledky predikují spokojenost (vysoké skóre), ale ve skutečnosti uvedli, že jsou se studiem nespokojeni (0).
- **False negative (FN)** udává počet studentů, kteří jsou spokojeni (1), ale jejich akademické výsledky predikují nespokojenost (nízké skóre).
- **True negative (TN)** vyjadřuje počet studentů, kteří jsou nespokojeni se studiem (0) a jejich akademické výsledky tuto nespokojenost správně predikují (nízké skóre).

Pro nalezení optimálního **cut-off skóre** akademických výsledků (college mark) jsme analyzovali, jak různé hranice tohoto skóre ovlivňují přesnost predikce spokojenosti studentů se studiem. Tabulka 1 tedy dále ukazuje, jak dobře model rozlišuje mezi spokojenými a nespokojenými studenty na základě následujících ukazatelů:

- **Senzitivita** vyjadřuje, jaké procento spokojených studentů (1) model správně identifikoval při daném cut-off skóre.
- **Specificita** naopak ukazuje, kolik procent nespokojených studentů (0) bylo modelem správně klasifikováno jako nespokojené.

K vyhodnocení celkové přesnosti modelu používáme **Youdenův index (J)**, který kombinuje senzitivitu a specificitu. Vyšší hodnota Youdenova indexu znamená lepší schopnost modelu odlišit spokojené a nespokojené studenty.

Youdenův index však předpokládá rovnoměrné zastoupení obou skupin, což v našem datasetu není splněno, protože **91 % studentů je spokojených**, zatímco pouze **9 % nespokojených**. Tento nepoměr by mohl zkreslit hodnocení modelu, proto navíc doplňujeme **statistiku I**, která **zohledňuje nerovnoměrné rozložení dat** a poskytuje přesnější odhad efektivity cut-off skóre. Tímto postupem jsme identifikovali **optimální cut-off skóre**, které

umožňuje co nejpřesnější predikci spokojenosti studentů na základě jejich akademických výsledků

**Tabulka 1:** Ukazatele pro vyhodnocení cut-off skóru

Cut off skóre	TP	TN	FP	FN	Sensitivita	Specificita	J	I
1	215	0	20	0	100 %	0 %	0,00	91 %
2	214	0	20	1	100 %	0 %	0,00	91 %
3	213	0	20	2	99 %	0 %	-0,01	91 %
4	212	0	20	3	99 %	0 %	-0,01	90 %
5	212	0	20	3	99 %	0 %	-0,01	90 %
6	212	0	20	3	99 %	0 %	-0,01	90 %
7	212	0	20	3	99 %	0 %	-0,01	90 %
8	212	1	19	3	99 %	5 %	0,04	91 %
9	212	1	19	3	99 %	5 %	0,04	91 %
10	212	1	19	3	99 %	5 %	0,04	91 %
11	212	1	19	3	99 %	5 %	0,04	91 %
12	212	1	19	3	99 %	5 %	0,04	91 %
13	212	2	18	3	99 %	10 %	0,09	91 %
14	212	2	18	3	99 %	10 %	0,09	91 %
15	212	2	18	3	99 %	10 %	0,09	91 %
16	212	2	18	3	99 %	10 %	0,09	91 %
17	212	2	18	3	99 %	10 %	0,09	91 %
18	212	2	18	3	99 %	10 %	0,09	91 %
19	212	2	18	3	99 %	10 %	0,09	91 %
20	212	2	18	3	99 %	10 %	0,09	91 %
21	212	2	18	3	99 %	10 %	0,09	91 %
22	212	2	18	3	99 %	10 %	0,09	91 %
23	212	2	18	3	99 %	10 %	0,09	91 %
24	212	2	18	3	99 %	10 %	0,09	91 %
25	212	2	18	3	99 %	10 %	0,09	91 %
26	212	2	18	3	99 %	10 %	0,09	91 %
27	212	2	18	3	99 %	10 %	0,09	91 %
28	212	2	18	3	99 %	10 %	0,09	91 %
29	212	2	18	3	99 %	10 %	0,09	91 %
30	212	2	18	3	99 %	10 %	0,09	91 %
31	210	2	18	5	98 %	10 %	0,08	90 %
32	210	2	18	5	98 %	10 %	0,08	90 %
33	210	2	18	5	98 %	10 %	0,08	90 %
34	210	2	18	5	98 %	10 %	0,08	90 %
35	210	2	18	5	98 %	10 %	0,08	90 %
36	209	2	18	6	97 %	10 %	0,07	90 %
37	209	2	18	6	97 %	10 %	0,07	90 %
38	209	2	18	6	97 %	10 %	0,07	90 %
39	209	2	18	6	97 %	10 %	0,07	90 %
40	209	2	18	6	97 %	10 %	0,07	90 %
41	209	2	18	6	97 %	10 %	0,07	90 %
42	209	2	18	6	97 %	10 %	0,07	90 %
43	209	2	18	6	97 %	10 %	0,07	90 %
44	209	2	18	6	97 %	10 %	0,07	90 %
45	209	2	18	6	97 %	10 %	0,07	90 %
46	209	2	18	6	97 %	10 %	0,07	90 %
47	209	2	18	6	97 %	10 %	0,07	90 %
48	209	2	18	6	97 %	10 %	0,07	90 %
49	209	2	18	6	97 %	10 %	0,07	90 %
50	209	2	18	6	97 %	10 %	0,07	90 %

