

Как я создал проект по машинному обучению.

Расскажу об этапах создания проекта, построении математической модели и применении модели для предсказания покупок пользователей портала DigitalEdu.



Постановка задачи

1 Цель:

Предсказать покупку курса по профессиональной переподготовке пользователями портала DigitalEdu.

2 Проблема:

Отсутствие информации о пользователях и их покупательском поведении.

3 Решение:

Собрать данные о пользователях, обработать их и построить математическую модель для предсказания их покупок.

Сбор и подготовка данных



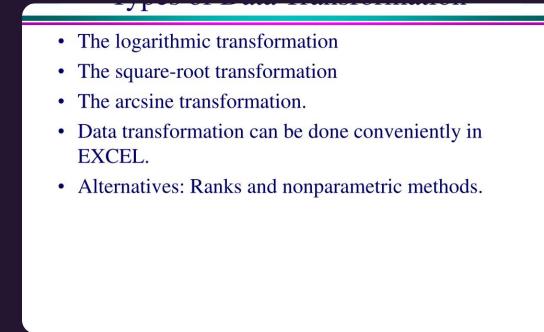
Сбор данных

Собрали данные о пользователях, такие как возраст, пол, образование, доход и оффлайн активность.



Подготовка данных

Обработали данные, удалили дубликаты, заполнили пропущенные значения и преобразовали категориальные признаки в численные.



Трансформация данных

Отмасштабировали и нормализовали данные для правильной работы статистических алгоритмов.

Обработка данных

Статистика

Использовали python библиотеки pandas и питру для проведения статистического анализа данных и поиска корреляций между признаками.

Визуализация

Создали визуализации данных с помощью библиотек seaborn и matplotlib, чтобы понять распределения признаков и их взаимосвязь.

Работа с выбросами

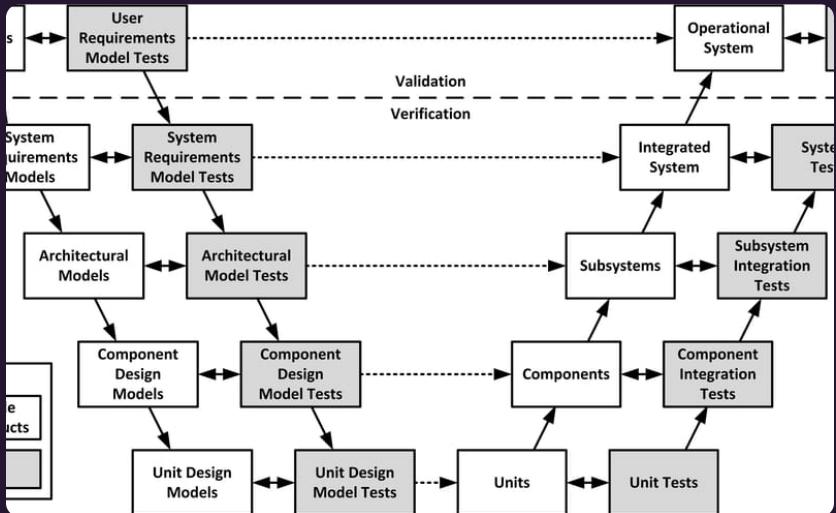
Использовали методы работы с выбросами и аномалиями данных, чтобы убрать погрешности в модели.



Построение математической модели



Тестирование модели



	All Columns Accuracy (%)	Reduced Accuracy (%)		
		Change (%)	Accuracy (%)	Change in Data (%)
M0	80.4	71.2	-9.2	-80
M1	91.3	87.0	-4.3	-86
M7	78.5	67.2	-11.3	-86
M8	92.0	90.7	-1.3	-86
M10	82.3	77.4	-3.9	-89
M11	99.9	94	-5.9	-82

Тестирование на новых данных

Создали тестовый датасет, чтобы проверить, насколько хорошо наша модель работает на новых данных.

Оценка точности модели

Провели оценку точности модели, используя метрики, такие как accuracy, precision, recall и F1-score.