

УДК 378

DOI 10.52452/18115942_2022_3_229

**ИСПОЛЬЗОВАНИЕ МЕТОДА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ
ПРОГНОЗИРОВАНИЯ АКАДЕМИЧЕСКИ РИСКОВАННЫХ СТУДЕНТОВ В
ЗАВИСИМОСТИ ОТ ИХ ТЕМПЕРАМЕНТА (НА ПРИМЕРЕ ФАКУЛЬТЕТА ИМИКН
В НИУ ВШЭ-НИЖНИЙ НОВГОРОД)**

© 2022г.

Е.В. Шадрина

Шадрина Е.В. – старший преподаватель кафедры прикладная математика и информатика
Национального исследовательского университета “Высшая школа экономики”, Нижний
Новгород, eshadrina@hse.ru.

Статья поступила в редакцию 30.05.2022

Статья принята к публикации 26.07.2022

АННОТАЦИЯ

В статье рассматривается влияние темперамента на академическую успеваемость студентов первого курса в НИУ ВШЭ-Нижний Новгород на примере факультета Информатики, Математики и Компьютерных наук. Материалом для исследования послужила информация о темпераменте, степени экстраверсии, стабильности и других личностных чертах студентов, полученная путем проведения опроса. В исследовании принимали участие студенты первого и второго курса факультета ИМиКН 2017-2018 учебного года. Далее в работе определяются психологические факторы, влияющие на средний балл и вероятность пересдач для студентов с разными темпераментами. Обнаружена некоторая связь между темпераментом и академическим успехом, делающая возможным предсказание «рискованных» студентов. Для этого применяются различные методы машинного обучения: kNN-метод (k-ближайших соседей) и дерево принятия решений. В результате проведенных вычислений наилучший результат показал метод дерева принятия решений. В итоге студенты первого года обучения классифицированы по трем группам (Good, Medium, Bad) по степени риска получить академическую задолженность. Практическим результатом исследования стали рекомендации

учебному офису факультета ИМиКН обратить внимание на рискованных студентов и оказать им помощь в учебном процессе. После окончания летней сессии результаты классификации были проверены. В статье также представлен алгоритм поиска рискованных студентов с учетом темперамента.

Ключевые слова: психология в высшем образовании, академические достижения, темперамент, студенты, методы машинного обучения, дерево принятия решения, он-лайн анкетирование.

Введение

Академическая успеваемость или неуспеваемость студентов, а также отчисление из вузов стало в настоящее время важным явлением, которое требует всестороннего изучения [4]. Успеваемость студентов – основной параметр, по которому можно оценить, насколько доходчиво знания были переданы и усвоены. Имея возможность заранее предсказать результаты экзаменов, можно принять превентивные меры: провести факультативы, дополнительные консультации по предметам, чтобы уменьшить число студентов с неудовлетворительными результатами, академическими задолженностями или отчисление из вуза. И если в маленьких группах склонности студентов могут быть определены благодаря высокой контактности педагога с группой, то при увеличении их числа отслеживание тех из них, кто с большой вероятностью не сдаст экзамен или зачет, становится затруднительным.

Факультет Информатики Математики и Компьютерных наук (ИМиКН) в Нижегородском кампусе Национального Исследовательского Университета Высшая школа экономики (НИУ ВШЭ-Нижний Новгород) растет одновременно с увеличением набора первокурсников, каждый из которых обладает своим уровнем мотивации и настойчивости, природной одаренности, усидчивости и прочих личностных особенностей. Все эти качества могут давать им преимущество или, наоборот, создавать препятствия на пути к академическому успеху [5].

Несмотря на то, что абитуриенты, поступающие на факультет, имеют высокий балл ЕГЭ (Единый Государственной Экзамен), процент отчисленных после первого года обучения очень высок - около 30%, что особенно свойственно высокоселективным вузам, где процент отчислений в случаях возникновения академической задолженности обычно выше, чем в неселективных [6].

Поскольку среди университетов существует высокая конкуренция за студентов, высшие учебные заведения вынуждены учитывать потребности студентов. Направление ИТ, например, зачастую выбирают абитуриенты, склонные к замкнутости и интроверсии, что

требует от преподавателей особенного подхода к данному типу аудитории. ВУЗы, учитывающие комплекс потребностей студентов (бытовые, личностные, академические) в большей степени, будут обладать несомненным конкурентным преимуществом. Не менее важно уметь понять, кто такие успешные студенты, что их интересует и каковы их намерения и амбиции. На факультете ИМиКН в НИУ ВШЭ-Нижний Новгород студенты с сильной подготовкой по информатике имеют возможность заниматься спортивным (олимпиадным) программированием, успешно выступают на международных олимпиадах.

Исходя из вышесказанного, мы убеждены, что проблема прогнозирования успеваемости студентов на основе их личных особенностей становится все более острой в современном образовании.

За последние 20 лет разными авторами были проведены исследования, выявляющие признаки, больше всего влияющие на академическую успеваемость и “выбытия” студентов из вуза [8, 9, 14]. Социологические теории принимают во внимание университетскую и семейную среду студента, механизмы социализации, влияние ключевых людей [3].

