

ИСПОЛЬЗОВАНИЕ ROC-АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ ОДАРЕННЫХ СТУДЕНТОВ ПО РЕЗУЛЬТАТАМ БАЛЛОВ АБИТУРИЕНТОВ ПРИ ПОСТУПЛЕНИИ

*Е.Н. Залесская, А.А. Чиркина, Е.А. Капорицова
Витебск, ВГУ имени П.М. Машерова*

В Республике Беларусь значительное внимание уделяется одаренным и талантливым студентам. При анализе контингента поступивших абитуриентов деканату доступна информация о баллах студентов при поступлении [1]. В дальнейшем под баллом студента при поступлении будем понимать сумму среднего балла аттестата, умноженного на 10, результатов ЦЭ и ЦТ абитуриентов. Эти данные являются показателем уровня подготовленности абитуриентов и могут выполнять диагностическую функцию для выявления одаренных студентов, работу с которыми необходимо начинать как можно раньше.

В связи с этим, актуальной задачей является использование информации о баллах при поступлении для определения порогового значения баллов, которое в дальнейшем можно использовать для выделения группы одаренных студентов. Таким образом, целью исследования является поиск взаимосвязи между количеством баллов при поступлении и успешностью обучения в университете для выявления способных студентов.

Материал и методы. В исследовании была использована информация о баллах студентов при поступлении, а также о средних баллах по итогам первого курса студентов факультета математики и информационных технологий ВГУ имени П.М. Машерова специальностей «Управление информационными ресурсами», «Информационные системы и технологии», «Прикладная информатика», «Программное обеспечение информационных технологий», «Программная инженерия», «Прикладная математика», «Компьютерная безопасность», «Кибербезопасность», «Физика», «Математика и информатика», «Физико-математическое образование» 2021–2023 годов набора (416 студентов). Для определения оптимального порога классификации по выделению группы одаренных студентов использовался ROC-анализ.

Результаты и их обсуждение. ROC-анализ находит широкое применение не только в медико-биологических исследованиях, но и в анализе образовательных данных, в том числе для диагностики способностей обучаемых по имеющейся информации о начальном уровне подготовки и результатам обучения. В процессе анализа выполняется построение ROC-кривой, которая показывает зависимость доли истинно положительных классификаций от доли ложных положительных классификаций для различных уровней изменяемого параметра. Расчеты, применяемые для построения характеристической кривой, основаны на формировании четырехпольной таблицы, которая имеет следующую структуру (таблица 1):

Таблица 1 – Таблица сопряженности (ошибок)

Результат по значению параметра	Результат по достижению конечной точки	
	Положительный	Отрицательный
Положительный	Истинно положительный	Ложноположительный
Отрицательный	Ложноотрицательный	Истинно отрицательный

Чувствительность (sensitivity, SE) определяется как доля истинно положительных наблюдений среди всех положительных результатов по достижению конечной точки.

Специфичность (specificity, SP) – доля истинно отрицательных наблюдений среди всех отрицательных результатов по достижению конечной точки. Долю ложноположительных случаев можно вычислить как разность единицы и специфичности.

Общая точность (accuracy, AC) – это доля правильных наблюдений в общем количестве результатов. Этот показатель рассчитывают, как отношение суммы истинно-положительных и истинно-отрицательных наблюдений теста к общему количеству результатов.

Для построения таблицы сопряженности была использована информация о баллах студентов при поступлении и результатах первых двух сессий студентов. В качестве изменяемого параметра будем считать факт достижения определенной суммы баллов и среднего балла аттестата соответственно выделенным пороговым значениям (240, 260, 280, 300, 320, 340, 360 баллов), в качестве конечных точек – средний балл по результатам первых двух сессий не менее 7 или не менее 9.

Для определения порогового значения количества баллов студентов при поступлении для каждого значения показателя были сформированы четырехпольные таблицы и по ним рассчитаны значения чувствительности (SE), специфичности (SP) и точности (AC).

Всего первый курс успешно завершили 416 студентов, по которым имеются данные о количестве баллов при поступлении. Из них по итогам первого года обучения имели средний балл, не менее девяти – 34 (8%) студента, не менее семи – 219 (53%) студентов. Данные представлены в таблице 2:

Таблица 2 – Распределение студентов по успеваемости по итогам первого года обучения в зависимости от порогового значения баллов при поступлении

Пороговое значение суммы баллов		Количество студентов	Количество студентов, если средний балл			
			≥ 7	< 7	≥ 9	< 9
240	больше	362	213	149	33	329
	меньше	54	6	48	1	53
260	больше	314	201	113	33	281
	меньше	102	18	84	1	101
280	больше	235	165	70	30	205
	меньше	181	54	127	4	177
300	больше	161	130	31	27	134
	меньше	255	89	166	7	248
320	больше	71	64	7	21	50
	меньше	345	155	190	13	332
340	больше	28	25	3	9	19
	меньше	388	194	194	25	363
360	больше	6	5	1	4	2
	меньше	410	214	196	30	380

