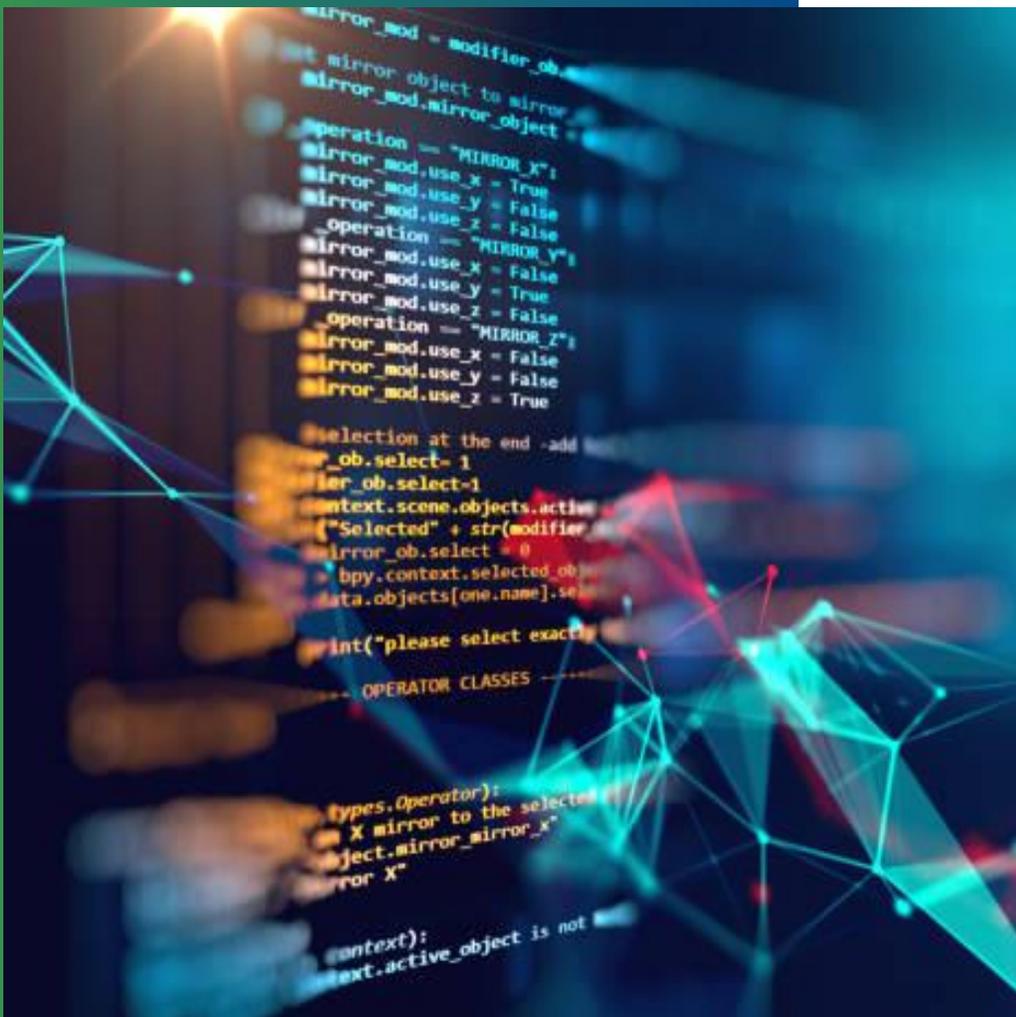




Обработка экспериментальных данных. Знания

Быстренина Ирина Евгеньевна,
кандидат педагогических наук, доцент
кафедры прикладной информатики
РГАУ-МСХА имени К.А. Тимирязева





Развитие науки происходит благодаря ее постоянному обновлению новыми знаниями. Для их получения используются методы научного исследования.

Совокупность всех методов получения нового знания и принципов построения научной теории, а также системная рефлексия (осмысление) познавательных возможностей этих методов называется **методологией** (от греч. *methodos* – путь, способ и *logos* – понятие, знание).