**Tabulka 1: pokračování**

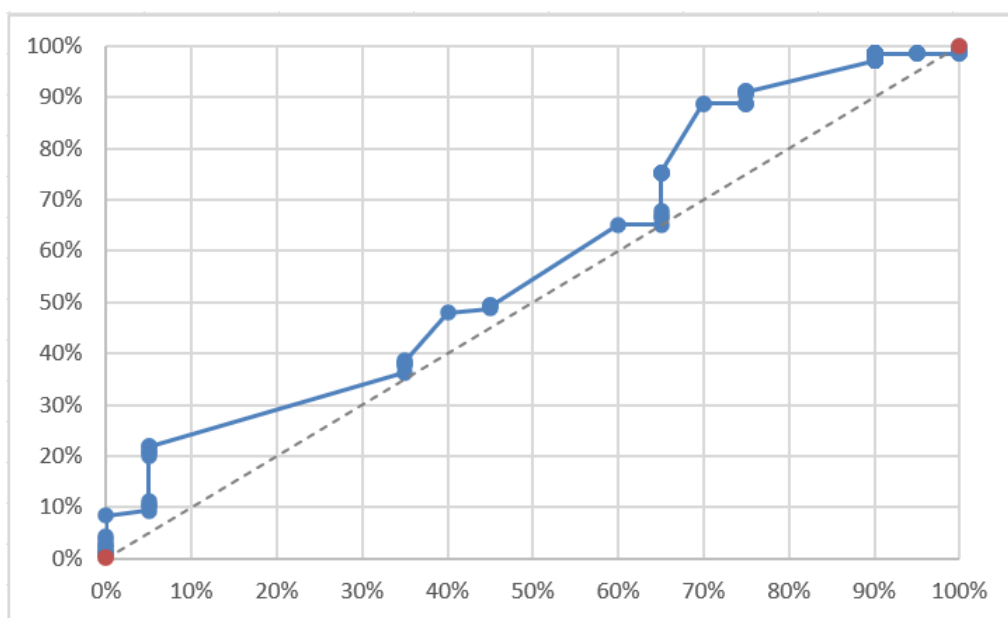
Cut off skóre	TP	TN	FP	FN	Sensitivita	Specificita	J	I
51	196	5	15	19	91 %	25 %	0,16	86 %
52	196	5	15	19	91 %	25 %	0,16	86 %
53	196	5	15	19	91 %	25 %	0,16	86 %
54	195	5	15	20	91 %	25 %	0,16	85 %
55	195	5	15	20	91 %	25 %	0,16	85 %
56	191	5	15	24	89 %	25 %	0,14	83 %
57	191	5	15	24	89 %	25 %	0,14	83 %
58	191	5	15	24	89 %	25 %	0,14	83 %
<b>59</b>	<b>191</b>	<b>6</b>	<b>14</b>	<b>24</b>	<b>89 %</b>	<b>30 %</b>	<b>0,19</b>	<b>84 %</b>
<b>60</b>	<b>191</b>	<b>6</b>	<b>14</b>	<b>24</b>	<b>89 %</b>	<b>30 %</b>	<b>0,19</b>	<b>84 %</b>
61	162	7	13	53	75 %	35 %	0,10	72 %
62	162	7	13	53	75 %	35 %	0,10	72 %
63	162	7	13	53	75 %	35 %	0,10	72 %
64	162	7	13	53	75 %	35 %	0,10	72 %
65	162	7	13	53	75 %	35 %	0,10	72 %
66	146	7	13	69	68 %	35 %	0,03	65 %
67	144	7	13	71	67 %	35 %	0,02	64 %
68	143	7	13	72	67 %	35 %	0,02	64 %
69	140	7	13	75	65 %	35 %	0,00	63 %
70	140	8	12	75	65 %	40 %	0,05	63 %
71	106	11	9	109	49 %	55 %	0,04	50 %
72	106	11	9	109	49 %	55 %	0,04	50 %
73	106	11	9	109	49 %	55 %	0,04	50 %
74	105	11	9	110	49 %	55 %	0,04	49 %
75	103	12	8	112	48 %	60 %	0,08	49 %
76	83	13	7	132	39 %	65 %	0,04	41 %
77	82	13	7	133	38 %	65 %	0,03	40 %
78	82	13	7	133	38 %	65 %	0,03	40 %
79	81	13	7	134	38 %	65 %	0,03	40 %
80	78	13	7	137	36 %	65 %	0,01	39 %
81	47	19	1	168	22 %	95 %	0,17	28 %
82	47	19	1	168	22 %	95 %	0,17	28 %
83	45	19	1	170	21 %	95 %	0,16	27 %
84	45	19	1	170	21 %	95 %	0,16	27 %
85	43	19	1	172	20 %	95 %	0,15	26 %
86	24	19	1	191	11 %	95 %	0,06	18 %
87	22	19	1	193	10 %	95 %	0,05	17 %
88	22	19	1	193	10 %	95 %	0,05	17 %
89	20	19	1	195	9 %	95 %	0,04	17 %
90	18	20	0	197	8 %	100 %	0,08	16 %
91	9	20	0	206	4 %	100 %	0,04	12 %
92	8	20	0	207	4 %	100 %	0,04	12 %
93	6	20	0	209	3 %	100 %	0,03	11 %
94	5	20	0	210	2 %	100 %	0,02	11 %
95	5	20	0	210	2 %	100 %	0,02	11 %
96	3	20	0	212	1 %	100 %	0,01	10 %
97	3	20	0	212	1 %	100 %	0,01	10 %
98	2	20	0	213	1 %	100 %	0,01	9 %
99	2	20	0	213	1 %	100 %	0,01	9 %
100	2	20	0	213	1 %	100 %	0,01	9 %

