

УДК 51-7

АНАЛИЗ ВЗАИМОСВЯЗИ ФАКТОРОВ, ВЛИЯЮЩИХ НА УСПЕ- ВАЕМОСТЬ ОБУЧАЮЩИХСЯ В СРЕДНИХ ОБРАЗОВАТЕЛЬНЫХ ОРГАНИЗАЦИЯХ

Матвеева А.О., студент гр. ИИМ-251, I курс
Научный руководитель: Ермакова И.А., д.т.н., профессор
Кузбасский государственный технический университет имени
Т. Ф. Горбачева, г. Кемерово

Успеваемость школьников - ключевой показатель эффективности учебного процесса. Её определяют многочисленные факторы: индивидуальные способности и мотивация учащихся, предшествующий опыт обучения, а также социально-семейный контекст. В настоящем исследовании оценивается влияние ряда доступных факторов на академические результаты школьников. Исследование основано на реальных данных о 25000 учащихся с различными характеристиками, и направлено на выявление наиболее значимых из них. На рисунке 1 приведены первые 15 строк исходных данных, данные взяты с сайта [1].

student_id	age	gender	school_type	parent_education	study_hours	attendance_percentage	internet_access	travel_time	extra_activities	final_grade
1	14	male	public	post graduate	3.1	84.3	yes	<15 min	yes	e
2	18	female	public	graduate	3.7	87.8	yes	>60 min	no	d
3	17	female	private	post graduate	7.9	65.5	no	<15 min	no	b
4	16	other	public	high school	1.1	58.1	no	15-30 min	no	e
5	16	female	public	high school	1.3	61.0	yes	30-60 min	yes	f
6	19	male	public	no formal	3.8	69.6	yes	>60 min	yes	d
7	14	female	private	post graduate	1.8	81.6	yes	30-60 min	no	f
8	18	female	private	post graduate	5.6	59.4	yes	>60 min	yes	d
9	15	other	private	high school	3.2	89.6	yes	15-30 min	yes	d
10	14	female	public	diploma	6.8	62.4	yes	>60 min	no	d
11	17	female	private	graduate	6.1	90.5	yes	15-30 min	no	c
12	19	other	public	post graduate	4.0	96.2	yes	<15 min	yes	d
13	18	female	private	high school	6.8	58.2	yes	>60 min	no	c
14	18	other	public	diploma	1.3	95.9	yes	30-60 min	yes	e
15	18	other	public	high school	4.9	85.3	yes	<15 min	no	d

Рисунок 1 – Исходные данные

Целью работы является построение и анализ регрессионной модели, которая описывает зависимость успеваемости (общий итоговый балл) от перечисленных ниже факторов.

В работу включены следующие независимые переменные:

- X1- age (возраст)
- X2 - gender (пол)
- X3 - school_type (тип школы)
- X4 - parent_education (уровень образования родителей)
- X5 - study_hours (среднее время ежедневной самостоятельной учёбы),
- X6 - attendance_percentage (процент посещения занятий)
- X7 - internet_access (наличие доступа в интернет дома)

- X8 - travel_time (время в пути до учебы)
- X9 - extra_activities (участие во внешкольных активностях),

В качестве целевой переменной используется final_grade (средний балл по основным предметам). Модель предполагает линейную связь между факторами и итоговым баллом. Все категориальные признаки были закодированы для последующего регрессионного анализа. Ниже представлена расшифровка:

- gender:
 - male (мужской): 0,
 - female (женский): 1,
 - other (не указан): 2
- school_type:
 - public (государственная): 0,
 - private (частная): 1
- parent_education:
 - no formal (без образования): 0,
 - high school (средняя школа): 1,
 - diploma (высшее образование, бакалавр): 2,
 - graduate (высшее образование, магистр): 3,
 - post graduate (высшее образование, аспирантура): 4,
- internet_access:
 - no: 0,
 - yes: 1
- travel_time:
 - <15 min: 0,
 - 15-30 min: 1,
 - 30-60 min: 2,
 - >60 min: 3
- extra_activities:
 - no: 0,
 - yes: 1
- final_grade:
 - f: 0,
 - e: 1,
 - d: 2,
 - c: 3,
 - b: 4,
 - a: 5

На рисунке 2 представлены подготовленные для регрессионного анализа данные.

	A	B	C	D	E	F	G	H	I	J
1	student_id	age	gender	school_type	parent_education	study_hours	attendance_percentage	internet_access	travel_time	extra_activities
2	1	14	0	0	4	3,1	84,3	1	0	1
3	2	18	1	0	3	3,7	87,8	1	3	0
4	3	17	1	1	4	7,9	65,5	0	0	0
5	4	16	2	0	1	1,1	58,1	0	1	0
6	5	16	1	0	1	1,3	61	1	2	1
7	6	19	0	0	0	3,8	69,6	1	3	1
8	7	14	1	1	4	1,8	81,6	1	2	0
9	8	18	1	1	4	5,6	59,4	1	3	1
10	9	15	2	1	1	3,2	89,6	1	1	1
11	10	14	1	0	2	6,8	62,4	1	3	0
12	11	17	1	1	3	6,1	90,5	1	1	0
13	12	19	2	0	4	4	96,2	1	0	1
14	13	18	1	1	1	6,8	58,2	1	3	0
15	14	18	2	0	2	1,3	95,9	1	2	1
16	15	18	2	0	1	4,9	85,3	1	0	0

Рисунок 2 – Подготовленные данные

Одним из факторов, который может оказывать существенное влияние на успеваемость обучающихся, является количество часов самостоятельной подготовки (study_hours). Можно предположить, что обучающиеся, уделяющие больше времени самостоятельному изучению материала, лучше усваивают учебный контент и демонстрируют более высокие итоговые результаты. Самостоятельная работа позволяет закреплять знания, повторять сложные темы и повышать уровень подготовки к контрольным мероприятиям.