Методология науки дает характеристику компонентов научного исследования его объекта, предмета анализа, задач исследования, совокупности исследовательских средств, необходимых для их решения, а также формирует представление о последовательности движения исследователя в процессе решения исследовательских задач.



Методы исследования, в отличие от методологии, – это сами способы изучения явлений, получения научной информации о них с целью установления закономерных связей, отношений и построения научных теорий.

Методы исследования



Теоретические	Эмпирические	Математические
<ul style="list-style-type: none">• анализ и синтез,• моделирование,• классификация,• типология и др.	<ul style="list-style-type: none">• наблюдение,• сравнение,• эксперимент и др.	<ul style="list-style-type: none">• регистрация,• ранжирование,• шкалирование,• статистические методы,• метод визуализации данных,• методы и модели динамического программирования,• методы и модели теории графов и сетевого моделирования и др.



Объект исследования – это объект любого характера (технического, социального, экономического, астрономического и т.д.), который изучается экспериментальным путем.

Эксперимент – это специальным образом спланированная и организованная процедура изучения некоторого объекта исследования, при которой на этот объект оказывают запланированные воздействия и регистрируют его реакции на эти воздействия.



Цифровые технологии анализа данных

KAGGLE

<https://www.kaggle.com/>

- Kaggle – система организации конкурсов по исследованию данных, а также социальная сеть специалистов по обработке данных и машинному обучению.

Среда организована как публичная веб-платформа, на которой пользователи и организации могут публиковать наборы данных, исследовать и создать модели, взаимодействовать с другими специалистами по данным и инженерами по машинному обучению, организовывать конкурсы по исследованию данных и участвовать в них. В системе размещены наборы открытых данных, предоставляются облачные инструменты для обработки данных и машинного обучения. Также реализованы обучающие ресурсы, имеется раздел для размещения вакансий работодателями, где тоже возможна организации конкурсов для отбора наилучших кандидатов.



☰ kaggle

🏠 Home

🏆 Compete

📁 Data

📄 Code

🗨️ Communities

🎓 Courses

⌵ More

🔍 View Active Events

🔍 Search

Data Tasks (1) Code (150) Discussion (11) Activity Metadata

Download (32 KB)

New Notebook

Data Explorer

32 KB

📄 heart.csv

📄 o2Saturation.csv

< heart.csv (11.06 KB)



Detail Compact Column

10 of 14 columns

About this file

Dataset on heart attack

age	sex	cp	trestbps	chol	
Age of the person	Gender of the person	Chest Pain type chest pain type	resting blood pressure (in mm Hg)	cholesterol in mg/dl fetched via BMI sensor	
29	0	0	94	126	0
65	1	5	145	233	1
37	1	2	138	258	0
41	0	1	138	284	0
56	1	1	128	236	0
57	0	0	129	354	0
57	1	0	148	192	0
56	0	1	148	294	0
44	1	1	129	265	0
52	1	2	172	199	1
57	1	2	158	168	0
54	1	0	148	239	0

Summary

📄 2 files

📄 15 columns



JUPYTER LAB

<https://jupyter.org/>

Jupyter Project – сообщество и одноименный инструмент для интерактивной разработки и представления проектов в области наук данных.

Jupyter Lab построен на основе языков программирования Python и использует такие библиотеки как:

- Matplotlib - построение графиков функций
- SciPy - система научных и инженерных расчётов
- NumPy – высокоуровневые математические функции для работы с многомерными массивами
- Pandas – анализ и обработка данных
- Scikit-learn – алгоритмы машинного обучения
- TensorFlow – библиотека машинного обучения от компании Google.



