

КЛАССИФИКАЦИЯ СТУДЕНТОВ ПО УРОВНЮ УСПЕВАЕМОСТИ С ПОМОЩЬЮ АППАРАТА ДИСКРИМИНАНТНОГО АНАЛИЗА

Статья посвящена разработке алгоритма выявления факторов, влияющих на успеваемость студента, в основе которого лежит математический инструментальный дискриминантный анализ, реализующего классификацию с «обучением» и позволяющего по значениям дискриминантных переменных отнести студента к классу с тем или иным уровнем успеваемости. Разработанная модель может быть полезной при формировании групп студентов-первокурсников, а также для повышения качества обучения путем воздействия на некоторые характеристики, которые влияют на дискриминацию.

Ключевые слова: дискриминантная модель, классифицирующая функция, обучающая выборка, качество распознавания, группировочный признак.

Коллектив Оренбургского колледжа статистики, экономики и информатики ставит основной целью создание образовательно-воспитательной среды, способствующей формированию и обучению высококвалифицированных специалистов в сфере СПО, способных, с одной стороны, к реализации полученных знаний в практической деятельности, а с другой – к приобретению и расширению полученных знаний, умеющих ориентироваться и адаптироваться в условиях рыночной экономики [1].

Для достижения этой цели образовательные системы с учетом профиля подготовки должны разрабатывать модель выпускника. Личностная составляющая модели специалиста обеспечивается самим учебным процессом, а также зависит от наследственности, семейного воспитания, состояния общества, наличия примеров для подражания [2].

Поэтому важно знать, какие факторы влияют положительно, а какие – отрицательно на формирование будущего специалиста. Понять, можно ли за счет устранения отрицательных воздействий на подготовку обучающегося улучшить уровень профессионализма молодого специалиста.

Другими словами, необходимо определить, какие факторные признаки могут влиять на успеваемость студентов и использовать эти факторы для классификации студентов по успеваемости, а также для повышения качества подготовки путем корректировки основных признаков.

Решить поставленные задачи позволяет метод дискриминантного анализа, в основе которого лежит понятие дискриминантной моде-

ли, которая оптимально разделяет множество объектов на подмножества и проводит классификацию новых объектов в тех случаях, когда неизвестно заранее, к какому из существующих классов они принадлежат.

Для классификации студентов Оренбургского колледжа статистики, экономики и информатики по успеваемости с помощью аппарата дискриминантного анализа проводилось четырехэтапное исследование.

На первом этапе проводился первичный отбор признаков на основе результатов анкетирования 114 студентов колледжа, обучающихся в выпускных группах.

Отбор признаков согласуется с известными исследованиями в области социологии, психологии и педагогики:

- исследования о негативном влиянии социальных сетей на успеваемость студентов [3];
- исследования о влиянии пола на успеваемость [4];
- исследование о воздействии личности родителей и их жизненной позиции на развивающуюся личность ребенка [5];
- исследование влияния здоровья растущего организма на гармоничное, соответствующие возрасту развитие [6];
- исследования о положительном влиянии чтения на мозг человека [7].

На успеваемость влияет так же уровень полученных знаний в школе. Здесь можно выделить два фактора:

- средний балл аттестата, характеризующий базовый уровень знаний, заложенный в школе;

– место проживания, т. е. тип населенного пункта.

Выбор последнего фактора обусловлен тем, что, как правило, в крупных городах имеется больше возможностей для получения качественного образования.

На втором этапе в ходе построения прогностической модели все студенты были разделены на три группы: учащиеся, у которых средний балл в колледже находится в интервале от 4,4 до 5,0, входят в группу А. Те, у кого средний балл успеваемости в колледже находится в интервале от 3,8 до 4,4 – в группу В и студенты, имеющие средний балл в интервале от 3,0 до 3,7 – в группу С.

Задача третьего этапа заключается в определении количества необходимых переменных модели. Данная задача решается путем процедуры последовательного отбора дискриминантных переменных.

