

Большие данные и машинное обучение

Введение

Арьков Валентин Юльевич

- профессор кафедры АСУ УУНиТ
- рабочий кабинет 6-215А
- приемная кафедры АСУ 6-322
- <https://github.com/Valentin-Arkov/Analytics>

Инструменты

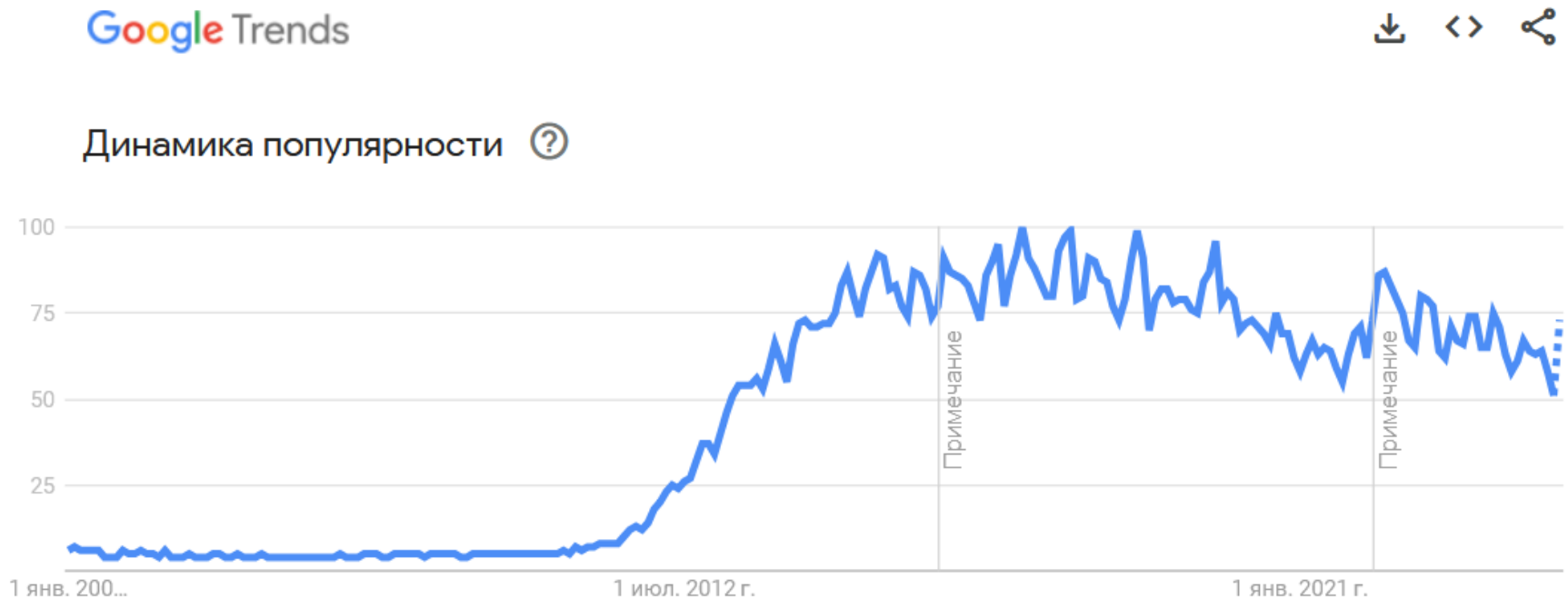
- Python
 - Jupyter Notebook / JupyterLab
 - Google Colab / Anaconda
- Библиотеки
 - SKLearn – SciKitLearn
 - Numpy
 - Pandas
 - Matplotlib

Нейросети – ИИ Ассистенты

- Интеллектуальные помощники
 - YandexGPT
 - GigaChat
 - Perplexity
 - DeepSeek
 - Character.ai / C.ai
- Мульти-модальный ИИ
 - Multimodal AI / Multimodal ML models
 - Текст / Изображение / Аудио / Видео

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

- Big Data



Термины

- Большие данные / Наука о данных
 - Big Data / Data Science
- Машинное обучение
 - Machine Learning
- Искусственный интеллект
 - Artificial Intelligence
- Многослойные нейронные сети
 - Глубокое / глубинное обучение
 - Deep Learning

Большие данные: 3Vs

- Объем (Volume)
 - Сложно обрабатывать традиционными средствами
- Скорость (Velocity)
 - Генерация и обработка данных
- Разнообразие (Variety)
 - Типы данных
 - Структурированные, неструктурированные, частично структурированные

Задание

- 3Vs – 5Vs – 7Vs – 9Vs – ...
- Большие данные / BI-аналитика

The screenshot shows the Russian Wikipedia page for 'Big data'. The article text discusses the complexity of data sets and the challenges of processing them. A chart titled 'Global Information Storage Capacity' is visible, showing the growth of storage capacity from 1986 to 2007. The chart compares analog and digital storage capacities, with digital storage showing exponential growth.

Global Information Storage Capacity in optimally compressed bytes

Year	Storage Type	Capacity (bytes)
1986	ANALOG	0.02 exabytes
1993	ANALOG	0.02 exabytes
2000	ANALOG	0.02 exabytes
2007	ANALOG	0.02 exabytes
2007	DIGITAL	19 exabytes

Legend: 1986 ANALOG 0.02 exabytes, 1993 ANALOG 0.02 exabytes, 2000 ANALOG 0.02 exabytes, 2007 ANALOG 0.02 exabytes, 2007 DIGITAL 19 exabytes.

Machine Learning

- Supervised Learning
 - Обучение с учителем
 - Labelled data
 - Размеченные данные
- Unsupervised Learning
 - Обучение без учителя
 - Неразмеченные данные
- Reinforced learning
 - Обучение с подкреплением
 - Награды и наказания

Профессии

- Data Engineer
 - Инженер по работе с данными
- Data Analyst
 - Аналитик данных
- Data Scientist
 - Исследователь данных
- ML Engineer
 - Инженер по машинному обучению
- ...

Программы обучения

- Цифровой марафон
- <https://it-marathon.21-school.ru/>
- Школа 21
- <https://21-school.ru/>