File Edit View Run Kernel Tabs Settings Help

Files

- notebooks
- Data.ipynb (an hour ago)
- Fasta.ipynb (a day ago)
- Julia.ipynb (a day ago)
- Lorenz.ipynb (seconds ago)**
- R.ipynb (a day ago)
- iris.csv (a day ago)
- lightning.json (9 days ago)
- lorenz.py (3 minutes ago)

Running

Commands

Cell Tools

Tabs

Lorenz.ipynb x Terminal 1 x Console 1 x Data.ipynb x README.md x

Code Python 3

In this Notebook we explore the Lorenz system of differential equations:

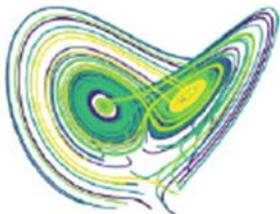
$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= -\beta z + xy\end{aligned}$$

Let's call the function once to view the solutions. For this set of parameters, we see the trajectories swirling around two points, called attractors.

```
In [4]: from lorenz import solve_lorenz
t, x_t = solve_lorenz(N=18)
```

Output View x

sigma 10.00
beta 2.67
rho 28.00



lorenz.py x

```
9 def solve_lorenz(N=10, max_time=4.0, sigma=18.0, beta=8./3, rho=28.0):
10     """Plot a solution to the Lorenz differential equations."""
11     fig = plt.figure()
12     ax = fig.add_axes([0, 0, 1, 1], projection='3d')
13     ax.axis('off')
14
15     # prepare the axes limits
16     ax.set_xlim((-25, 25))
17     ax.set_ylim((-35, 35))
18     ax.set_zlim((5, 55))
19
20     def lorenz_deriv(x_y_z, t0, sigma=sigma, beta=beta, rho=rho):
21         """Compute the time-derivative of a Lorenz system."""
22         x, y, z = x_y_z
23         return [sigma * (y - x), x * (rho - z) - y, x * y - beta * z]
24
25     # Choose random starting points, uniformly distributed from -15 to 15
26     np.random.seed(1)
27     x0 = -15 + 30 * np.random.random((N, 3))
```

