

ФОРМИРОВАНИЕ ГРУППОВОЙ И ИНДИВИДУАЛЬНОЙ ТРАЕКТОРИЙ УСПЕВАЕМОСТИ ПО ДАННЫМ E-LEARNING-ПЛАТФОРМ

Владова А. Ю.¹

(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва, Финансовый
университет при Правительстве РФ, Москва)

Поддержание высокого уровня образования является одной из основных задач управления вузом. Несмотря на постоянный мониторинг успеваемости учащихся, менеджмент образовательных учреждений недостаточно использует методы прогноза при формировании траекторий успеваемости. Предлагаемый подход отличается от известных тем, что анализируют признаки, содержащие оценки за различные работы, выполненные на e-learning-платформе, изменяют размерность пространства признаков за счет нормализации оценок по единой шкале, а также создания новых динамических признаков и признаков, содержащих решение задачи классификации учащихся, находящихся в группе академического риска и задачи прогнозирования экзаменационных оценок. Далее по дополненным данным, выделяют группы учащихся со схожими траекториями успеваемости для индивидуализации консультаций. Таким образом, формирование групповой и индивидуальной траекторий успеваемости с учетом результатов прогноза направлено на упреждающее повышение уровня академической успеваемости учащихся в университетских условиях.

Ключевые слова: академическая успеваемость, статистический анализ, линейный дискриминантный анализ, регрессия, траектории обучения, дистанционная платформа.

1. Введение

Управление образовательным учреждением учитывает успеваемость учащихся как одну из важнейших характеристик, по которой можно судить о достигнутых результатах или существующих проблемах. Успеваемость, как правило, мониторят с помощью множества аудиторных, самостоятельных, контрольных и курсовых работ, ответах на семинарах. Часть этих активностей

¹ Алла Юрьевна Владова, д.т.н., в.н.с. (avladova@ipu.ru).

часто автоматизирована с помощью e-learning-платформ [7]. Результаты мониторинга успеваемости используют для управления учебным процессом, при принятии решений о присуждении государственных академических и именных стипендий, выдаче дипломов с отличием и других задачах.

2. Анализ литературы

Оценка успеваемости является одним из важнейших компонентов электронного обучения. Авторы исследования [12] оценили успеваемость в области медицинского образования через показатели получения и восприятия знаний учащимися, уровень уверенности, удобство пользования e-learning-платформой и готовность рекомендовать электронное обучение. Изъян предлагаемого подхода заключается в качественном характере анализируемых признаков, их сильной зависимости от мнения разных экспертов. Исследователи [10] дополнили набор количественными признаками, которые позволяет собирать e-learning-платформа, а именно: набор онлайн-действий студента, время, необходимое, чтобы ответить на вопрос или отправить задание, пропущенные студентами вопросы. Кроме того, исследователи отдельно мониторили деструктивное поведение, которое включает низкую вовлеченность, чрезмерные опоздания, высокий процент отсева, списывание на заданиях и тестах, низкую эффективность обучения, уничижительные комментарии в онлайн-дискуссиях или по электронной почте. Недостатком данного подхода можно считать то, что данные брались без дополнительных преобразований, влияющих на производительность модели.

Исследование [14] предлагает модель прогнозирования эффективности онлайн-обучения, основанную на слиянии и выборе признаков. Модель использует взаимосвязь между признаками и объединяет учащихся в соответствии с категорией, которая достигла лучшей эффективности. Авторы [3, 4] в результате проведенной разведки данных для выявления скрытых взаимосвязей создали наборы синтетических признаков, позволяющих повысить эффективность моделей прогнозирования параметров организационных систем.

Авторы статьи [6] выделяют линейный дискриминантный анализ при вычислении вероятностного профиля принадлежности к целевым кластерам, на основе этого строится итоговое заключение.

Авторы [11] отмечают, что один из методов определения эффективности обучения предполагает использование линейной регрессии. Для этого они строят регрессионную модель по трем признакам (тест оценки знаний, активность в ходе курса и итоговый экзамен) и анализируют ее параметры. Перед построением регрессионной модели авторы изучили корреляционную матрицу, чтобы определить, насколько тесно связаны признаки. Минусы подхода заключаются в небольшой размерности пространства признаков и отсутствии учета динамики успеваемости.

