

Копытова Ирина Владимировна, студентка,  
Кубанский государственный университет,  
г. Краснодар

## МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ УСПЕВАЕМОСТИ УЧАЩИХСЯ

**Аннотация.** Прогноз успеваемости учащихся посвящен применению методов интеллектуального анализа образовательных данных с целью прогнозирования успеваемости студентов. Эта проблема является основной задачей персонализированного образования и привлекает все большее внимание в области интеллектуального анализа данных с целью оптимизации образовательных процессов, улучшения эффективности обучения и повышения общего успеха студентов

**Ключевые слова:** интеллектуальный анализ данных, машинное обучение, анализ текстовых данных, кластерный анализ, образовательные данные, прогноз успеваемости учащихся.

В современном образовательном контексте, где данные становятся все более доступными и объемными, внимание к интеллектуальному анализу образовательных данных растет. В данном исследовании внимание направлено на использование методов интеллектуального анализа для улучшения прогнозирования успеваемости студентов. Отслеживание и анализ различных аспектов академической деятельности, включая оценки, активность на платформах для обучения и анкетные данные, предоставляет возможность выявления скрытых закономерностей и позволяет создать более точные модели предсказания успеха.

История исследований в области анализа данных отражает постоянный прогресс в понимании и оптимизации образовательных процессов. Начиная с ранних работ по статистическому анализу успеваемости, современные исследования все больше ориентированы на использование передовых методов для прогнозирования академических результатов.

Существующие методы прогнозирования успеваемости включают как традиционные статистические подходы, так и инновационные методы машинного обучения. Статистические модели, такие как регрессионный анализ, давно использовались для анализа взаимосвязей между различными факторами и успехом студентов. Однако, в последнее десятилетие, с развитием машинного обучения, возможности прогнозирования были значительно расширены.

Для использования различных методов интеллектуального анализа, используется обширный и многоуровневый набор данных, который охватывает разнообразные аспекты студенческой активности и успеваемости.

Источниками данных являются:

1. Электронные журналы: информация об оценках, результаты тестов и заданий, а также динамика успеваемости фиксируются в электронных журналах. Эти данные предоставляют количественные показатели успеха студентов в различных дисциплинах.

2. Платформы для обучения: данные об активности студентов на образовательных платформах включают в себя частоту входа в систему, время, проведенное на платформе, использование различных образовательных ресурсов и т.д. Эта информация предоставляет контекст и позволяет выявить образовательные привычки студентов.



3. Анкетные данные: дополнительные анкетные данные включают в себя информацию о предыдущем образовании, уровне мотивации, и предпочтениях студентов. Эти данные добавляют качественную размерность к анализу, позволяя учесть индивидуальные особенности.

4. Временные ряды успеваемости: для каждого студента фиксируется динамика успеваемости в течение времени, что позволяет проводить анализ изменений и выявлять тренды в академической деятельности.

Данные представлены в формате таблиц, где каждая строка соответствует отдельному студенту, а столбцы содержат информацию о различных параметрах и показателях. Каждый набор данных содержит уникальные идентификаторы студентов для обеспечения конфиденциальности и четкой идентификации.

Выбор методов анализа данных важен для достижения целей исследования, которые включают прогнозирование успеваемости студентов и выявление влияющих факторов. Применяются разнообразные методы, охватывающие различные аспекты студенческой активности:

#### 1. Машинное обучение.

На этапе применения методов машинного обучения фокусируются на использовании различных моделей для прогнозирования успеваемости студентов. Выбор моделей машинного обучения зависит от характера данных и целей исследования. Можно экспериментировать с различными типами моделей, включая Support Vector Machines (SVM), Decision Trees, и Neural Networks. SVM эффективны в работе с разделяющими данными, Decision Trees позволяют идентифицировать важные признаки, а Neural Networks способны моделировать сложные нелинейные зависимости [1].

Обучение моделей включает в себя подготовку данных, разделение их на тренировочный и тестовый наборы, и настройку параметров моделей. Используются стандартные процедуры, такие как методы стохастического градиентного спуска или обратного распространения ошибки.

Оценка эффективности моделей является критическим этапом. Применяются методы кросс-валидации для избежания переобучения моделей и обеспечения их обобщающей способности. Точность, полнота, F-мера и кривые ROC-AUC используются для качественной оценки результатов.

Результаты моделей анализируются с учетом различных метрик и кривых. Выявляются ключевые признаки, влияющие на успешное предсказание успеваемости, а также оценивается вклад каждой модели в общую предсказательную силу. Этот анализ позволяет лучше понять, какие факторы существенны для точного прогнозирования успеваемости студентов.

