

**Государственное бюджетное
общеобразовательное учреждение Школа №1554**

**Примеры реализации методики
практико-ориентированного подхода в
преподавании курса
«Большие данные»**

**учитель информатики
И.С. Смирнова**

Москва, 2023 год

Цель и основная задача

- Формирование необходимого понятийного аппарата в рамках учебной программы курса.
- Формирование устойчивой мотивации в изучении предмета.
- Формирование навыков самообразования.

Основная задача — разработка практических работ по темам курса, подбор, предварительная обработка наборов данных.

Примеры реализации

✓ Весь материал сгруппирован по темам

☰ 2223 Анализ данных Лента Задания Пользователи Оценки

Все темы

- Машинное обучение
- Статистика для ан...
- Математика для а...
- Инструменты анал...

Машинное обучение

 Введение в машинное обучение Срок сдачи: 19 янв.

Статистика для анализа данных

 Регрессия Срок сдачи: 30 дек. 2022 г.

 Корреляция Срок сдачи: 26 дек. 2022 г.

 Нормальное распределение Срок сдачи: 24 нояб. 2022 г.

 Меры центральной тенденции и изменчи... Срок сдачи: 14 нояб. 2022 г.

✓ Практика на каждом уроке

Примеры реализации

Теоретический материал, практическое задание, тест

✓ Постепенное нарастание сложности заданий

 Нестационарные временные ряды

 Прогнозирование временных рядов

 Анализ временных рядов

Практическая работа



↓
Прогнозирование вре...
Colaboratory



↑
Прогнозирование-вре...
<https://disk.yandex.ru/i/mnU4>

Материал для изучения

Срок сдачи: 21 нояб. 2022 г.

25
Сданы

3
Назначено

Временные ряды
<https://forms.gle/3HGnhRHVx>

↑
Тест

✓ Подробные пошаговые инструкции

Опубликован

Задание 1.

1) Считайте данные. Установите индекс таблицы равным столбцу *Datetime*. Преобразуйте тип данных *Datetime* с *object* на *datetime64*.

2) Чтобы проверить, в хронологическом ли порядке расположены даты и время, посмотрите атрибут индекса таблицы *is_monotonic*. Если порядок

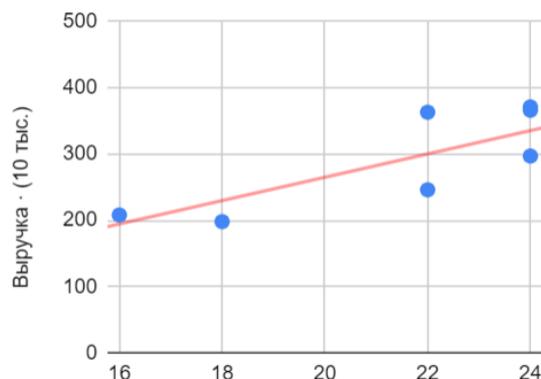
✓ Тесты
контроля
знаний

Пример реализации

Фрагменты задания «Уравнения регрессии»

	A	B	C	D
1	Название	Площадь, кв.м	S от метро, м	Выручка, (10 тыс.)
2	Вилка	30	80	469
3	Ложка			66
4	Авоська			71
5	Небоська			88
6	Конкор			
7	Аккорд			
8	Рекорд			
9	Пикап			
10	Самокат			

Выручка относительно параметра "Площадь, кв.м"