Авторы [8] используют алгоритм кластеризации k -means, который выявляет учащихся с низкой вовлеченностью. Количество кластеров задано априори и недостаточно отражает учащихся с удовлетворительным уровнем взаимодействия. Исследователи отмечают [13], что кластеризация пользователей в группы с общими интересами очень полезна при изучении нескольких языков. Они и авторы [2] также используют алгоритм k -means из-за его простоты. В то же время известен метод DBScan [9], чья способность приспосабливаться к различным распределениям данных и идентифицировать кластеры различных форм и размеров делает его универсальным и мощным алгоритмом для выявления сложных закономерностей и структур в наборах данных.

Для выявления деструктивного поведения клиентов, которые недовольны предоставляемыми услугами, авторы статьи [1] рассмотрели различные методы классификации, включая деревья принятия решений, наивный байесовский классификатор, случайный лес, искусственные нейронные сети, линейный дискриминантный анализ, метод опорных векторов и их ансамбли. Эффективность классификаторов оценена с помощью таких показателей, как площадь под кривой AUC, точность, F1-мера и время классификации. Авторы указывают, что для четырех исследованных датасетов (после применения метода главных компонент для уменьшения пространства признаков) получены фактически идентичные оценки F1, точности и AUC.

Таким образом, установлено, что при построении моделей успеваемости авторы недостаточно учитывают динамику изменения показателей успеваемости, минимально используют преобразования имеющихся данных при повышении точности моделей, а также отмечено отставание в применении современных классификационных моделей при формировании групповой и индивидуальной траекторий успеваемости. Основное противоречие заключается в том, что при формировании индивидуальных траекторий успеваемости учащихся не учитывают результаты прогноза и групповые траектории.

В результате анализа литературы сформулированы новизна подхода и практическая применимость.

Новизна формирования групповой и индивидуальной траекторий успеваемости учащихся отличается от известных тем, что на основе значительного количества независимых от эксперта разновременных оценок генерируют динамические признаки, далее формируют дополнительные признаки решением задачи классификации с выделением группы учащихся под академическим риском и задачи регрессии с предсказанием индивидуальной траектории успеваемости, а в заключение решают задачу кластеризации по поиску учащихся со схожими траекториями успеваемости с неизвестным априори количеством кластеров.

Практическая применимость подхода сформулирована следующим образом:

- высокая предсказательная способность на основе исторических данных обучения;
- прогнозы и факторы, влияющие на эти прогнозы, должны быть понятны для преподавателей и администраторов, принимающих решения;
- результаты прогнозирования по модели не должны сильно изменяться на незначительных изменениях входных данных;
- отсутствие предвзятости и дискриминации по отношению к различным группам учащихся;
- адаптируемость к изменяющимся потребностям образовательной среды.

3. Статистический анализ данных мониторинга успеваемости

3.1. ОПИСАНИЕ ДАННЫХ

Исходный датасет содержит оценки по разным элементам дисциплины «Анализ данных», прочитанной двумя преподавателями. Баллы за тесты с автоматической проверкой сформированы в том числе по произвольным, созданным преподавателем, шкалам. Датасет с промежуточными и итоговыми баллами учащихся за первый семестр содержит 20 признаков: два текстовых признака с фамилиями и именами; номер группы; и две группы количественных признаков с оценками во время первой и второй половины семестра за домашние задания, контрольные, проекты и тесты, размещенные на платформе электронного обучения. Остальные количественные признаки содержат баллы, выставленные преподавателями за навыки программирования, активность на семинарах, а также итоговые баллы. Для визуализации количественных признаков использована диаграмма «ящик с усами» (рис. 1).