В качестве группировочного признака был взят средний балл успеваемости в колледже. Для проведения классификации студентов по уровню успеваемости были отобраны следующие дискриминантные переменные:

X1 – средний балл аттестата;

X2 – пол (1 – женский пол, 2 – мужской пол);

X3 – прописка (1 – г. Оренбург, 2 – город Оренбургской области (не районный – центр), 3 – город, районный центр, 4 – село, поселок, районный центр, 5 – село, поселок (не районный центр));

X4 – количество детей в семье;

X5 – возраст матери, количество полных лет;

X6 – возраст отца, количество полных лет;

X7 – количество больничных листов в течение года;

X8 – время, проводимое в интернете в течение дня, часы;

X9 – семья (1 – один родитель (неполная

семья), 2 – два родителя (полная семья), 0 – нет родителей (сирота));

X10 – количество друзей;

X11 – количество прочитанных в месяц книг;

X12 – средний балл в колледже;

Средние арифметические значения и моды признаков, рассчитанные по обучающим выборкам, представлены в таблице 1. По полученным результатам можно дать интерпретацию группам.

Из данной таблицы следует, что студенты с более высоким средним баллом аттестата относятся к группам В и А, причем значения в группе А выше, чем в группе В. Если рассматривать признак X2 (пол), то можно увидеть, что у девушек успеваемость выше, чем у юношей. Относительно сведений о том, где учился студент: в городе, районном центре или в селе, можно сказать, что студенты с высоким средним баллом в колледже обучались в областном центре, а студенты, у которых балл ниже, обучались преимущественно в районных центрах и обычных сельских школах.

Данные о количестве детей в семье так же могут помочь отнести студента к группе с высоким средним баллом в колледже, т. е. чем больше в семье детей, тем ниже средний балл у студента. Данные о возрасте матери и отца являются значимыми, чем больше возраст матери и отца, тем лучше занимается ребенок. Студенты с низким средним баллом в колледже чаще всех отсутствуют на занятиях по болезни. Чем больше часов в день проводит студент в интернете, тем ниже его средний балл.

Что касается данных о характеристике семьи, заметна связь, что в полной семье дети учатся лучше, чем дети из неполных семей или дети-сироты. Данные о круге общения студента могут помочь отнести студента к группе успеваемости, чаще студенты с высоким средним баллом не имеют такого количества друзей, как студенты со средним и низким баллом.

Таблица 1. Средние арифметические значения признаков

Группа	Средние значения и моды признаков											Количество объектов
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	
С	3,82	2	3	2,03	44,37	52,17	1,77	3,13	2	4,83	0,07	30
В	4,09	2	2	1,52	43,93	52,36	0,59	2,23	2	3,89	0,68	56
А	4,42	1	1	1,14	45,46	52,64	0,21	1,39	2	3,36	1,46	28
Общее среднее	4,10	2	2	1,56	44,42	52,38	0,81	2,26	2	4,01	0,71	114

По количеству прочитанных в течение месяца книг можно точно отнести студента, который читает одну и более книг в месяц к группе с высоким средним баллом в колледже, а студенты, которые не читают книг вообще, попадают в группу с низким уровнем успеваемости.

Эти различия позволяют сделать вывод, что по рассматриваемым показателям можно классифицировать абитуриентов и таким образом предсказать, будет ли успешным их обучение.

На четвертом этапе были построены классифицирующие функции.

Объект относится к группе с наибольшим значением функции. Функция называется простой классифицирующей функцией, потому что предполагает лишь равенство групповых ковариационных матриц.

На основе функций, полученных на четвертом этапе, проводится повторная классификация объектов обучающих выборок, результаты которой представлены в таблице 2.

После повторной классификации к группе А относятся 23 студента из 28, а 5 остальных относятся к группе В. К группе В относятся 47 объектов, 3 отнесены к группе А, 6 отнесены к группе С. К группе С относятся 22 объекта, а 8 отнесены к группе В. Качество распознавания составило 80,7 %. Учитывая, что для анализа использованы реальные исходные данные, можно сделать вывод о том, что дискриминация объектов осуществляется достаточно точно.