Масштабное исследование, выявившее наиболее влиятельные факторы, было проведено, например, Д.Ф. Супербайем с соавторами в Бельгийском университете [16]. Более свежее исследование было проведено Ластом и соавторами в Бельгийском и Французском университетах с использованием продвинутых методов интеллектуального анализа данных и машинного обучения [13]. В НИУ ВШЭ-Нижний Новгород было проведено похожее исследование в 2014 году [2]. В целом, все эти исследования показали, что наибольший вес имеют факторы, связанные с академическим поведением (посещаемость лекций и семинаров, умение записывать лекции в аудитории, уверенность в выборе университета/факультета/направления подготовки, выполнение домашнего задания, посещение факультативов), а также факторы личной истории (образование родителей, внеучебные активности в школе, выпускные оценки в школе). Они также показали, что особенности характера тоже имеют вес, но не столь значительный по сравнению с описанными факторами. В свою очередь, О.В. Польдин с соавторами в своей статье объясняет, что общительные и активные студенты вследствие общения развиваются в личностном плане и как следствие, становятся более успешными [15]. В работе [10] выявлена связь, показывающая, что с течением времени происходит социальная изоляция студентов с академическими задолженностями, создающая для них дополнительные риски исключения из вуза.

В упомянутых выше работах отмечается влияние дружеских связей на успеваемость студентов, и авторами данной работы было замечено: кто умеет дружить, тот умеет учиться.

С другой стороны, авторами было замечено, что студенты выбирающие IT направление, более других подвержены замкнутости и им тяжело дается раскрыться в новом коллективе как с одноклассниками так и с преподавателями. Таким образом, исходя из личного опыта и опыта коллег, возникла идея более внимательно изучить психологические аспекты личности студентов.

В данной работе мы будем опираться на типологию четырех темпераментов Юнга [1]:

- холерик – характеризуется яркостью и силой протекания эмоциональных процессов. Холерики вспыльчивы, страстны, энергичны.
- сангвиник – индивид отличается сравнительно слабой интенсивностью процессов психики с быстрой сменой одних процессов другими. Сангвиники жизнерадостны, работоспособны, легко справляются с различными задачами.
- флегматик – отличается медлительностью, вялостью движений, не энергичностью. Чувства флегматика ровны и спокойны
- меланхолик – характеризуется глубиной эмоциональных проявлений, но медленностью протекания психических процессов. Такие люди болезненно реагируют на внешние обстоятельства, часто проявляют себя как пассивные и вялые.

Целью данной работы стало исследование влияния особенностей личности (то есть темперамента как основу личности) на академическую успеваемость и определение студентов с наибольшей вероятностью не сдать экзамены.

Наша главная гипотеза заключается в том, что существует некоторая зависимость между темпераментом и академической успеваемостью студента. Материалом для исследования послужила информация о темпераменте, степени экстраверсии, стабильности и других личностных черт студентов, полученная путем проведения опроса. В исследовании принимали участие студенты первого и второго курса факультете ИМиКН 2017-2018 учебного года.

Получение и обработка данных

Известный ученый Борис Миркин обращает внимание на важность классификации в задачах анализа данных. Классификация (Classification) — построение классификации, структурирующее рассматриваемое множество явлений в совокупность отдельных классов, отражающих важные свойства этих явлений. В настоящее время этот термин также применяется к задачам отнесения отдельных объектов к заранее заданным классам. [7].

Главная задача нашего исследования - классифицировать студентов на три группы:

- Good - это “не рискованные” студенты, у них не будет академических задолженностей;

- Medium - студенты “со средней вероятностью не сдать экзамен”, они могут успешно сдать сессию, благодаря мерам поддержки, принятым в университете;
- Bad - это “студенты с высоким риском” не сдать экзамен, бросить учебу.

Таким образом, нам необходимо создать набор данных, в котором каждый студент описывается в соответствии с рядом личностных характеристик.

Для получения данных о темпераменте студентов и ряде других особенностей характера был проведен опрос во внутренней системе НИУ ВШЭ-Нижний Новгород LMS (Learning Management System). В нем приняли участие 90 студентов первого и 50 студентов второго курса факультета Информатики Математики и Компьютерных Наук НИУ ВШЭ-Нижний Новгород.



Рисунок 1. Классификация темпераментов по экстраверсии и рациональности

Опрос был составлен на основе теста-опросника Г.Айзенка [1] для определения типов темперамента каждого студента по показателям экстраверсия-интроверсия и нестабильность-стабильность. Были также включены вопросы об активности в школе, усидчивости, умении расставлять приоритеты, все это личные характеристики студента. Блок по выявлению темперамента состоял из 12 вопросов, каждый из которых описывал один из темпераментов: холерик, сангвиник, флегматик или меланхолик. Используя ответы студентов, мы рассчитали такие параметры, как экстраверсия и рациональность (рис. 1), затем определялся темперамент каждого студента. Всего в анкете студентам предстояло ответить на 17 личных вопросов. Данные о промежуточном и итоговом контроле (оценки за экзамены) после первой половины учебного года мы взяли из внутренней базы данных НИУ ВШЭ АСАВ (Абитуриент Студент Аспирант Выпускник). Далее, после объединения всей информации мы извлекли 13 двоичных переменных для каждого студента (переменные вида 0/1). Основная переменная для принятия решения, используемая для валидации нашей модели, представляет собой среднюю оценку по всем дисциплинам (средний балл) и информацию о сдаче/несдаче экзаменов из общего

рейтинга студентов. На основе этой переменной и происходит классификация по группам Good, Medium, Bad.