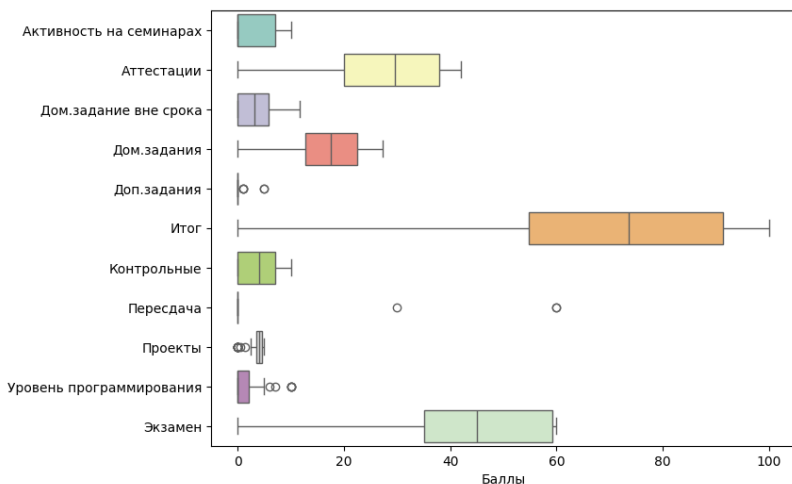


Рис. 1. Сравнение баллов по различным видам работ в семестре

Диаграмма содержит набор ящиков с усами, отображающих распределение баллов учащихся по различным видам работ. Каждый ящик представляет один вид работы и показывает медиану, верхний и нижний квартили, а также наличие выбросов в баллах. Таким образом, диаграмма позволила увидеть разные оценочные шкалы для разных видов работ, незначительное количество учащихся, выполнивших дополнительные задания или пересдавших экзамен. Медиана экзаменационных баллов равна 45, а итоговых баллов – 73,5.

3.2. ПОСТАНОВКА ЗАДАЧИ

Математическая постановка задачи может быть выражена следующим образом: пусть n – количество учащихся, а m – количество признаков успеваемости каждого учащегося. Тогда для каждого i -го учащегося имеем вектор признаков успеваемости $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$, где x_{ij} представляет собой значение признака. Увеличение размерности пространства признаков происходит за счет нормализации существующих и создания новых признаков, расширяющих пространство $x_i \in R^d$, где $d > m$.

Для выделения учащихся в группу академического риска обозначим бинарную переменную принадлежности i -го студента к группе академического риска (1 – находится в группе академического риска, 0 – не находится) как y_i . Тогда средние значения признаков для каждой группы обозначим μ_0 и μ_1 . Направление оптимальной гиперплоскости, разделяющей две группы учащихся по академическому риску, задаем с помощью критерия линейного дискриминантного анализа:

- (1) $w = S_w^{-1} \Delta\mu$,
- (2) $S_w = cov(x_1) + cov(x_0)$,
- (3) $\Delta\mu = \mu_0 - \mu_1$,

где w – вектор весов, который максимизирует отношение двух классов, S_w – обобщенная ковариационная матрица.

Линейный дискриминантный анализ является более устойчивым к выбросам, чем логистическая регрессия, и достигает хорошей точности даже при малых размерах выборки. Но этот метод требует подготовки данных следующим образом: результи-

рующий признак должен быть категориальным; входные признаки должны быть распределены близко к стандартному нормальному распределению, чтобы их дисперсии были близки.

Прогнозирование итогового балла за семестр проведем с помощью линейно-регрессионной модели, где для каждого i -го студента строим модель успеваемости:

$$(4) \quad y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_d x_{id} + e_i,$$

где y_i – итоговый балл за семестр; $\alpha_0, \alpha_1 x_{i1}, \alpha_2 x_{i2}, \dots, \alpha_d$ – коэффициенты регрессии; e_i – случайная ошибка.

Для поиска учащихся со схожими траекториями успеваемости при тех же вводных (n – количество учащихся, m – количество признаков для каждого студента, тогда X – матрица оценок учащихся размерности $n \times m$, где каждая строка представляет собой набор оценок для одного студента) найти k групп учащихся, объединенных вместе на основании схожести их оценок, с целью формирования групповых траекторий успеваемости.

Таким образом, математическая постановка задачи включает в себя увеличение размерности пространства признаков, построение модели для прогнозирования успеваемости и последующую индивидуализацию консультаций на основе схожести траекторий успеваемости учащихся.

4. Реализация

4.1. ОБЪЕДИНЕНИЕ ДАННЫХ ПО ГРУППАМ УЧАЩИХСЯ

Набор оценок на обучающей платформе формируется для каждой группы учащихся. Поэтому реализовано слияние наборов оценок разных групп учащихся в единый датасет. После слияния отсутствующие значения заменены на ноль. Этот шаг обеспечил согласованность данных и позволил проводить дальнейший анализ без сбоев, вызванных отсутствием данных.

Процесс слияния и очистки данных закладывает основу для проведения всестороннего анализа и получения значимой информации путем ее консолидации из нескольких источников.

4.2 АНАЛИЗ ПРОПУСКОВ И НУЛЕЙ

Пропуски в данных связаны с отсутствием у студента баллов за активность. Причиной отсутствия баллов может быть пропуск

контрольной по уважительной или неуважительной причине, отсутствие ответов на семинарах, невыполнение курсовой работы, а также непроведение какого-то типа активности преподавателем в данной группе. Визуализация нулевых значений признаков представлена на рис. 2.