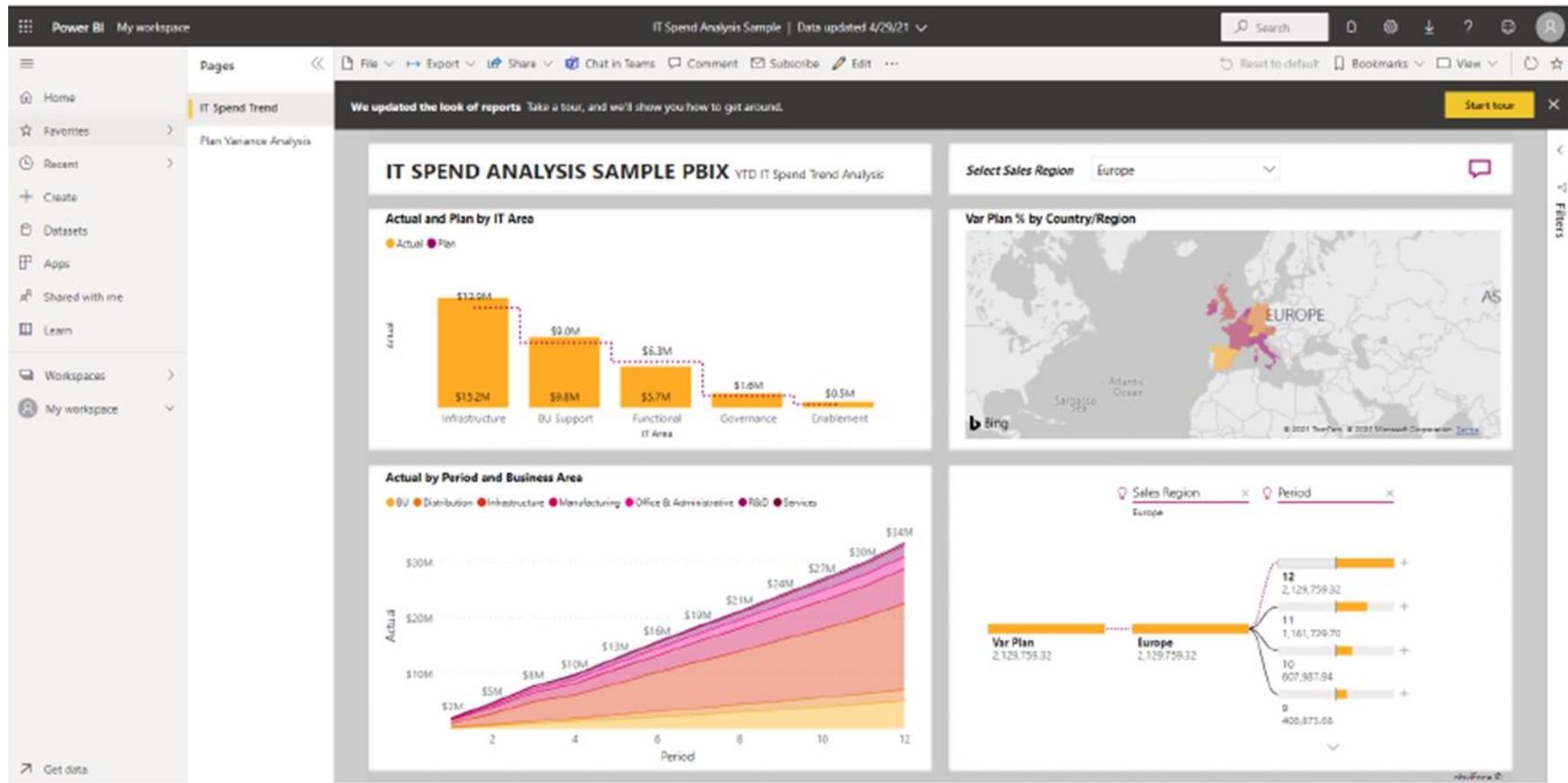


GOOGLE DATA STUDIO

<https://datastudio.google.com/>

Google Data Studio выполняет три основные функции:

- Connect – позволяет собрать нужные вам данные (импортировать с того места, где они находятся) и при необходимости подготовить (т.е. создать рассчитываемые показатели, убрать лишнее, форматировать в нужный формат).
- Visualize – позволяет визуализировать и анализировать данные. Предоставляет возможность рисовать графики, диаграммы, накладывать данные на карту, используя данные из разных источников.
- Share - позволяет подготовить данные и сделать выводы. Сделанной вами визуализацией вы сможете легко поделиться, как и в других сервисах Google.





MICROSOFT POWER BI

<https://app.powerbi.com/>

Построй анализ данных интеллектуальными средствами в масштабах предприятия, включая:

- Получение данных
- Обработка данных
- Модерирование (установления связей, создание вычислительных столбцов и мер);
- Создание визуализаций (отчетов).



Факторы производства:





Для сбора и обработки информационных ресурсов, формирования баз данных и знаний необходим информационно-технологический комплекс, включающий:

- информационную систему,
- хранилище данных,
- информационный портал,
- торговый портал,
- программно-технический комплекс,
- средства телекоммуникации и связи.



Системы, основанные на знаниях

Это компьютерные программы, спроектированные таким образом, чтобы воспроизводить работу экспертов в заданных областях знания.

Есть 6 основных типов систем, основанных на знаниях.

- **Экспертные системы** моделируют процесс принятия решения, свойственный человеческому мозгу.
- **Нейронные сети** моделируют работу человеческого мозга на биологическом уровне (на уровне поведения отдельных клеток мозга, нейронов).
- **Системы, основанные на прецедентах**, моделируют человеческую способность принимать решение по аналогии с уже имеющимися прецедентами.
- **Системы, построенные на генетических алгоритмах**, как явствует из названия, моделируют процесс развития биологических организмов на клеточном уровне.
- **Интеллектуальные агенты** — это программы, для которых задана конечная цель, но они могут самостоятельно выбирать пути достижения этой цели, гибко изменяя алгоритм своего поведения.
- **Системы добычи данных**. Термин «добыча данных» используют для того, чтобы описать процесс обнаружения знаний путем нахождения ранее неизвестных взаимосвязей между данными.



Спасибо за внимание!