Распределение темпераментов в выборке показано на рис. 2. Меланхолики (27%) и сангвиники (26%) составляют примерно одну четвертую опрошенных. Количество холериков (34%) в три раза превышает количество флегматиков (13%).



Рисунок 2. Распределение темпераментов в выборке студентов

Поскольку число опрошенных первокурсников было в два раза больше, чем число второкурсников, мы решили составить тренировочную выборку из общего числа студентов второго курса и половины первокурсников, всего 100 человек. Выборка, на которой будет проверяться работа модели, состоит из 40 студентов первого курса.

Студенты из тренировочной выборки в зависимости от среднего балла и наличия пересдач были разделены на категории Good, Medium и Bad, соответственно с низкой, средней и высокой вероятностью не сдать какой-либо из экзаменов (рис. 3). В категорию «Good» вошли студенты первой трети рейтинга без пересдач, в категорию «Medium» - второй трети рейтинга без пересдач, а в категорию «Bad» вошли все остальные студенты (рис. 3а, рис. 3б).

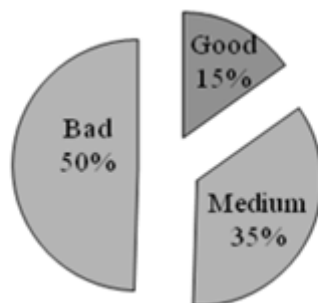


Рис 3. Распределение студентов по категориям

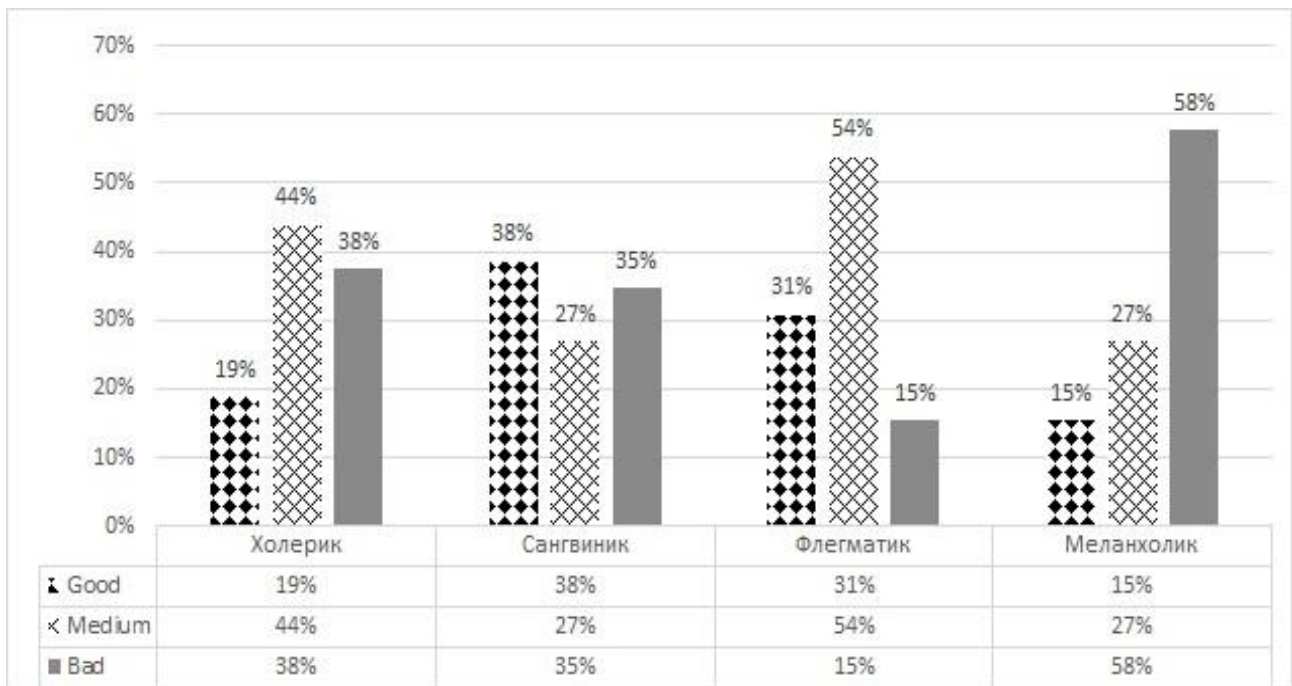


Рис 3а. Распределение темпераментов по категориям (в %)

