

Университет ИТМО

Практическая работа №1
по дисциплине «Визуализация и моделирование»

Автор: Голуб Анна Леонидовна

Поток: ВИМ 1.2

Группа: К3243

Факультет: ИКТ

Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

В ходе практической работы был выбран датасет [Goodreads-books](#) с информацией о книгах и их рейтингах на популярном ресурсе Goodreads. Информация о столбцах датасета представлена в таблице.

название столбца в датасете	краткое описание данных	тип данных
bookID	уникальный номер книги в датасете	натуральное число
title	название книги, использованное при публикации	строка
authors	список авторов книги	строка с перечислением имен авторов через /
average_rating	средний рейтинг книги на сайте	число от 0.0 до 5.0
isbn	ISBN-идентификатор книги	натуральное число (11 знаков)
isbn13	ISBN-идентификатор книги (13 знаков)	натуральное число (13 знаков)
language_code	язык оригинала книги	строка
num_pages	число страниц в книге	натуральное число
ratings_count	сколько раз книге был поставлен рейтинг	натуральное число
text_reviews_count	число письменных отзывов на книгу	натуральное число
publication_date	дата первой публикации книги	дата в формате ММ/DD/YYYY
publisher	издательство, опубликовавшее книгу	строка

Датасет можно использовать для предсказания среднего рейтинга книги на основании данных о ее авторах, числе странице, языке оригинала и т. д. В дополнение, с использованием этих данных можно проверить, есть ли существенная корреляция между рейтингом книги и ее характеристиками, в частности:

- пол и национальность автора
- язык оригинала
- число страниц
- длина названия
- название издательства, где книга была издана