	H	I	J	K
	матрица корреляции			
		x1	x2	x3
x1				
x2				
x3				
x4				

✓ Реализация и в ЭТ, и на языке программирования

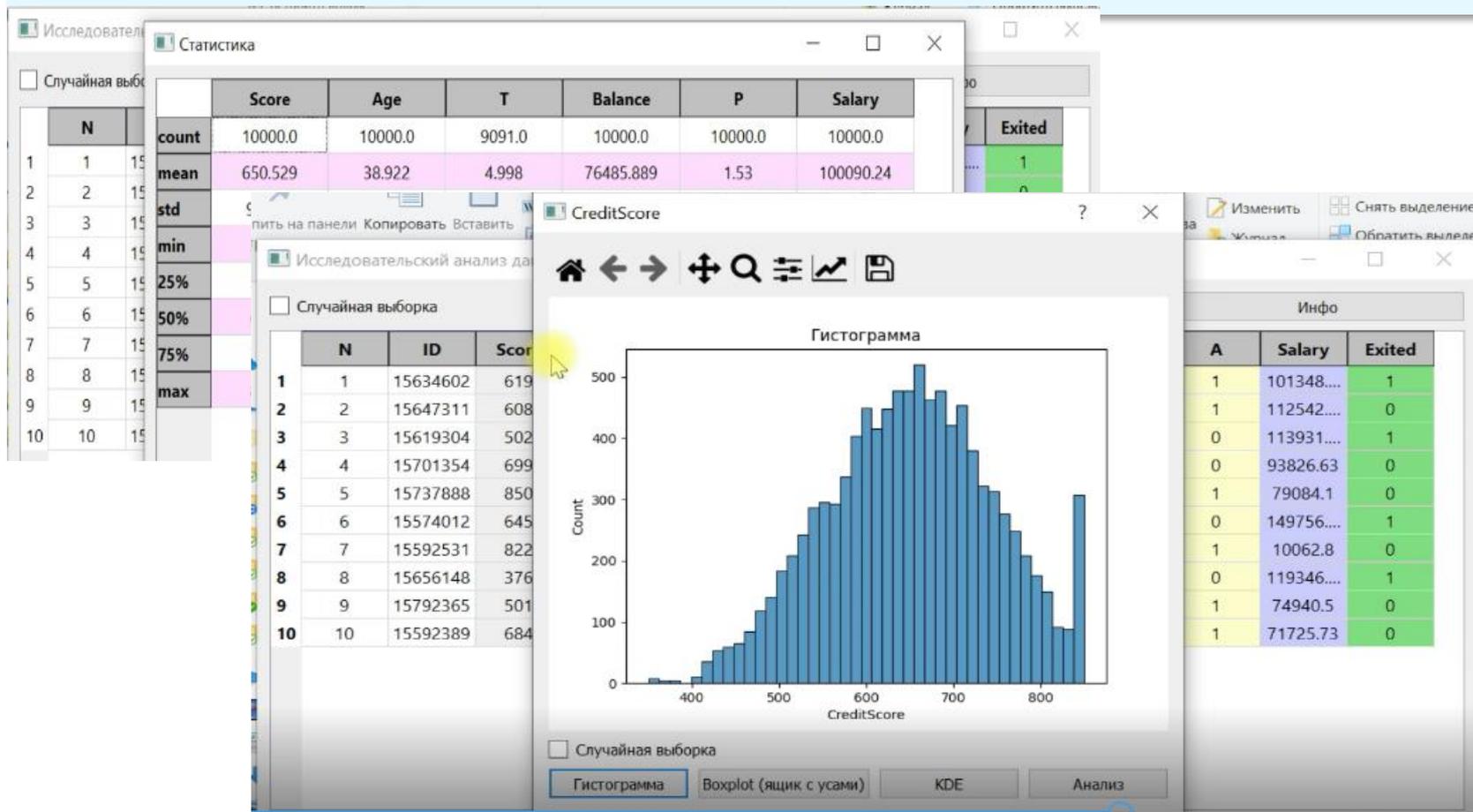
```
# чтение данных
df = pd.read_csv('/content/shops.csv', delimiter=';')
# создание признаков
X = df['S от метро, м']
y = df['Выручка, (10 тыс.)']
```

$$a = \frac{\sum_{xy}}{S_{xx}}, \quad b = \bar{y} - \bar{x} \cdot a$$

$$R = \frac{\text{сумма произведений отклонений } y}{\sqrt{\text{сумма квадратов откл. } y \cdot \text{сумма квадрата}}}$$

Пример реализации

- ✓ Использование собственного приложения для анализа данных



Пример реализации

Фрагменты задания «Кластеризация на реальных данных»