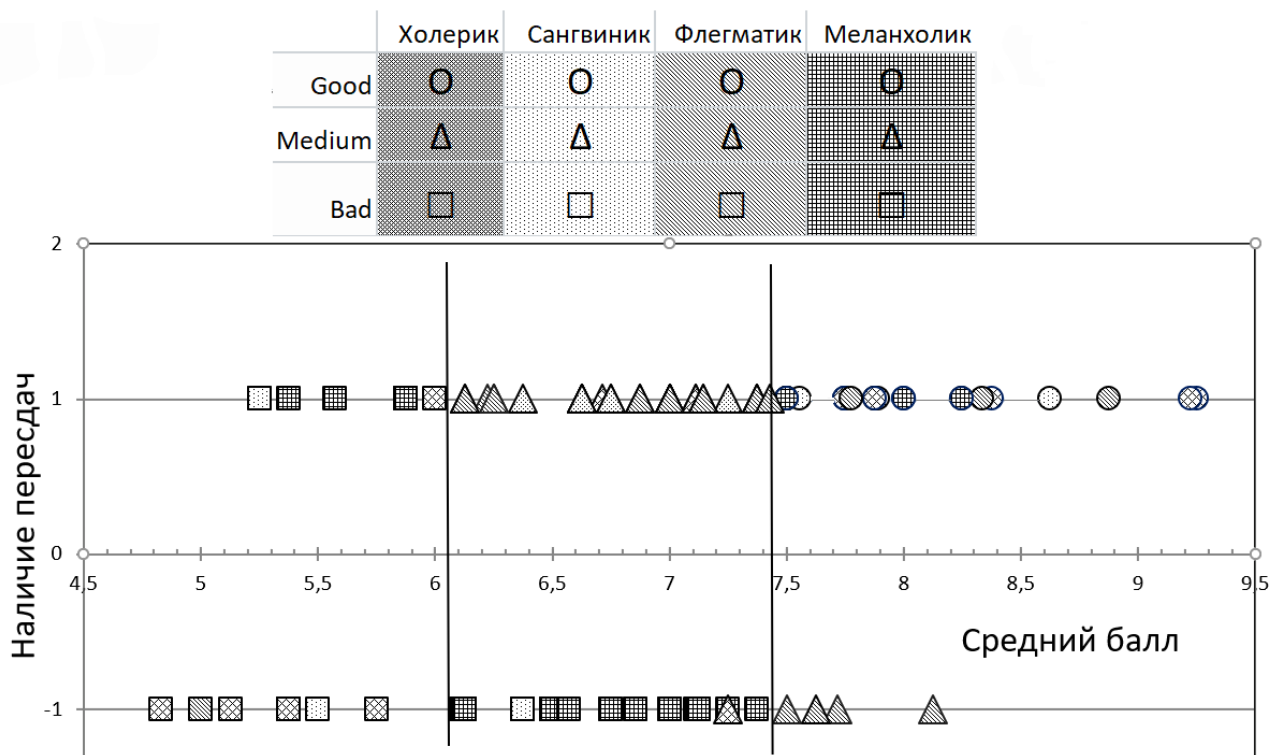


Рис 3б. Распределение темпераментов по категориям(Good/Medium/Bad) в зависимости от среднего балла и наличия/отсутствия пересдач

На рисунке 3 видно, что половина студентов попала в категорию “Bad”. Такие результаты разбиения объясняются тем, что почти каждый второй из опрошенных имел пересдачу в семестре. Для факультета ИМиКН это характерная ситуация из-за ряда непростых дисциплин, изучаемых в начале первого и второго курса. Стоит отметить, что распределение

в категорию «Bad» не означает, что студент все экзамены сдает неудовлетворительно и стоит на пороге отчисления. Многие из попавших в «Bad» имеют хороший средний балл, но долг по одной из дисциплин. Так как целью работы является выявление всех студентов, у которых ожидаются пересдачи, то даже студенты с хорошей успеваемостью, но наличием задолженностей, должны быть определены в категорию Bad. На рис. 3а видны как общие тенденции, так и различия между темпераментами. Картина распределения по категориям у холериков и флегматиков похожая: наибольшее количество студентов попало в категорию Medium (44% холерики, 54% флегматики). Среди сангвиников 38% приходится на студентов из категории Good. Для меланхоликов наибольшее количество студентов (58%) попадает в категорию Bad.

Рассмотрим фактор наличия/отсутствия пересдачи более подробно. На рис. 3б показано распределение темпераментов по категориям (Good, Medium, Bad) в зависимости от среднего балла и наличия/отсутствия пересдачи. На рис. 3б явно видно:

- круглые отличники (Good, обозначены фигурой круг) имеют высокий средний балл и отсутствие пересдач;
- студенты категории Medium (обозначены фигурой треугольник) имеют средний балл от 6,3 до 7,4 и не имеют пересдач;
- студенты категории Bad (обозначены фигурой квадрат) имеют низкий средний балл (ниже 6) и/или пересдачи; большинство студентов категории Bad пересдачи имеют.

Методы

Задача распределения студентов на категории с высокой, средней и низкой вероятностью академических неудач является задачей классификации, основанной на обучении с учителем [12].

На первом этапе работы мы выбрали такие параметры, которые некоторым образом влияют на средний балл каждого студента и возможные пересдачи.

На втором этапе работы мы искали алгоритм, который оптимальным образом (с наибольшим количеством угаданных студентов из категории Bad) осуществляет классификацию студентов тренировочной выборки. Нам необходимо, чтобы алгоритм определял каждого студента в нужную категорию (Bad/Medium/Good) с учетом его (студента) формализованных признаков. Для этого были использованы такие методы машинного обучения как kNN-алгоритм (k Nearest Neighbor или k ближайших соседей) и дерево принятия решений (Decision tree) [11, 16]. Поскольку мы заранее не знали, какой метод даст наиболее

точный результат, мы проверили оба и нашли такие параметры, при которых точность угадывания неуспешных студентов стала наилучшей.

На третьем этапе работы мы определили лучшую модель и использовали ее для предсказания успешности/неуспешности студентов первокурсников.