Ещё одним важным фактором является посещаемость занятий (attendance_percentage). Регулярное присутствие на занятиях способствует более полному восприятию учебного материала, активному участию в образовательном процессе и своевременному получению обратной связи от преподавателей. Высокий уровень посещаемости, как правило, связан с более высокой академической успеваемостью.

Возраст обучающихся (age) также может влиять на результаты обучения. Различные возрастные группы могут отличаться уровнем мотивации, ответственностью и опытом обучения, что в совокупности способно отражаться на итоговом балле. Однако влияние данного фактора может быть менее выраженным по сравнению с учебной активностью студента.

Пол обучающегося (gender) рассматривается как социально-демографический фактор, который в ряде исследований связывается с различиями в стиле обучения и академической мотивации. Тем не менее его влияние на успеваемость не всегда является статистически значимым и требует проверки.

Тип образовательного учреждения (school_type) также может играть определённую роль в формировании учебных результатов. Различия в учебных программах, материально-техническом обеспечении и образовательной среде между государственными и частными школами могут отражаться на уровне подготовки школьников.

Уровень образования родителей (parent_education) традиционно рассматривается как один из факторов, влияющих на академические достижения обучающихся. Более высокий уровень образования родителей может способ-

ствовать формированию благоприятной учебной среды, поддержке учебной мотивации и развитию образовательных навыков у студентов.

Наличие доступа к интернету (internet_access) является важным условием для получения дополнительной информации, использования электронных образовательных ресурсов и самостоятельной подготовки. Однако при отсутствии мотивации данный фактор сам по себе не гарантирует повышения успеваемости.

Время, затрачиваемое на дорогу до учебного заведения (travel_time), также может влиять на учебные результаты. Длительная дорога может приводить к усталости и сокращению времени, доступного для самостоятельной подготовки, что потенциально отражается на итоговом балле.

Участие во внеучебных активностях (extra_activities) может оказывать двойственное влияние на успеваемость. С одной стороны, такие активности способствуют развитию навыков самоорганизации и социальной адаптации, с другой - могут уменьшать время, доступное для учебной деятельности.

Корреляционный и регрессионный анализ были проведены с использованием программных средств анализа данных. Результаты корреляционного и регрессионного анализа представлены на рисунке 3.

	<i>Коэффициенты</i>	<i>P-Значение</i>
Y-пересечение	-2,114729951	0
Переменная X 1	-0,005293724	0,004702809
Переменная X 2	0,002198024	0,573318622
Переменная X 3	-0,015729742	0,013731128
Переменная X 4	-0,003069086	0,102633796
Переменная X 5	0,543738411	1,19348E-05
Переменная X 6	0,0271266	0,000731128
Переменная X 7	0,009517163	0,285645414
Переменная X 8	0,00268001	0,350937493
Переменная X 9	-0,005298519	0,406484621

Рисунок 3 - Результаты регрессионного и корреляционного анализа

По заданным данным построена множественная линейная регрессия с зависимой переменной final_grade (средний балл по основным предметам).

По P-значениям определяем, какие факторы значимы и влияют на итоговый средний балл учащегося [2]. Сравним полученные P-значения со стандартным уровнем значимости $\alpha=0,05$. Если P-значение фактора меньше, чем 0,05, то этот фактор значим, то есть оказывает влияние на итоговую оценку. Значимые факторы выделены в таблице на рисунке 3 зеленым цветом. Наибольшее влияние на итоговую оценку обучающегося оказывают факторы:

- X1- возраст
- X3 - тип школы

- X5 - среднее время ежедневной самостоятельной учёбы,
- X6 - процент посещения занятий

Остальные факторы, которые не выделены цветом на рисунке 3, не оказывают влияния на итоговую оценку обучающегося:

- X2 - пол
- X4 - уровень образования родителей
- X7 - наличие доступа в интернет дома
- X8 - время в пути до учебы
- X9 - участие во внешкольных активностях

Также хочется отметить, что возраст имеет отрицательный коэффициент регрессии, следовательно, чем выше возраст обучающегося, тем ниже его итоговый балл. Тип школы тоже имеет отрицательный коэффициент регрессии, это говорит о том, что при обучении в частной школе успеваемость обучающегося будет ниже, чем в государственной образовательной организации.

Среднее время ежедневной самостоятельной учёбы и процент посещаемости занятий имеют положительный коэффициент регрессии, следовательно, чем больше времени обучающийся уделяет самостоятельной подготовке и посещению занятий, тем больше его средний итоговый балл.

Анализ данных свидетельствует о том, что среди рассмотренных факторов ключевое влияние на академическую успеваемость оказывают:

- Часы самостоятельной подготовки
- Регулярность посещения занятий
- Тип школы
- Возраст

В заключение можно сказать, что на итоговую оценку обучающегося влияют такие факторы как: возраст, тип школы, посещаемость занятий и часы самостоятельной подготовки. Остальные факторы, рассмотренные в этом анализе, не оказывают влияния на итоговую оценку. Практическое значение полученных результатов состоит в том, что для повышения успеваемости обучающегося важно правильно планировать время самостоятельной учёбы и стремиться к высокой посещаемости.

Список литературы

1. Набор данных об успеваемости учащихся [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/datasets/kundanbedmutha/student-performance-dataset/data>, (Дата обращения 15.11.2023), свободный.
2. Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман – М.: Высш. школа, 1979 [Электронный ресурс]. – Режим доступа: <http://elenagavrile.narod.ru/ms/gmurman.pdf>, свободный. (Дата обращения 15.11.2023) – с 394: Приложение 7.