

ИСПОЛЬЗОВАНИЕ АНАЛИЗА ДАННЫХ В ОБРАЗОВАНИИ

Г. Г. Трубач

*Белорусский государственный университет, Минск; gen-
nady.trubach@gmail.com;*

науч. рук. – А. Н. Курбацкий, д-р. техн. наук, проф.

Анализ данных используется в настоящее время для решения многих задач из разных областей. С его помощью можно исследовать, фильтровать, преобразовывать и моделировать данные для извлечения полезной информации и принятия решений. Одной из сфер анализа данных является интеллектуальный анализ образовательных данных (Educational Data Mining, EDM) направленный на исследование образовательных данных в целях улучшения качества образования. В данной работе приводится исследование и пример использования анализа образовательных данных.

Ключевые слова: анализ образовательных данных, анализ успеваемости, инструменты анализа данных, источники образовательных данных, качество образования.

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ОБРАЗОВАТЕЛЬНЫХ ДАННЫХ

Основные сведения

Интеллектуальный анализ образовательных данных (Educational Data Mining, EDM) (ИАОД) – это направление исследований, связанное с применением методов интеллектуального анализа данных (Data mining), машинного обучения и статистики к информации, производимой образовательными учреждениями (рис. 1). Такой анализ пытается извлечь закономерности из данных, генерируемых в процессе обучения, таких как, например, посещаемость, успеваемость или данных из систем дистанционного обучения [1, с. 2].

Главной целью ИАОД является повышение качества образования и подготовки специалистов. Выделяют четыре категории пользователей, вовлеченных в него [1, с. 4]:

1. Обучающиеся, заинтересованные в осознанном формировании индивидуальной образовательной траектории; правильном выборе факультативных дисциплин и дисциплин по выбору; профессиональной ориентации и точном выборе сферы будущей профессиональной деятельности.

2. Преподаватели, заинтересованные в разделении студентов на группы и подбор для каждой группы оптимальной технологии и наиболее эффективных методов обучения; оптимизации структуры и содержания лекционного курса; прогнозировании успешности учебной деятельности.

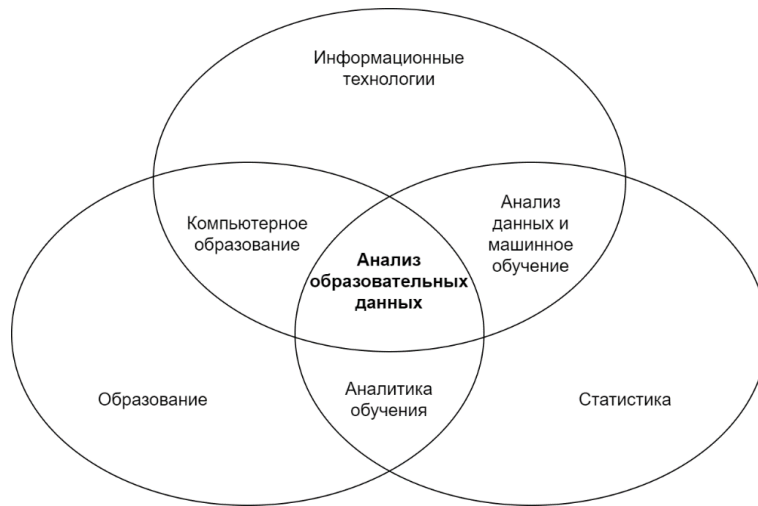


Рис. 1. Интеллектуальный анализ образовательных данных

3. Исследователи, заинтересованные в разработке методов объективной оценки эффективности и результативности учебного процесса, разработке новых технологий и методов обучения, совершенствовании существующих и разработке новых методов и алгоритмов анализа образовательных данных.

4. Администраторы, заинтересованные в принятии научно обоснованных управленческих решений.