Поиск значимых параметров

Коэффициент корреляции позволяет оценить зависимость между двумя или более параметрами и довольно успешно используется в социологии образования и интеллектуальном анализе данных [12]. Для анализа данных мы представили полученные данные опроса в удобном формате: ответы студентов преобразованы в вид булевых переменных: параметр равен 1, если студент обладает характеристикой, и 0 в противном случае. Используя коэффициент корреляции, мы выбрали из всех параметров те, которые оказывают наибольшее влияние на средний балл и пересдачи. Результаты расчетов приведены в таблице 1. Существенным, для нашего исследования, является значение коэффициента корреляции, превышающие 0,2 по абсолютному значению (в таблице выделено курсивом) и несущественным, значение коэффициента корреляции находящееся в пределах от 0 до 0,2.

Таблица 1. Зависимость между академическими успехами и темпераментом студентов

Параметр	Корреляция со средним баллом	Корреляция с наличием пересдачи
Активность в школе	0.06	0.09
<i>Усидчивость</i>	<i>0.19</i>	<i>-0.24</i>
<i>Умение расставлять приоритеты</i>	<i>0.29</i>	<i>-0.18</i>
Проживание в общежитии	0.06	-0.05
Проживание с друзьями	-0.01	0.01
Проживание с родителями	-0.01	0.02
Одиночное проживание	-0.04	0

<i>Экстраверсия</i>	0.07	-0.18
Рациональность	-0.004	-0.08
<i>Холерик</i>	-0.1	0.044
<i>Сангвиник</i>	0.05	-0.16
<i>Флегматик</i>	0.13	0.017
<i>Меланхолик</i>	-0.04	0.2

Как видно из таблицы 1 параметр «Усидчивость» и «Умение расставлять приоритеты» в значительной степени зависят от успеваемости, что вполне объяснимо и понятно. При этом темперамент и вытекающие из него психологические характеристики менее коррелируют с успехами в учебе. Но поскольку мы исследуем влияние темпераментов (Холерик/Сангвиник/Флегматик/Меланхолик) на академический успех, то темпераменты выделены курсивом, так как их абсолютное значение корреляции хотя бы с одним из параметров превышает 0,1.

Таким образом, связь между академическими успехами и темпераментом существует и в некоторых случаях весьма значительна. Например, холерики имеют ниже средний балл ($\text{cor} = -0.1$), в то время как усидчивость флегматиков позволяет иметь оценки выше ($\text{cor} = 0.13$). Сангвиники реже ходят на пересдачи ($\text{cor} = -0.16$), в то время как шанс для меланхолика попасть на нее значителен ($\text{cor} = 0.2$).

Сравнение классификации с помощью kNN-метода и Деревя принятия решений

Первым был протестирован kNN-метод.

- используя все признаки с высокой корреляцией, а также средний балл. Наилучший результат был достигнут при восьми ближайших соседях (74%);
- все признаки без использования среднего балла. Лучший результат при пяти ближайших соседях (60%). Это является хорошим показателем, потому что вероятность попадания в нужную категорию в два раза больше случайного угадывания категории (которая составляет 33%);

Следует отметить, что в методе kNN психологические признаки уменьшают точность вычислений, поэтому полученная классификация для наших целей не подходит.

Вторым был протестирован метод дерева принятия решений:

- при использовании всех признаков для построения точность достигала 74%;
- без учета среднего балла дерево строилось на признаках «Экстраверсия» и «Рациональность» и точность достигла 62%;
- на основе только среднего балла дерево дало точность 76%;
- наилучший результат был получен при построении дерева на признаках «Усидчивость», «Умение расставлять приоритеты», «Экстраверсия», «Рациональность» и тип темперамента, и точность достигла 84%.