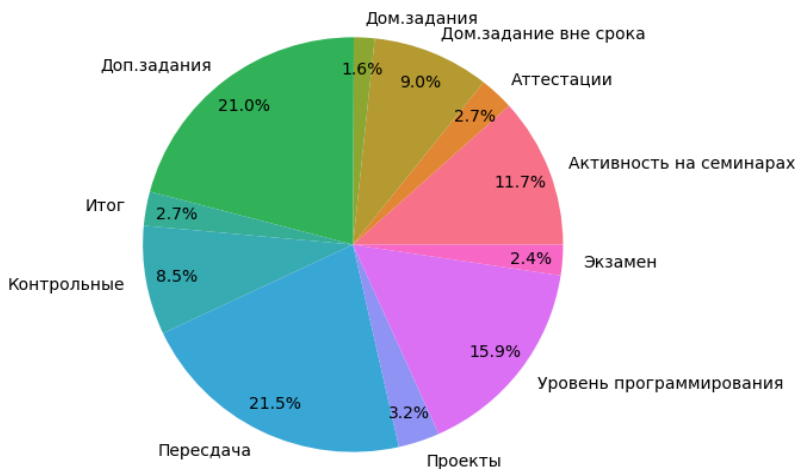


Рис. 2. Визуализация нулевых значений

Признаками со значительным количеством нулей являются выполнение дополнительных заданий, пересдача экзамена, уровень программирования и активность на семинарах.

4.3 ИЗМЕНЕНИЕ ПРОСТРАНСТВА ПРИЗНАКОВ

Разница в производительности методов машинного обучения напрямую зависит от предварительной обработки входных данных. Поэтому для повышения производительности модели используют различные методы проектирования признаков [4]. В работе на основе имен и окончаний фамилий добавлен бинарный атрибут Пол. Для анонимизации учащихся введен признак под названием Index, состоящий из первых букв фамилии, имени, пола и номера группы учащегося. Установлено, что баллы за тестовые, проектные и домашние задания варьируются от 0 до 5,

баллы за промежуточные оценки варьируются от 0 до 22, экзаменационные баллы – от 0 до 60, итоговые баллы – от 0 до 100. Таким образом, все баллы нормированы в диапазон от 0 до 1 делением на соответствующие максимальные значения. Это преобразование приводит к появлению нормированных признаков, которые характеризуют академическую успеваемость по шкале от 0 до 1 (рис. 3).

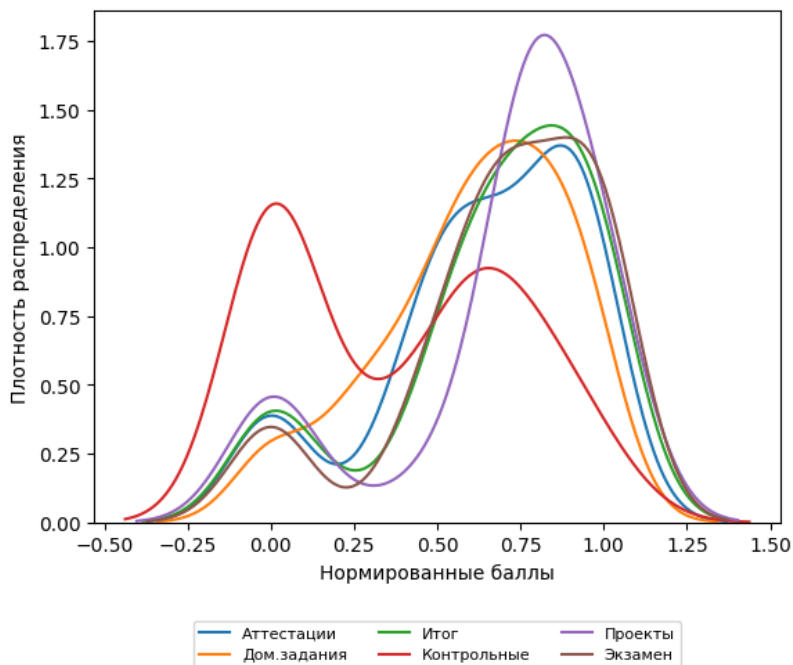


Рис. 3. Распределения нормированных признаков

Более низкая (по сравнению с левой) правая вершина распределения баллов за контрольные работы говорит о том, что успеваемость большинства учащихся по этой активности недостаточно хороша.