Этапы анализа

Множество методов интеллектуального анализа данных были включены в различные образовательные контексты. В каждом случае основной целью является преобразование данных в значимую информацию о процессе обучения и поиск закономерностей для лучшего принятия решений относительно образовательного процесса. Таким образом ИАОД включает следующие этапы [1, с. 8]:

1. Предварительная обработка данных.
2. Выявление закономерностей в данных тем или иным методом анализа данных.
3. Проверка обнаруженных закономерностей.
4. Применение найденных закономерностей для прогнозирования будущих событий в обучающей среде.
5. Использование построенных прогнозов при поддержке принятия решений и выработки образовательной политики.

ИСТОЧНИКИ ДАННЫХ

Для исследования данных можно использовать данные внутренних электронных систем и систем дистанционного обучения.

В частности, можно использовать такие данные внутренних систем, как посещаемость, успеваемость, данные о распределении, данные библиотеки и другие доступные данные, позволяющие исследовать успеваемость студента и его отношение к образовательному процессу.

Также можно использовать данные систем дистанционного обучения. Например, в БГУ на данный момент активно используется система Moodle, поддерживающая ведение журналов посещаемости и успеваемости, а также хранящая другие образовательные данные.

ПРИМЕР ИСПОЛЬЗОВАНИЯ

Рассмотрим пример на основе данных академической успеваемости, опубликованных на сайте Kaggle [2]. Они включают в себя такие атрибуты, как: пол; национальность; место рождения; степень образования; уровень успеваемости: L – низкий уровень (0 – 69), M – средний уровень (70 – 89), H – высокий уровень (90 – 100); тема курса; семестр: первый или второй; количество посещенных курсов; количество пропущенных дней и другие. Для исследования будем использовать язык Python.

Проанализировав эти данные, можно увидеть, что студенты во втором семестре учатся лучше, чем в первом: низких оценок (L) больше в первом семестре, высоких (H) больше во втором (рис. 2).

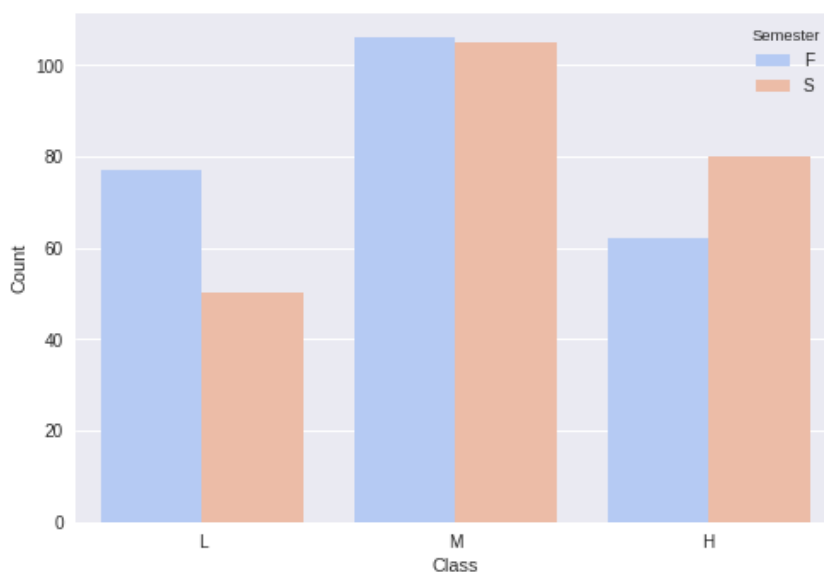


Рис. 2. Успеваемость студентов по семестрам

Данный результат говорит о том, что, возможно, нужно перераспределить нагрузку между семестрами, и в первом семестре давать более доступный материал, чем во втором.

Также можно заметить, что успеваемость студентов коррелирует с посещаемостью ими различных образовательных ресурсов (рис. 3).