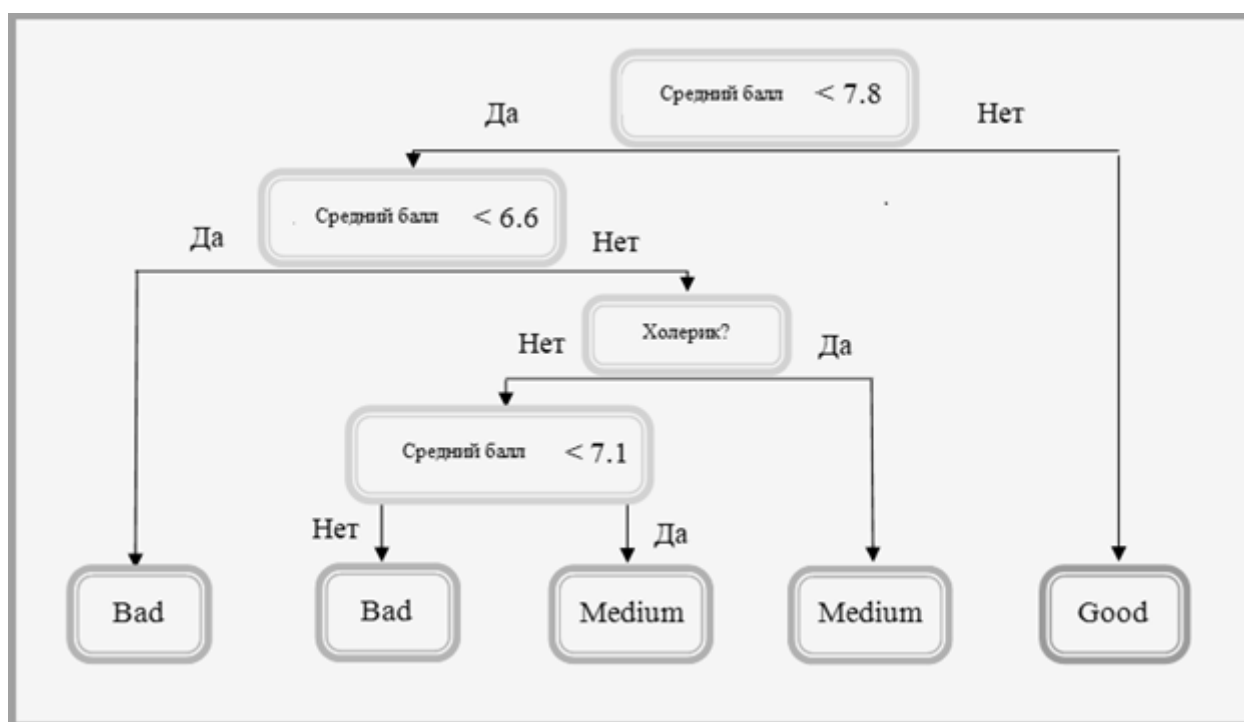


Рисунок 4. Дерево принятия решений с параметрами классификации

Как можно заметить, разная комбинация параметров может дать прирост точности угадывания по сравнению с использованием всех признаков.

Точность ответов при использовании обоих методов колебалась около 75%, но лучший результат показало дерево принятия решений. Поэтому для анализа был выбран именно этот метод. Итоговое дерево принятия решения с параметрами представлено на рис. 4.

Результаты исследования

С помощью дерева принятия решений, полученного на предыдущем шаге, студенты первого курса были разделены на три группы: Bad, Medium и Good. В целях сокрытия личной информации о студентах, фамилия каждого была заменена условным обозначением Студент1, Студент2, ..., СтудентN. Мы получили, что в категорию Bad (категория рискованных студентов) попали 22 студента, в категорию Medium – 12 студентов, в категорию Good – 6

студентов. После итоговой сессии в 4 модуле и всех пересдач осенью, появилась возможность проверить результаты исследования.

В таблице 2 представлены данные о реальном рейтинге и наличии пересдач студентов по итогам первого года обучения.

Таблица 2 Обобщение данных

Предсказанная Категория	Всего студентов	Студентов с пересдачами		Студентов отчислено или ИУП		Высокий средний балл строго больше 7,5)	
		чел	%	чел	%	чел	%
Bad	22	14	64%	6	27%	0	0%
Medium	12	3	25%	1	8%	4	33%
Good	6	0	0%	0	0%	6	100%

В категорию Bad попали 22 студента, 14 студентов (64%) имели пересдачи, по результатам всех осенних пересдач 6 студентов (27%) студентов остались с академическими задолженностями или были отчислены. Следует обратить внимание, что у 4 из 6 студентов остался долг по одной дисциплине (Математический анализ или Линейная алгебра) и им был предложен ИУП (Индивидуальный учебный план), но эти студенты потеряли бюджетное место.

В категорию Medium попало 12 студентов, 3 студента (25%) - имели пересдачи, но успешно исправили долги и перешли на 2 курс без академической задолженности. Одна студентка из категории Medium была отчислена по собственному желанию, а не за академическую задолженность, так как решила кардинально сменить сферу деятельности.

Среди студентов, попавших в категорию Good (6 человек) пересдач не было и никто из студентов этой категории не был отчислен.

Таким образом, можно сделать вывод о том, что результат распределения студентов по категориям хорошо согласовывается с реальной ситуацией. Следует также обратить внимание на средний балл студентов: студентов с высоким средним баллом (строго больше 7,5) больше в категории Good, меньше в категории Medium и нет в категории Bad.

В результате проведенного исследования были даны следующие рекомендации учебному офису факультета информатики, математики и компьютерных наук НИУ ВШЭ – Нижний Новгород:

- обратить внимание на студентов из категории Bad (беседа со студентами и их законными представителями);
- предложить студентам категории Bad и Medium дополнительные занятия в виде факультатива;
- с помощью механизма учебных ассистентов разбирать и проверять домашние задания;
- уделить дополнительное внимание студентам из категории Medium с невысоким средним баллом после первого учебного модуля.

В заключение, опишем алгоритм поиска рискованных студентов с учетом темперамента:

1. Проведение опроса студентов 1 и 2 курса (сентябрь-октябрь нового учебного года) с целью получить данные о психологических характеристиках
2. Перевод результатов опроса в булевы переменные (в соответствии с параметрам в таблице 1)
3. Формирование тренировочной выборки из студентов 2 курса
4. Получение информации о рейтинге и наличии пересдач (для студентов из тренировочной выборки)
5. Вычисление корреляции психологических характеристик со средним баллом и наличием пересдач
6. Выбор значимых для нашего исследования параметров
7. Классификация студентов из тренировочной выборки по категориям Good, Medium, Bad в зависимости от рейтинга и наличия пересдач
8. Построение дерева принятия решений по значимым параметрам в соответствии с полученной классификацией
9. Формирование тестовой выборки из студентов 1 курса
10. Получение информации о рейтинге и наличии пересдач после 1 модуля (ноябрь текущего учебного года)
11. Классификация студентов 1 курса с помощью построенного дерева принятия решений
12. Рекомендации учебному офису обратить внимание на успеваемость студентов, попавших в категорию Bad, а также в категорию Medium с невысоким средним баллом (не позже начала декабря текущего учебного года).