Na základě **Youdenova indexu (J)** a **statistiky I** lze **cut-off skóre** stanovit na **59 anebo 60 bodů**, protože obě hodnoty dosahují stejných výsledků a volba mezi nimi by neměla mít významný dopad na výsledky.

Studenti s výsledkem **59 a více** mají vyšší pravděpodobnost být spokojeni se studiem, zatímco nižší skóre naznačuje vyšší pravděpodobnost nespokojenosti. **Senzitivita dosahuje 89 % a specificita 30 %**, což naznačuje vysokou schopnost správně identifikovat spokojené studenty, ale omezenou schopnost správně určit ty nespokojené.

Efektivitu škály ukazuje **ROC křivka**, která srovnává naši metodu s náhodným rozdělením, znázorněným tečkovanou čarou na Grafu 1.

**Graf 1:** ROC křivka



ROC křivka ukazuje, jak dobře test rozlišuje mezi spokojenými a nespokojenými studenty. Čím více se křivka přibližuje levému hornímu rohu, tím je test přesnější. V našem případě křivka není výrazně vychýlená od diagonály, což naznačuje **omezenou efektivitu metody**.

Plocha pod křivkou **AUC (area under the curve)** dosahuje **58,07 %**, což znamená, že test mírně překonává náhodné rozřazení (50 %), ale jeho schopnost přesně predikovat spokojenost studentů není vysoká. **Diagonální čára v grafu představuje hranici náhodného rozhodování**, kde by test nebyl efektivní. AUC blížící se 100 % by značila vysokou přesnost predikce, což v našem případě bohužel neplatí.

## Závěr

Na základě **ROC analýzy** jsme určili, že **hranice akademických výsledků pro predikci spokojenosti studentů je 59 anebo 60 bodů**. Studenti s tímto skóre a vyšším mají větší pravděpodobnost být spokojeni se studiem. Ukazatel **AUC (58,07 %)** indikuje, že metoda je o něco lepší než náhodné rozřazení, ale její schopnost přesné predikce je omezená. Proto je vhodné výsledky brát s rezervou a zvážit i další faktory ovlivňující spokojenost studentů.

## Použitá literatura

- Oja, M. (2011). Student Satisfaction and Student Performance. *Journal of Applied Research in the Community College*, 19(1), 47–53.
- Schreiner, L. A. (2009). Linking student satisfaction and retention. *Noel-Levitz, Coralville, IA*.
- Weerasinghe, I. M. S., & Fernando, R. L. (2017). *Students' Satisfaction in Higher Education* (SSRN Scholarly Paper 2976013). Social Science Research Network. <https://papers.ssrn.com/abstract=2976013>