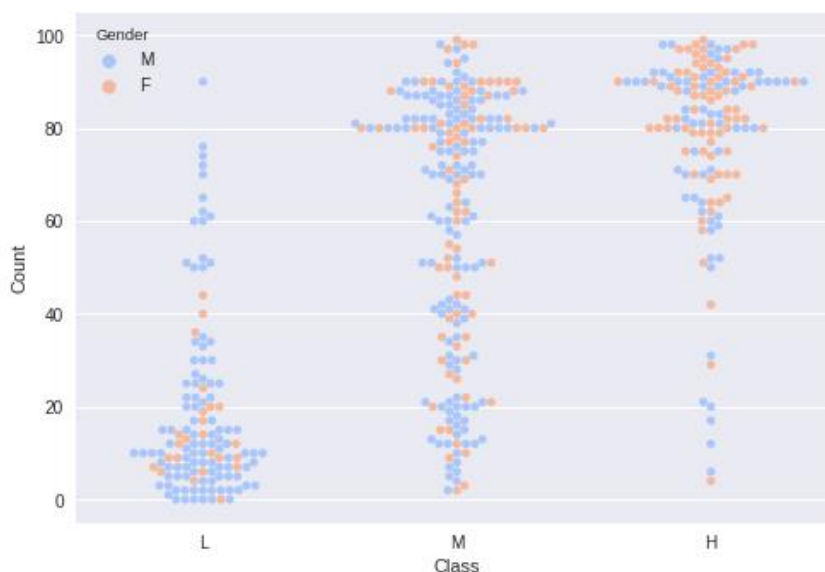


Рис. 3. Посещаемость образовательных ресурсов

Используя градиентный бустинг можно выделить ключевые параметры (рис 4), которые можно использовать для дальнейшего анализа и прогнозирования.

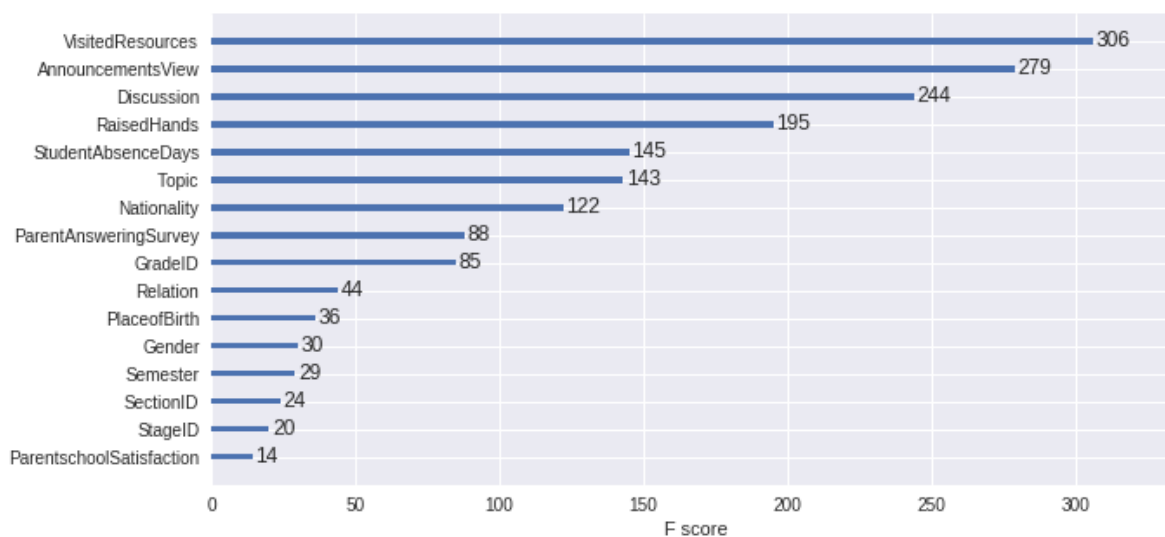


Рис. 4. Ключевые параметры

Библиографические ссылки

1. Белоножко П. П., Карпенко А. П., Храмов Д. А. Анализ образовательных данных: направления и перспективы применения // Интернет-журнал «НАУКОВЕДЕНИЕ». 2017. Т. 9, №4. URL: <http://naukovedenie.ru/PDF/15TVN417.pdf> (дата обращения: 29.05.2020).
2. Aljarah I. Students' Academic Performance Dataset // Kaggle. 2016. URL: <https://www.kaggle.com/aljarah/xAPI-Edu-Data> (дата обращения: 29.05.2020).