Для учета динамики успеваемости по одноименным активностям вычислены разницы между нормированными баллами

во второй и первой половине семестра. Распределения значений полученных признаков представлены на рис. 4.

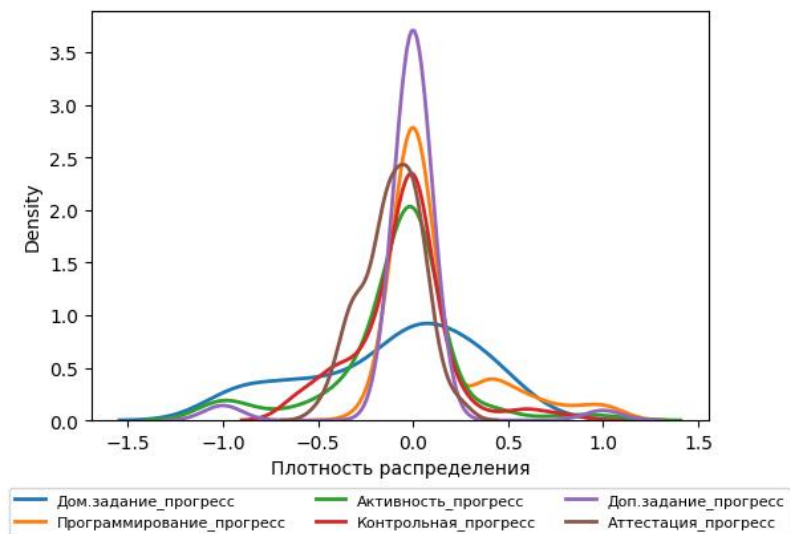


Рис. 4. Распределения признаков, учитывающих динамику успеваемости

Характер распределений новых признаков стал ближе к нормальному с незначительным смещением математического ожидания в отрицательную область у всех признаков кроме уровня программирования. Наличие отрицательных значений свидетельствует о снижении баллов у ряда учащихся во второй половине семестра. Нормализация распределений признаков улучшает сходимость алгоритмов, особенно когда диапазоны значений исходных признаков значительно отличаются.

Использование динамических признаков улучшило интерпретируемость модели за счет уменьшения числа входных признаков и более точном определении значимости каждого признака для классификации. Введение зашумленных признаков улучшило обобщающую способность модели.

4.4. ВЫДЕЛЕНИЕ ГРУППЫ УЧАЩИХСЯ ПОД АКАДЕМИЧЕСКИМ РИСКОМ ПОД АКАДЕМИЧЕСКИМ РИСКОМ

Для выделения группы учащихся под академическим риском использован линейный дискриминантный анализ. Метод создает линейную комбинацию признаков, которая максимально разделяет два класса учащихся, позволяя классифицировать новых студентов как тех, кто скорее всего сдаст экзамен, и тех, кто, возможно, не сдаст его. При подготовке данных целевая переменная преобразована в бинарную.

Полученный набор данных разделен в пропорции 80/20 на обучающий и тестовый. Причем в обучающий и в тестовый наборы попали студенты, не сдавшие экзамен. Для избежания переобучения и улучшения обобщающей способности модели с помощью библиотеки FunctionTransformer значения признаков обучающего набора данных изменены добавлением нормально распределенного случайного шума [5]. Диаграмма рис. 5 показывает желтыми треугольниками обученные точки и сравнивает местоположение имеющихся тестовых точек в виде серых прямоугольников и прогнозируемых точек в виде зеленых треугольников.

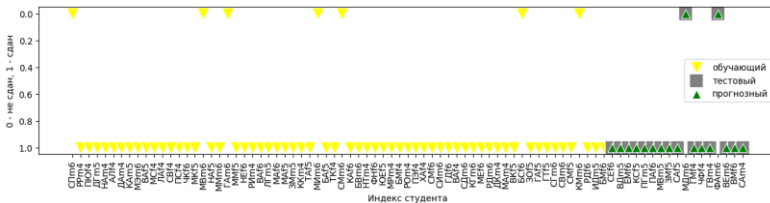


Рис. 5. Классификация учащихся на сдавших и не сдавших экзамен

Результаты классификации экзаменационных оценок выглядят следующим образом:

- константа дискриминации $C = 9,241$;
- уравнение дискриминационной функции: $D(x) = -5,908x_1 - 2,163x_2 - 4,470x_3 + 0,893x_4 + 1,912x_5 - 0,237x_6 + 0,174x_7 + C$, где x_1, \dots, x_7 – признаки, расположенные по возрастанию влияния, такие как группа учащегося (что говорит о неоднородности групп,

а также влиянии экспертных оценок), приращение оценок по домашним заданиям, пол, приращение активности на семинарах и дополнительных заданиях и др.;

- среднее гармоническое значение точности и полноты, представленное оценкой $F1 = 92,86\%$, означает, что модель способна хорошо разделять классы, дает низкую долю ложноположительных предсказаний и может правильно классифицировать большинство положительных и отрицательных случаев;

- оценка среднего гармонического значения точности и полноты для исходных данных (без динамических признаков) составляет $83,61\%$.

Эти результаты отражают потенциальную эффективность модели при двухклассовой классификации успеваемости.

4.5. ПРЕДСКАЗАНИЕ ИНДИВИДУАЛЬНОЙ ТРАЕКТОРИИ УСПЕВАЕМОСТИ

Нормализованные признаки разбиты на обучающую (80%) и тестовую (20%) части. Экземпляр модели линейной регрессии обучается на большей выборке и делает прогнозы для тестовой части. Результат прогнозирования части экзаменационных оценок визуализирован с помощью точечной диаграммы (рис. 5).

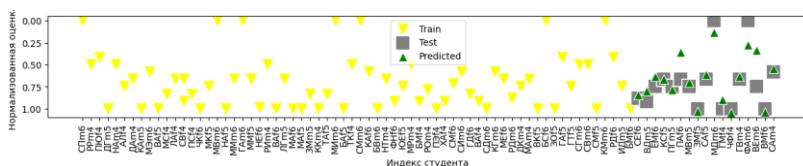


Рис. 5. Прогноз результатов экзамена

Для оценки работоспособности модели линейной регрессии на тестовых данных рассчитаны метрики ошибок: $MAE = 0,108$, $MSE = 0,024$, $RMSE = 0,156$, $R^2 = 0,712$. Модель гребневой регрессии снизила значения ошибок: $MAE = 0,099$, $MSE = 0,023$, $RMSE = 0,102$, $R^2 = 0,727$. Для имеющегося набора данных модель показывает высокую точность, поскольку известные тестовые и прогнозируемые значения достаточно близки, а ошибки не превышают 10%.

Индивидуальные траектории учащихся с прогнозом следующей оценки представлены на рис. 6.

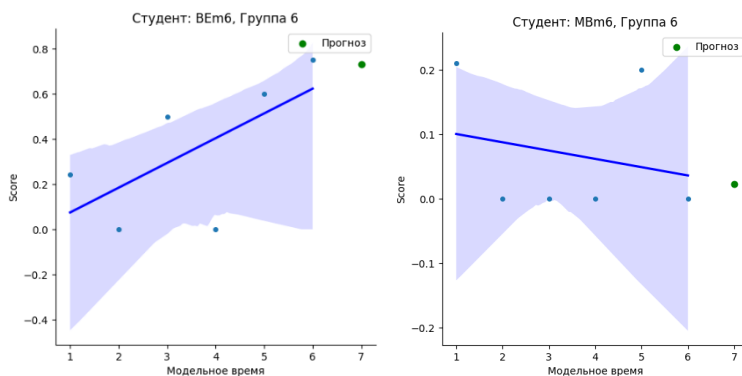


Рис. 6. Прогноз индивидуальных траекторий успеваемости:
а) растущий тренд; б) снижающийся

Графики на рис. 5 демонстрируют разный подход к обучению у двух учащихся одной группы. Второй учащийся нуждается в дополнительных консультациях.

4.6. ФОРМИРОВАНИЕ ГРУППОВЫХ ТРАЕКТОРИЙ УСПЕВАЕМОСТИ

При создании персонализированных учебных планов, выявлении успешных/неуспешных стратегий обучения удобно выделить группы учащихся со схожими траекториями успеваемости. Для этого использован метод DBScan на том же размеченном наборе данных. Метод кластеризует данные по различной плотности и отличает плотные регионы от разреженных областей, обрабатывая выбросы и шум. Результаты кластеризации представлены на рис. 7.

DBScan делит данные на три категории: основные точки, которые имеют достаточное количество соседних точек данных в пределах заданного радиуса эpsilon (ϵ); граничные точки, которые находятся в пределах ϵ -радиуса от основной точки, но не имеют достаточного количества соседей, чтобы считаться основными точками; и шумовые точки, которые не принадлежат

ни к одному кластеру из-за их изолированности и недостаточной близости к другим точкам.