Заключение

Изучая влияние темперамента на успеваемость студентов в НИУ ВШЭ-Нижний Новгород на факультете ИМиКН, мы определили наиболее важные психологические факторы. Ими оказались «Усидчивость» и «Умение расставлять приоритеты» - наличие этих факторов

резко увеличивало средний балл и уменьшало вероятность пересдач. Было замечено, что вспыльчивые холерики имеют более низкий средний балл, в то время как спокойствие и размеренность флегматиков помогают им учиться лучше. Сангвиники реже ходят на пересдачи, тогда как шанс меланхолика на повторную сдачу экзаменов значительно выше. Чем больше у студента выражена экстраверсия, тем выше его средний балл.

Мы убеждены в том, что наше исследование будет полезно другим университетам для:

- 1) выявления академически неуспешных студентов и фокусирования ресурсов факультета на “рискованных” студентах. Одна из важнейших целей факультета ИМиКН в НИУ ВШЭ-Нижний Новгород - обучить и перевести без академических долгов на второй курс как можно больше принятых на программу студентов первого курса. Иначе ресурсы государства (в случае бюджетного обучения) или личные средства студента (в случае коммерческого обучения), затраченные на образование, будут использованы нерационально.
- 2) формирования индивидуальной образовательной траектории. В НИУ ВШЭ-Нижний Новгород на сегодняшний день, существуют гибкие возможности перехода с одной образовательной программы на другую, используя ИУП. Мы предполагаем, что формирование учебных групп, учитывая личные характеристики студента, повысит успеваемость каждого студента.

Мы понимаем, что, в нашем исследовании есть некоторые ограничения. В данной работе был представлен небольшой пул исходных данных (140 ответов студентов). Но факультет ИМиКН в НИУ ВШЭ-НН не такой большой, чтобы мы могли говорить о “настоящей” Big Data и использовать все возможности алгоритмов машинного обучения.

В дальнейшем мы планируем использовать лонгитюдные данные для проверки результатов исследования в НИУ ВШЭ-Нижний Новгород. Так же нам было бы интересно использовать другие методы интеллектуального анализа данных для прогнозирования академически рискованных студентов.

Благодарности

Мы благодарим НИУ ВШЭ - Нижний Новгород за предоставленные данные для проведения исследования.

Литература

1. Айзенк, Г., & Вильсон, Г. Как измерить личность. М.: Когито-центр. 2000. 156–159.

2. Булычева П.А., Ошмарина О.Е., Шадрина Е.В. Выявление академически неуспешных студентов на первом году обучения в университете на примере НИУ ВШЭ-Нижний Новгород//Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки. 2016. № 2(42). С. 136-143.
3. Горбунова Е.В. Выбытия студентов из вузов: исследования в России и США//Вопросы образования //Educational Studies Moscow. 2018. № 1. С. 110–131
4. Горбунова Е.В., Груздев И.А., Фруммин И.Д. Студенческий отсев в российских вузах: к постановке проблемы // Вопросы образования / Educational Studies Moscow. 2013. № 2. С. 67–81.
5. Кейек-Франсен Д. Практики успешности студентов: от очного обучения к масштабному и обратно//Вопросы образования. 2018. № 4. С. 116–138.
6. Кочергина Е. В., Прахов И. А. Взаимосвязь между отношением к риску, успеваемостью студентов и вероятностью отчисления из вуза//Вопросы образования. 2016. № 4. С. 206–228.
7. Миркин Б. Г. Введение в анализ данных: учебник и практикум для бакалавриата и магистратуры//М.: Издательство Юрайт. 2014. 174 с.
8. Arulselvan, A., Boginski, V., Mendoza, P., & Pardalos, P. Predicting the nexus between post-secondary education affordability and student success: An application of network-based approaches. ASONAM'09. International Conference on Advances in Social Networks Analysis and Mining. 2009.149–154.
9. Campbell, J.P. Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study. Purdue University. 2007.
10. Dokuka, S.V., Valeeva, D.R., Yudkevich, M.M. How Academic Failures Break Up Friendship Ties: Social Networks and Retakes//Educational Studies Moscow. 2017. №1 .8-21.
11. Friedl, M. A., & Brodley, C. E. Decision tree classification of land cover from remotely sensed data//Remote sensing of environment. 1997. 61(3), 399–409.
12. Luan, J. Data mining and its applications in higher education// New directions for institutional research, 2002. (113), 17–36.
13. Lust, T., Meskens, N., & Ahues, M. Predicting academic success in Belgium and France Comparison and integration of variables related to student behavior. arXiv preprint arXiv:1408.4955. 2014.
14. Pal, S. Mining educational data using classification to decrease dropout rate of students. arXiv preprint arXiv:1206.3078. 2012.

15. Poldin, O., Valeeva, D., & Yudkevich, M. How social ties affect peer group effects: Case of university students//SSRN Electronic Journal, January 2013. DOI: 10.2139/ssrn.2207666
16. Superby, J.-F., Vandamme, J., & Meskens, N. Determination of factors influencing the achievement of the first-year university students using data mining methods. Workshop on Educational Data Mining. 2006. 32, 234.

