

Университет ИТМО

Практическая работа №3
по дисциплине «Визуализация и моделирование»

Автор: Редичкина Анна Максимовна

Поток: ВИМ 1.1

Группа: К3242

Факультет: ИКТ

Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

1 Описание датасета

'Graduate Admission 2' содержит в себе информацию об академической успеваемости школьников, их учебных достижениях. На основе этих параметров был рассчитан шанс поступления в один из пяти университетов, куда хочет попасть абитуриент. В датасете 8 колонок и 500 строк.

2 Содержание колонок

Таблица 1: Columns

Название колонки	Описание данных	Тип данных	Шкала
Serial No.	Порядковый номер абитуриента	целочисленный тип данных	Номинальная
GRE Score	Результаты экзамена GRE (тест, который необходимо сдавать для поступления в аспирантуру, магистратуру или иной последипломный курс в вузе США и ряда других стран). Максимальный балл - 340.	целочисленный тип данных	Интервальная
TOEFL Score	Результаты экзамена на знание английского языка (максимальный балл - 120).	целочисленный тип данных	Интервальная
University rating	Позиция университета, куда подает заявление абитуриент, в рейтинге (топ-5 университетов).	целочисленный тип данных	Порядковая
SOP	Statement of Purpose Strength (оценка за заявление о намерениях). Максимальный балл - 5.	числа с плавающей точкой	Интервальная
LOR	Letter of Recommendation (оценка за рекомендательное письмо). Максимальный балл - 5.	числа с плавающей точкой	Интервальная
CGPA	Средний балл аттестата школьника. Максимальный балл - 10.	числа с плавающей точкой	Интервальная
Research	Данные показывают наличие исследовательского опыта (1- есть, 0 - нет)	целочисленный тип данных	Номинальная
Chance of Admit	Шанс попадания в университет (от 0 до 1)	числа с плавающей точкой	Интервальная

3 Проблемы датасета

[h]

Таблица 2: Проблемы датасета

Название колонки	Тип данных	Проблема	Решение
Serial No.	целочисленный тип данных	Данные не влияют на вероятность поступления	Удалить столбец
GRE Score	целочисленный тип данных	-	-
TOEFL Score	целочисленный тип данных	-	-
University rating	целочисленный тип данных	-	-
SOP	числа с плавающей точкой	-	-
LOR	числа с плавающей точкой	-	-
CGPA	числа с плавающей точкой	-	-
Research	целочисленный тип данных	-	-
Chance of Admit	числа с плавающей точкой	-	-

4 Гипотезы

1. GRE и CGPA лучше других показателей определяют шансы на поступление.
2. Опыт в исследовательской деятельности важен для поступления.
3. Оценка за заявление о намерениях (SOP) меньше всего влияет на поступление.
4. Оценки за SOP и LOR меньше всего связаны с другими показателями абитуриентов.
5. Знание английского языка - один из самых важных показателей при приеме в университет.