Для построения ROC-кривой были рассчитаны значения чувствительности и специфичности для каждого значения порога отсечения (таблица 3):

Таблица 3 – Операционные характеристики в зависимости от порогового значения количества баллов при поступлении

Сумма баллов по ЦТ, ЦЭ	Средний балл не меньше 7				Средний балл не меньше 9			
	SE	SP	AC	SE+SP	SE	SP	AC	SE+SP
240	0,97	0,24	0,63	1,21	0,97	0,14	0,21	1,11
260	0,92	0,43	0,69	1,34	0,97	0,26	0,32	1,23
280	0,75	0,64	0,70	1,40	0,88	0,46	0,50	1,35
300	0,59	0,84	0,71	1,44	0,79	0,65	0,66	1,44

320	0,29	0,96	0,61	1,26	0,62	0,87	0,85	1,49
340	0,11	0,98	0,53	1,10	0,26	0,95	0,89	1,21
360	0,02	0,99	0,48	1,02	0,12	0,99	0,92	1,11

Характеристический график представлен на рисунке 1:

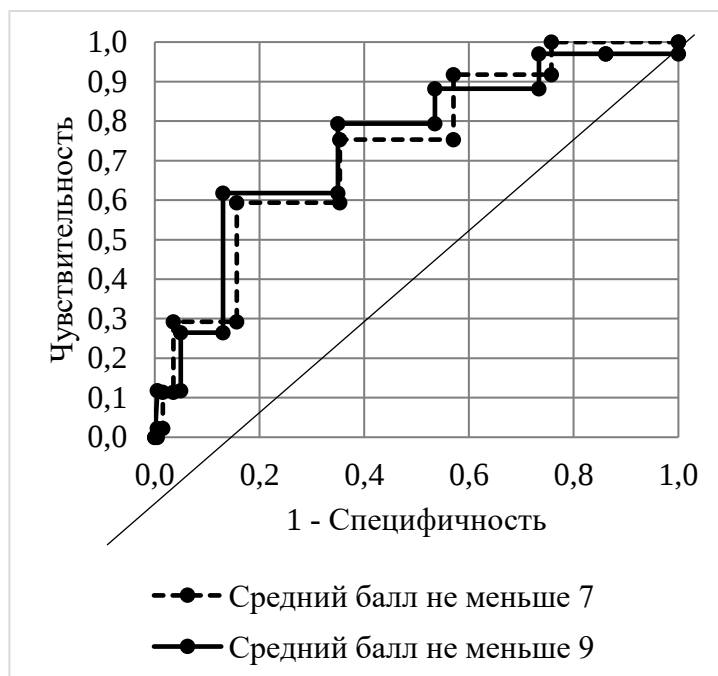


Рисунок 1 – Характеристическая кривая

Для идеальной модели оптимальным является максимальное значение для чувствительности и специфичности. Однако, большему значению чувствительности соответствует меньшее значение специфичности и наоборот, так как эти величины антагонистичны. Для того, чтобы найти точку разделения, которой будут соответствовать оптимальные значения чувствительности и специфичности, можно воспользоваться требованием максимальной суммарной чувствительности и специфичности модели, для чего нужно выбрать точку, которой соответствует максимальное значение произведения данных показателей.

Таким образом, если мы хотим выделить студентов, которые с большой долей вероятности будут успевать на 7, 8, 9, 10, оптимальной точкой разделения будет точка, соответствующая 300 баллам. В этом случае выполняется оптимальное соотношение чувствительности и специфичности. Если же мы хотим выделить будущих отличников, оптимальной точкой разделения будет точка, соответствующая 320 баллам.

С 2021 по 2024 гг. на факультет поступило 220 студентов, имеющих количество баллов при поступлении больше 300 и отнесенных к одаренным, что составляет 33,4% от общего количества студентов, поступивших на факультет за этот период (659 человек).

Закключение. Использование ROC-анализа для определения порогового значения количества баллов при поступлении позволяет в самом начале обучения выделить группу одаренных студентов для дальнейшей усиленной работы с ними. На факультете математики и информационных технологий таким значением является количество баллов при поступлении, большее 320.

1. Залесская, Е.Н. Анализ успеваемости студентов факультета математики и информационных технологий ВГУ имени П.М. Машерова, набравших высокие баллы по результатам вступительных испытаний / Е.Н. Залесская, А.А. Чиркина, С.А. Шпаков // Наука – образованию, производству, экономике: материалы 75-й Региональной научно-практической конференции преподавателей, научных сотрудников и аспирантов, Витебск, 3 марта 2023 г. – Витебск: ВГУ имени П. М. Машерова, 2023. – С. 680–684. URL: <https://rep.vsu.by/handle/123456789/37027> (дата обращения: 27.01.2025).