На основе полученных обучающих выборок проводилась классификация студентов выпускной группы из 27 человек, подлежащих группировке. В результате классификации из 27 студентов неправильно классифицированы 3 студента. Если привести данные к процентному соотношению, то процент ошибок в данной классификации новых студентов будет равен 11,1%, соответственно процент правильной классификации составляет 88,89%, что говорит о доста-

Таблица 2. Результаты классификации объектов обучающих выборок

	Точность классификации	С p=.26316	В p=.49123	А p=.24561
С	73.33334	22	8	0
В	83.92857	6	47	3
А	82.14286	0	5	23
Суммарная точность классификации	80.70175	28	60	26

точно высоким качестве распознавания новых объектов на основе полученной в результате исследования модели. Ошибки в классификации могут быть вызваны нарушением многомерной нормальности распределения некоторых дискриминантных переменных, особенно это касается классификации объектов в «пограничных» зонах, где вероятности принадлежности объекта к двум или нескольким классам приблизительно равны.

Таким образом, в результате исследования были выявлены факторы, влияющие на успеваемость студентов. Ряд факторов, такие, как количество прочитанных в месяц книг, количество часов, проведенных в интернете, круг общения и склонность к заболеваниям, можно корректировать. В частности, укрепить здоровье или прекратить тратить время на общение в сети без необходимости. Это поможет студентам повысить качество знаний.

Данный метод может быть использован при формировании групп, что исключает возможность создания проблемных групп с преобладанием студентов, требующих особого внимания. Равномерное распределение студентов в группе с разным уровнем успеваемости положительно влияет на качество подготовки как отдельных студентов, так и группы в целом. Возможен и другой подход – формировать группы с разными прогнозируемыми уровнями успеваемости с дальнейшим дифференцированным подходом к обучению каждой группы.

08.06.2014

Список литературы:

1. Материалы официального сайта Оренбургского колледжа статистики, экономики и информатики [Электронный ресурс]. 2014. – Режим доступа: <http://www.oksei.ru>.
2. Лифинцева, М.В. Содержание и организация воспитательной работы в образовательных учреждениях в условиях интеграции социально-педагогической среды (на примере медицинского колледжа): дис. канд. пед. наук : 13.00.01 / М. В. Лифинцева. – Ставрополь, 2004. – С. 89–90.
3. Kirschner, P. A., Karpinski, A. C. Facebook® and Academic Performance /Journal Computers in Human Behavior, 2010 – volume 26, № 6. – pages 1237–1245

4. Луцкова, Л.Н. Исследование факторов, влияющих на учебную мотивацию студентов медицинского вуза. [Электронный ресурс] / Л.Н. Луцкова, Н.А. Русина // Медицинская психология в России: электрон. науч. журн. 2012. N 1. Режим доступа: [http:// medpsy.ru](http://medpsy.ru).
5. Овчарова, Р.В. Родительство как психологический феномен : учеб. пособие / Р.В. Овчарова. – Москва : Московский психолого-социальный институт, 2006. – С. 63–84. – ISBN 5-89502-771-7.
6. Тищенко, В.А. Влияние физического развития студентов на их профессиональную подготовку / В.А. Тищенко // Физическое воспитание студентов, 2012. – № 02. – С. 96–100.
7. MedDaily – «МедДэйли» новости медицины [Электронный ресурс] // Проект объединенной редакции NEWSru.com– Москва : MEDDAILY.RU, 2010–2014. – Режим доступа: [http:// meddaily.ru](http://meddaily.ru). – 05.03.2014.

Сведения об авторе:

Панова Наталья Федоровна, старший преподаватель Оренбургского государственного университета

Денисова Нина Владимировна, студентка Оренбургского государственного университета

460018, г. Оренбург, пр-т Победы, 13, ауд. 6309, e-mail: panovan@yandex.ru