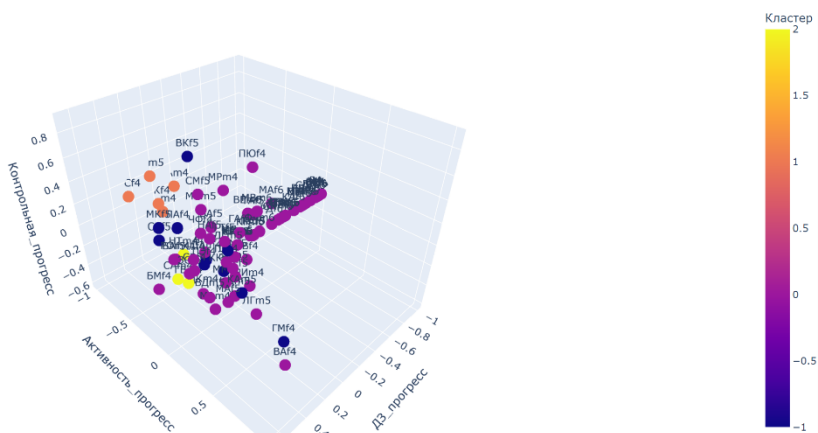


Рис. 7. Кластеризация учащихся по траекториям успеваемости

После применения DBSCAN для кластеризации данных, получены следующие результаты по количеству точек в каждом кластере: кластер -1 : 10 точек, кластер 0 : 66 точек, кластер 1 : 5 точек, кластер 2 : 3 точки. Эти результаты могут быть объяснены следующим образом: кластер -1 содержит шумовые точки. Значимые кластеры 0 , 1 и 2 вместе содержат 74 точки. Точки в этих кластерах характеризуются общим поведением.

Таким образом, кластеры определены на основе локальной плотности точек и пространственных отношений, что преодолевает традиционные ограничения кластеризации.

5. Выводы

Обзор литературы посвящен прогнозированию траекторий успеваемости учащихся с использованием методов машинного обучения. Недостатками практической реализации предлагаемых методов являются значительное количество качественных признаков при небольшой размерности пространства признаков, сильное влияние экспертов, недостаточный учет динамики успеваемости.

Проведен статистический анализ признаков e-learning-платформы, включающий оценку дисбаланса данных, выявление пропусков, построение плотности распределений и изучение корреляционной матрицы.

Предлагаемый подход включает 4 этапа: вычисление приращений разновременных одноименных признаков и нормализация набора данных, формирование рискованной группы учащихся, которые могут не сдать предмет, прогноз оценки по предмету для каждого учащегося, формирование множества групп учащихся с близкими компетенциями.

Для повышения точности прогнозных моделей изменена размерность пространства признаков за счет приведения баллов за разные активности к общей шкале и создания новых признаков, таких как индексы учащихся и приращение баллов по разным активностям в процессе обучения.

При решении задачи кластеризации для поиска учащихся со схожими траекториями успеваемости, в которой количество кластеров заранее неизвестно, в качестве дополнительных признаков использованы результаты решения задачи классификации – выделения учащихся под академическим риском – и решения задачи прогнозирования – расчет итоговой оценки студента.

Результаты прогнозирования по модели на зашумленных данных и с использованием дополнительно сгенерированных данных изменились в пределах 7%.

Вопрос адаптируемости предложенного четырехэтапного подхода к изменяющимся потребностям образовательной среды требует дальнейшего исследования.

Упреждающее выявление и поддержка учащихся с риском неуспеваемости может помочь учебным заведениям адаптировать мероприятия и поддержку для учащихся, которые больше всего в этом нуждаются.

Признавая разнообразные потребности студенческих групп и соответствующим образом адаптируя поддержку, исследование направлено на создание более персонализированной и эффективной учебной среды.

Примеры данных и кода представлены на странице <https://github.com/avladova/Student-performance-prediction>.