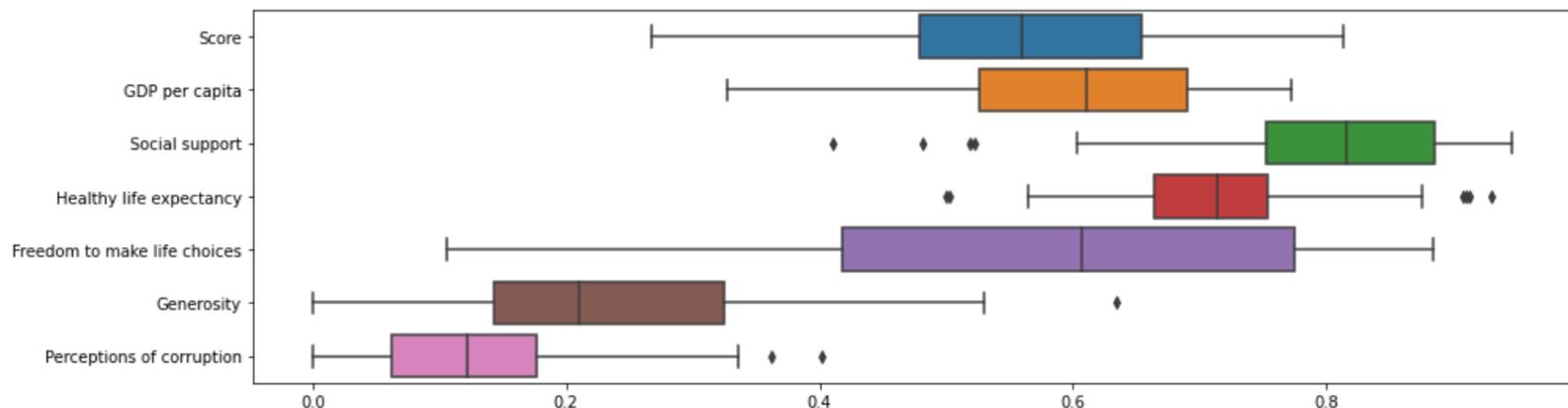
11-Кластеризация_на_реальных_данных ☆

✓ Интересные задания на реальных данных

Кластеризация на реальных данных

Данные

Всемирный доклад о счастье (англ. World Happiness Report) – ежегодный доклад, публикуемый подразделением ООН по поиску решений стабильного развития (англ. UN Sustainable Development Solutions Network). <https://ru.wikipedia.org/wiki/>



В каком кластере окажется Россия?

```
Data_clastered["region"] = regions
```

Пример реализации

✓ Презентационные материалы по каждой теме

Типичные проблемы анализа данных

✓ Структурная разница между признаками

	A	B	C	D	E	F
1	Компания	Доход, млрд руб.	Доля Р, %	ОП	Инт (да/нет)	Сектор экономики
2	Авер	19,0	21,85	2	нет	Химия
3	Ант	29,4	18,00	3	нет	Химия
4	Астон	23,9	19,00	3	нет	Металлургия
5	Бмарт	18,4	13,95	2	да	Химия
6	Брек	25,7	11,15	3	да	Металлургия
7	Бумо	12,1	8,45	2	да	Металлургия
8	Виж	23,9	15,10	4	да	Торговля
9	Вурд	27,2	29,00	5	да	Торговля
10						

Количественные,
те для которых
осмыслена операция
усреднения их
значений

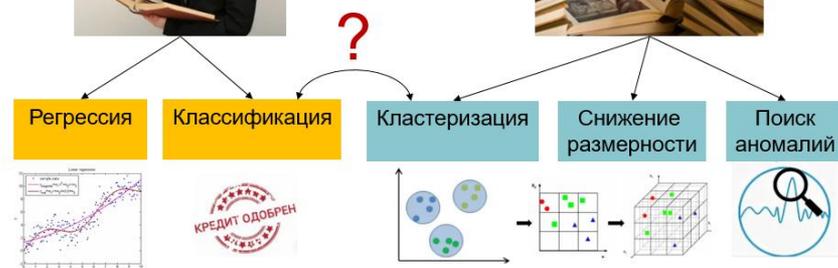
**Качественные
(категориальные)**

Представлены
непересекающимися
категориями значений

Бинарные

Номинальные

Задачи машинного обучения



Python для комбинаторики

```
from itertools import *
```

```
s = 'ТИХОРЕЦК'
```

```
for p in permutations(s, 4):
```

```
    print(p, end=' ')
```