USING THE METHOD OF DATA MINING TO PREDICT ACADEMICALLY RISKY STUDENTS DEPENDING ON THEIR TEMPERAMENT (USING THE EXAMPLE OF THE IM&CS FACULTY AT NRU HSE-NIZHNY NOVGOROD)

Shadrina E.V.

senior teacher, Department of Applied Mathematics and Informatics, National Research University
Higher School of Economics Nizhny Novgorod, Nizhny Novgorod, Russia. E-mail:
eshadrina@hse.ru

ABSTRACT

The article discusses the influence of temperament on the academic performance of the first-year students at HSE-Nizhny Novgorod on the example of the Faculty of Informatics, Mathematics and Computer Science. Analysis was held with the help of statistics methods and methods of data mining. The baseline data for the study is information about students, collected using the online support system for the educational process at HSE - LMS (Learning Management System). The material for the study was the information about temperament, degree of extroversion, stability, and other personality traits of students, obtained by conducting a survey. The study involved students of the first and second years of the faculty of the IM&CS 2017-2018 academic year. Further, the work identifies psychological factors affecting the average score and the probability of re-training for students with different temperaments. Some connection was found between temperament and academic success, which makes possible the prediction of "risky" students. For this, various machine learning methods are used: the kNN-method (k-nearest neighbors) and the decision tree. As a result of the calculations, the best result was shown by the decision tree method. As a result, first-year students are classified into three groups (Good, Medium, Bad) according to the degree of risk of getting academic debt. The practical result of the research was the recommendations to the educational office of the Faculty of IM&CS to pay attention to risky students and assist them in the

educational process. After the end of the summer session, the classification results were checked. The article also presents an algorithm for finding risky students, taking temperament into account.

Keywords: psychology of higher education, academic achievements, human temperament, students, data mining, decision tree, on-line questionnaire.

Acknowledgements

We would like to thank the education office specialists of HSE University Nizhny Novgorod for providing data for conducting the research.

References

1. Arulselvan, A., Boginski, V., Mendoza, P., & Pardalos, P. Predicting the nexus between post-secondary education affordability and student success: An application of network-based approaches. ASONAM'09. International Conference on Advances in Social Networks Analysis and Mining. 2009, pp. 149–154.
2. Bulycheva, P., Oshmarina, O., Shadrina, E.: Identifying academically “unsuccessful” first-year students: a case study of Higher School of Economics – Nizhny Novgorod. In: Vestnik of Lobachevsky State University of Nizhny Novgorod. Series: Social Sciences, 2016, no. 2(42), pp. 136-143.
3. Campbell, J.P. Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study. Purdue University, 2007.
4. Dokuka, S.V., Valeeva, D.R., Yudkevich, M.M. How Academic Failures Break Up Friendship Ties: Social Networks and Retakes//Educational Studies Moscow, 2017, no. 1 pp.8 -21.
5. Friedl, M. A., & Brodley, C. E. Decision tree classification of land cover from remotely sensed data//Remote sensing of environment, 1997, no. 61(3), pp. 399–409.
6. Gorbunova E. Vybytiya studentov iz vuzov: issledovaniya v Rossii i SShA [Elaboration of Research on Student Withdrawal from Universities in Russia and the United States]. Voprosy obrazovaniya / Educational Studies Moscow, 2018, no. 1, pp. 110-131.
7. Gorbunova E. V., Gruzdev I. A., Frumin I. D. Studencheskii otsev v rossiiskikh vuzakh: k postanovke problemy [Student Dropout in Russian Higher Education Institutions: the Problem Statement]. Voprosy obrazovaniya [Educational Studies], 2013, no. 2, pp. 67–81.
8. Izenk G. & Vilson G. Kak izmerit lichnost [How to measure a personality]. M.: Kogito tsentr, 2000, pp.156–159.

9. Keyek-Franssen D. Praktiki uspehnosti studentov: ot ochnogo obucheniya k masshtabnomu i obratno [Practices for Student Success: From Face-to-Face to At-Scale and Back]. Voprosy obrazovaniya / Educational Studies Moscow, 2018, no. 4, pp. 116-138.
10. Kochergina E., Prakhov I. Vzaimosvyaz' mezhdou otnosheniem k risku, uspevaemost'yu studentov i veroyatnost'yu otchisleniya iz vuza [Relationships between Risk Attitude, Academic Performance, and the Likelihood of Drop-outs]. /Voprosy obrazovaniya [Educational Studies] Moscow, 2016, no 4, pp. 206-228.
11. Luan, J. Data mining and its applications in higher education// New directions for institutional research, 2002, no. 113, pp. 17–36.
12. Lust, T., Meskens, N., & Ahues, M. Predicting academic success in Belgium and France Comparison and integration of variables related to student behavior. arXiv preprint arXiv:1408.4955. 2014.
13. Mirkin B. G. Vvedeniye v analiz dannyh: uchebnik I praktikum dlya bakalavriata I magistratury [Introduction to data analyses: theory and practice for bachelor and master courses], M.: Izdatelstvo Urite, 2014, 174 p.
14. Poldin, O., Valeeva, D., & Yudkevich, M. How social ties affect peer group effects: Case of university students//SSRN Electronic Journal, January 2013. DOI: 10.2139/ssrn.2207666
15. Pal, S. Mining educational data using classification to decrease dropout rate of students. arXiv preprint arXiv:1206.3078. 2012.
16. Superby, J.-F., Vandamme, J., & Meskens, N. Determination of factors influencing the achievement of the first-year university students using data mining methods. Workshop on Educational Data Mining. 2006, no. 32, pp. 234.