Литература

1. АРЗАМАСЦЕВ С. А., БГАТОВ М., СЕМЕНЧИКОВ Д. *Предсказание оттока абонентов: сравнение методов машинного обучения* // Computer Tools in Education. – 2018. – №5. – С. 5–23.
2. ВЛАДОВА А.Ю. *Кластерный анализ изменений пространственного положения трубных секций магистрального нефтепровода по данным внутритрубных обследований* // Безопасность труда в промышленности. – 2018. – №1. – С. 22–25.
3. ВЛАДОВА А.Ю. *Формирование пространства признаков и авторегрессионных моделей для прогноза отступлений железнодорожного полотна* // Проблемы управления. – 2023. – Вып. 2. – С. 54–64.
4. ВЛАДОВА А.Ю., ШЕК Е.Д. *Подготовка данных для машинного анализа ключевых показателей эффективности территориальных менеджеров* // Бизнес-информатика. – 2021. – №3. – С. 48–59.
5. ЗАДАДАЕВ С.А., ОРЛОВА И.В. *Опыт применения эконометрического инструментария для прогнозирования показателей национальных целей развития РФ* // Фундаментальные исследования. – 2022. – №10-1. – С. 54–59. – DOI: 10.17513/ft.43343.
6. КУРАВСКИЙ Л.С., ЮРЬЕВ Г.А., ЗЛАТОМРЕЖЕВ В.И. и др. *Оценка действий экипажа воздушного судна на основе модели рисков человеческого фактора* // Экспериментальная психология. – 2020. – Т. 13, №2. – С. 153–181. – DOI: 10.17759/exppsy.2020130211.
7. ВАКНТАДЗЕ N., ЗАЙКИН O., ЗЫЛАНСКИ A. и др. *Incentive Model of a Project Learning Process* // 7th Int. Conf. on Frontiers of Industrial Engineering, ICFIE-2020: Virtual, Singapore, 2020. – P. 73–81. – DOI: 10.1109/ICFIE50845.2020.9266738.
8. БАЯЗИТ А., ИЛГАЗ Н., ЕРДЕН С. *Profiling students via clustering in a flipped clinical skills course using learning analytics* // Medical Teacher. – 2022. – №7(45). – С. 724–731.

9. KIM J., CHO J. *DBSCAN-Based Tracklet Association Annealer for Advanced Multi-Object Tracking* // *Sensors*. – 2021. – Vol. 21. – P. 5715.
10. LIU M., YU D. *Towards intelligent E-learning systems* // *Education and Information Technologies*. – 2023. – No. 7(28). – P. 7845–7876.
11. NAGY V., DUMA L. *Measuring efficiency and effectiveness of knowledge transfer in e-learning* // *Heliyon*. – 2023. – No. 7(9).
12. OLUWADELE D., SINGH Y., ADELIYI T. *E-Learning Performance Evaluation in Medical Education – A Bibliometric and Visualization Analysis* // *Healthcare*. – 2023. – No. 11. – P. 232.
13. TROUSSAS C., VIRVOU M., ALEPIS E. *Comulang: towards a collaborative e-learning system that supports student group modeling* // *SpringerPlus*. – 2013. – No. 1(2). – P. 387.
14. YE M., SHENG X. *SA-FEM: Combined Feature Selection and Feature Fusion for Students' Performance Prediction* // *Sensors*. – 2022. – No. 22(22). – P. 8838.

DEVELOPING GROUP AND INDIVIDUAL PERFORMANCE PATHS BASED ON E-LEARNING PLATFORM DATA

Alla Vladova, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Financial University under the Government of Russian Federation, Dr.Sc. (avladova@ipu.ru).

Abstract: Maintaining a high level of education is a key task in university management. Despite continuous monitoring of student performance, educational institution management fails to adequately utilize performance forecasting methods when shaping student learning paths. The proposed approach differs from existing ones in several aspects. Firstly, it analyzes features containing grades for various assignments completed by students on the e-learning platform, expanding the feature space by normalizing grades on a single scale and creating new features: an index and changes in performance for different types of assignments. Secondly, it identifies students at academic risk. Thirdly, it predicts exam scores for each student using a linear regression model. Fourthly, it groups students with similar learning trajectories for personalized consultations. The approach to predicting exam results for individual students demonstrates a commitment to providing comprehensive support beyond simple assessment. Through analysis, modeling, and personalized consultations, the research aims to proactively enhance academic performance in university settings.

Keywords: academic performance, statistical analysis, linear discriminant analysis, regression, learning trajectories, distant learning system.

УДК 65.012.4 + 681

ББК 65.9

DOI: 10.25728/ubs.2024.111.07

*Статья представлена к публикации
членом редакционной коллегии Я.И. Квинто.*

Поступила в редакцию 27.03.2024.

Опубликована 30.09.2024.