Применение методов машинного обучения в этом контексте не только позволяет создать эффективные модели прогнозирования, но и выявить сложные взаимосвязи в данных, которые могут быть ценными для понимания академического успеха студентов.

#### 2. Анализ текстовых данных.

Этот анализ включает в себя обработку эссе студентов, комментарии преподавателей и другие текстовые материалы, предоставляя более глубокое понимание факторов, не всегда выраженных количественными показателями.

Прежде чем анализировать текстовые данные, проводится предварительная обработка, включающая в себя удаление стоп-слов и нормализацию текста. Эти шаги не только сокращают размерность данных, но и улучшают качество анализа, фокусируясь на существенных элементах текста.

Проводятся методы анализа тональности для выявления эмоциональной окраски текстов. Это позволяет определить не только содержание текста, но и то, как студенты



относятся к учебным материалам, своей активности и результатам. Эмоциональные аспекты могут сыграть ключевую роль в понимании индивидуальных подходов к обучению [2].

Применение методов обработки естественного языка также включает в себя выделение ключевых слов и выявление тематических акцентов в текстах. Это помогает выявить основные концепции, с которыми студенты взаимодействуют, и определить, на каких аспектах учебного материала стоит сосредотачиваться.

Кроме анализа студенческих эссе, также рассматриваются комментарии преподавателей к академической деятельности студентов. Эти комментарии могут содержать ценные наблюдения и рекомендации, которые дополняют количественные данные.

Результаты анализа текстовых данных интегрируются с результатами количественного анализа, создавая комплексный образ академической среды. Этот подход обогащает понимание влияния контекстуальных факторов на успеваемость студентов и предоставляет ценные инсайты для персонализированных стратегий обучения.

### 3. Кластерный анализ.

Кластерный анализ направлен на выделение групп студентов со схожими характеристиками и успехами. Данный анализ описывает методы, используемые для выделения кластеров, интерпретацию результатов и возможные стратегии персонализированного обучения.

Выбор признаков для кластеризации включает в себя как количественные, так и качественные переменные. Учитываются оценки, активность на платформе, анкетные данные, текстовые характеристики и даже динамика успеваемости в различных периодах. Этот широкий набор признаков позволяет выделить различия и схожесть между студентами [4].

Результаты кластерного анализа интерпретируются с учетом характеристик, которые объединяют студентов в один кластер. Это может быть схожесть по уровню активности, способу восприятия материала, мотивации или даже стилю обучения. Интерпретация позволяет выявить сценарии успеха и особенности каждой группы.

На основе результатов кластерного анализа, разрабатываются стратегии персонализированного обучения для каждой группы студентов. Это может включать в себя адаптацию учебных программ, изменение методов оценки, предоставление дополнительных ресурсов для определенных групп и другие подходы, направленные на улучшение обучения и результатов.

Применение кластерного анализа дает возможность систематизировать и группировать студентов, что позволяет эффективно адаптировать стратегии обучения к индивидуальным потребностям каждой группы.

Выбор методов зависит от характера данных и поставленных задач. Например, машинное обучение подходит для предсказания количественных показателей, а анализ текстовых данных может раскрывать дополнительные контекстуальные аспекты. Комбинированный подход позволяет получить глубокое понимание взаимосвязей в данных и выявить множество факторов, влияющих на успеваемость студентов.

Исследование раскрывает потенциал методов интеллектуального анализа для оптимизации образовательных процессов и создания более эффективных стратегий обучения. Однако требуются дальнейшие исследования для углубленного понимания влияния различных факторов на успеваемость студентов и оптимального использования полученных данных в практических образовательных сценариях.

### *Список литературы:*

1. Евстигнеева О.А. Анализ успеваемости студентов: применение методов машинного обучения и кластерного анализа / О.А. Евстигнеева, В.Г. Кольцова // Вестник высшей школы. – 2020. – № 3. – С. 78-92.



2. Лавренов Е.К. Роль текстовых данных в прогнозировании успеваемости в образовательной среде / Е.К. Лавренов, Г.С. Новиков // Педагогика и психология. – 2018. – № 2. – С. 56-68.

3. Носикова Д.А. Применение кластерного анализа для выделения групп студентов с различными образовательными потребностями / Д.А. Носикова, Н.П. Крылова // Образовательные технологии. – 2017. – № 4. – С. 112-126.

4. Рудинский Л.Е. Интеграция методов машинного обучения и кластерного анализа для анализа образовательных данных / Л.Е. Рудинский, А.И. Петров // Информационные технологии в образовании. – 2021. – № 2. – С. 88-104.

5. Harb H., Abazid M. An overview of the business process re-engineering in higher education / H. Harb, M. Abazid // Asian Journal of Management Sciences & Education. 2018. Vol. 7, № 2. P. 99-106.