[permutations]

permutations

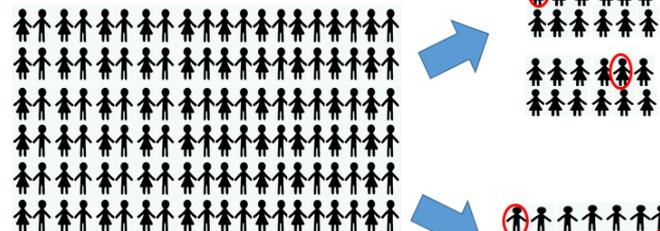
перестановки без повторов

```
('Т', 'И', 'Х', 'О') ('Т', 'И', 'Х', 'Р')
('Т', 'И', 'Х', 'Е') ('Т', 'И', 'Х', 'Ц')
('Т', 'И', 'Х', 'К') ('Т', 'И', 'О', 'Х')
```

Стратифицированная выборка

- Стратифицированная выборка (stratified sample)

Группы максимально однородны!

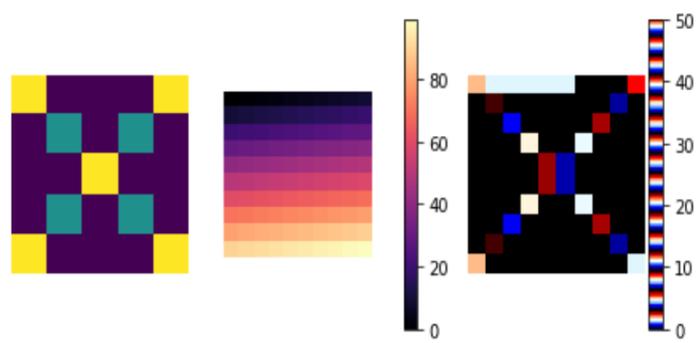


Пример реализации

✓ Обратная связь

Василий Устенко 100/100 Черновик < > Не возвращено [Вернуть](#)

+ Код + Текст [Последнее изменение: 15 дек. 2022 г.](#) Подключиться ⚙️



```
[ ] import numpy as np #
import matplotlib.pyplot as plt #

fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(8, 4)) #
ax1.imshow(z) #
ax1.axis('off')
ax2.imshow(z)
```

Файлы
Сдано: 15 дек. 2022 г., 14:12
[Посмотреть историю](#)
👤 Василий Устенко ...

Оценка
100/100

Личные комментарии
Отличная работа!

Что еще предстоит сделать?

В планах по развитию методики – создание электронного учебного пособия и собственной онлайн-платформы для анализа данных.

На сегодняшний день командами учащихся ведется работа над двумя проектами, на основе которых будут создаваться web-приложения.

Результаты

- ✓ Повышение мотивации при изучении предмета.
- ✓ Повышение качества знаний при прохождении промежуточной аттестации.
- ✓ Увеличение количества учащихся, сдающих предпрофессиональный экзамен.
- ✓ Большая вовлеченность в самостоятельную работу над проектами в области науки о данных.
- ✓ Активное участие в соревнованиях по Data Science и искусственному интеллекту, во Всероссийском учебном фестивале RuCode.

Спасибо за внимание!

disath@mail.ru

Материалы, используемы в презентации:

Уравнение регрессии

https://docs.google.com/spreadsheets/d/11-LNyh1cmOxjPWaA2LjDLCaKohlXN9-aUfDvZOc_Dbg/edit?usp=sharing

Кластеризация на реальных данных (шаблон)

<https://colab.research.google.com/drive/1OILc0P9KsKQBy2JLAAj2oJYpWSgV7fQV?usp=sharing>

Анализ временных рядов

<https://colab.research.google.com/drive/1JBWLOzDe4h2rCld6ftzkxHHed1ORgBOU?usp=sharing>