

Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт математики и фундаментальной информатики
Базовая кафедра вычислительных и информационных технологий

УТВЕРЖДАЮ
Заведующий кафедрой
_____ / В.В. Шайдуров

«__» _____ 2023 г.

БАКАЛАВРСКАЯ РАБОТА

Направление 02.03.01 Математика и компьютерные науки

ИССЛЕДОВАНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА УСПЕВАЕМОСТЬ И МОТИВАЦИЮ СТУДЕНТОВ С ПОМОЩЬЮ МЕТОДОВ СТАТИСТИЧЕСКОГО И ФАКТОРНОГО АНАЛИЗА

Руководитель	доцент, кандидат физико- математических наук	И. В. Баранова
Выпускник		Д.А. Попов
Нормоконтролер		Т.Н. Шипина

Красноярск 2023

ВВЕДЕНИЕ

Главной задачей любого высшего учебного заведения, является подготовка и выпуск квалифицированных специалистов. Для выполнения данной задачи применяются различные методы контроля успеваемости обучающихся.

Одним из таких методов может стать Data Mining (интеллектуальный анализ данных). Data Mining — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных для интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. С помощью инструментов интеллектуального анализа можно получить полезные знания об успеваемости тех или иных студентов, выяснить, какие параметры влияют на средний балл студентов. В данной работе исследуются данные о студентах для построения предсказательных моделей, а также выполняется анализ показателей успеваемости студентов, поиск скрытых закономерностей и выделение признаков, наиболее сильно влияющих на целевую переменную.

Эффективность анализа данных и точность полученных предиктивных моделей во многом зависит от качества входных данных, поэтому их подготовка является важным шагом. Данные, полученные на этапе сбора “сырые”. Они могут содержать выбросы, недопустимые значения, повторяющиеся строки и т.д. Признаки (предикторы, объясняющие переменные) могут сильно коррелировать между собой.

Целями данной работы являются:

- Исследование факторов, влияющих на успеваемость и мотивацию студентов;
- Предсказание успешности обучения студентов в вузе на основе выделенных факторов;
- Выявление групп риска студентов, которые могут быть отчислены.

Для достижения поставленных целей необходимо выполнить следующие задачи:

- Изучить постановки задачи регрессии, бинарной классификации;
- Изучить основные методы машинного обучения;
- Изучить основные методы статистического анализа, регрессионного анализа, корреляционного анализа;
- Реализовать основные методы решения поставленных задач: линейную регрессию, логистическую регрессию, парную корреляцию;
- Построить и обучить различные модели предсказания успешности студентов;
- Решить практическую задачу анализа данных;
- Выполнить анализ полученных результатов.

В первой главе бакалаврской работы приводятся основные понятия и определения из статистического анализа данных. Во второй главе поставлена задача и приведена информация об имеющейся статистике. В третьей главе рассмотрены методы анализа данных, необходимые для решения поставленных задач. В четвертой главе приведены результаты практической реализации поставленных задач, и их анализ.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

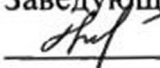
1. Айвазян, С.А. Прикладная статистика: исследование зависимостей / С.А. Айвазян, С.А., Енюков, И.С., Мешалкин, Л.Д. – Москва: Финансы и статистика, 1985. – 487с.
2. Е. С. Шмарихина, Исследование факторов успеваемости обучающихся [сайт] – 2023. URL: <https://cyberleninka.ru/article/n/issledovanie-faktorov-uspevaemosti-obuchayuschisya> (дата обращения: 20.06.2023)
3. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. – Новосибирск: ИМ СО РАН, 1999. – 270 с.
4. Кобзарь, А.И. Прикладная статистика / А.И. Кобзарь. – Москва: Физматлит, 2012. – 816 с.
5. Официальная документация Python / Python [сайт] – 2023. – URL: <https://docs.python.org/3/> (дата обращения: 20.06.2023)
6. Официальная документация библиотеки Imbalanced-learn / Imbalanced-learn [сайт] – 2023. – URL: https://imbalanced-learn.org/stable/user_guide.html#user-guide (дата обращения: 20.06.2023)
7. Официальная документация библиотеки Pandas / Pandas [сайт] – 2023. – URL: <https://pandas.pydata.org/docs/> (дата обращения: 20.06.2023)
8. Официальная документация библиотеки Sklearn / Sklearn [сайт] – 2023. – URL: https://scikit-learn.org/stable/user_guide.html (дата обращения: 20.06.2023)
9. Себер, Дж. Линейный регрессионный анализ / Дж. Себер. – Москва: Мир, 1980. – 456 с.
10. Т. Е. Хавенсон, А. А. Соловьева. Связь результатов Единого государственного экзамена с успеваемостью в вузе [сайт] – 2013. – URL: <https://publications.hse.ru/pubs/share/folder/bm8thy8qcj/119778116.pdf> (дата обращения: 20.06.2023)
11. Чубукова И.А. Курс лекций по Data Mining [сайт] – 2023. – URL: http://lnfm1.sai.msu.ru/~milkyway/Books/Chubukova-Data_Mining.pdf (дата обращения: 20.06.2023)

Министерство науки и высшего образования РФ
Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт математики и фундаментальной информатики
Базовая кафедра вычислительных и информационных технологий

УТВЕРЖДАЮ

Заведующий кафедрой

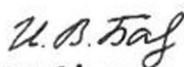
 / В.В. Шайдуров

«23» июня 2023 г.


БАКАЛАВРСКАЯ РАБОТА

Направление 02.03.01 Математика и компьютерные науки

ИССЛЕДОВАНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА УСПЕВАЕМОСТЬ И МОТИВАЦИЮ СТУДЕНТОВ С ПОМОЩЬЮ МЕТОДОВ СТАТИСТИЧЕСКОГО И ФАКТОРНОГО АНАЛИЗА

Руководитель  23.06.2023 доцент, кандидат физико-математических наук И. В. Баранова

Выпускник  23.06.2023 Д.А. Попов

Нормоконтролер  23.06.2023 Т.Н. Шипина

Красноярск 2023