

# Словарь терминов по курсу “Искусственный интеллект: создайте свою первую нейросеть”

## Занятие 4. Data Science сегодня

**Adversarial attacks** — целенаправленное воздействие на нейронную сеть, которое способно вызвать ошибки в её поведении. Например, подмешиваются специальные шумы, невидимые человеку, но замечаемые нейросетью.

**Airflow** — открытое программное обеспечение для создания, выполнения и мониторинга потоков операций по обработке данных.

**ClickHouse** — реляционное хранилище данных для онлайн обработки аналитических запросов (OLAP) с открытым исходным кодом, основное внимание в котором уделяется скорости и производительности запросов. Разработана на языке C++ в 2008–2012 г. командой Яндекса для Яндекс Метрики.

**Data Analyst** — специалист по подготовке аналитических отчётов о данных и моделях.

**Data Engineer** — специалист по сбору и хранению данных.

**Data Fest** — крупнейшая в Евразии бесплатная конференция, объединяющая исследователей, инженеров и разработчиков, связанных с data science.

**Data Scientist** — специалист по разработке и обучению моделей.

**Github** — сервис для хостинга (то есть хранения) репозитория. На Github есть возможность контролировать разные версии кода, управлять исходным кодом и работать вместе с командой, то есть делать то же, что и Git.

**Hadoop** — свободно распространяемый набор утилит, библиотек и фреймворков для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов.

**Hive** — система управления базами данных на основе платформы Hadoop. Позволяет выполнять запросы, агрегировать и анализировать данные.

**Java** — универсальный язык программирования, который часто используется для веб-разработки и для разработки под Android. Относится к объектно-ориентированным языкам программирования.

**Kafka** — распределённый программный брокер сообщений, проект с открытым исходным кодом, разрабатываемый в рамках фонда Apache. Написан на языках программирования Java и Scala.

**Kaggle** — система организации конкурсов по исследованию данных, а также социальная сеть специалистов по обработке данных и машинному обучению. Принадлежит корпорации Google.

**Keras** — открытая библиотека, написанная на языке Python и обеспечивающая взаимодействие с искусственными нейронными сетями.

**Linux** — операционная система (как Windows или MacOS), удобная для разработчиков. Позволяет полностью использовать возможности современных персональных компьютеров (ПК), снимая ограничения, присущие DOS и MS Windows по использованию памяти машины и ресурсов процессора.

**NumPy** — библиотека с открытым исходным кодом для языка программирования Python. Возможности: поддержка многомерных массивов, поддержка высокоуровневых математических функций, предназначенных для работы с многомерными массивами.

**ODS** (OpenDataScience, [ods.ai](https://ods.ai)) — открытое сообщество специалистов и любителей данных.

**Pandas** — программная библиотека на языке Python для обработки и анализа данных.

**Power BI** — набор программных сервисов Microsoft, которые работают вместе, превращая несвязанные источники данных компании в целостные интерактивные отчёты. При этом источником могут быть базы данных, файлы Excel, данные из облачных источников и интернета, текстовые файлы.

**PyData** — конференции, митапы, воркшопы, которые проходят по всему миру и посвящены анализу данных с применением языка программирования Python.

**R** (произносится “ар”) — язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом.

**Scala** — функциональный объектно-ориентированный язык программирования, который отличается компактным и лаконичным стилем. Scala работает на виртуальной машине Java (JVM) и особенно подходит для разработки сложных приложений.

**Scikit-learn** — один из наиболее широко используемых пакетов Python для Data Science и Machine Learning. Он позволяет выполнять множество операций и предоставляет множество алгоритмов. Scikit-learn также предлагает отличную документацию о своих классах, методах, функциях и описание используемых алгоритмов.

**Spark** — фреймворк с открытым исходным кодом для реализации распределённой обработки неструктурированных и слабоструктурированных данных.

**SQL** — структурированный язык запросов, созданный для того, чтобы получать из базы данных необходимую информацию. Данные хранятся в виде таблиц, они структурированы и разложены по строкам и столбцам, чтобы ими легче было оперировать. Такой способ хранения информации называют реляционными базами данных (от англ. relation — «отношения»).

**Tableau** — программное обеспечение для интерактивной бизнес-аналитики и визуализации данных. Оно помогает провести глубокий анализ большого количества информации и представить результаты в удобной и интуитивно понятной форме.

**TensorFlow** — открытая программная библиотека для машинного обучения, разработанная компанией Google для решения задач проектирования и обучения нейронных сетей.

**Docker** — программное обеспечение для автоматизации, развёртывания и управления приложениями в средах с поддержкой контейнеризации, контейнеризатор приложений.

**Алгоритм** — конечная совокупность точно заданных правил решения некоторого класса задач или набор инструкций, описывающих порядок действий исполнителя для решения определённой задачи.

**Линейная алгебра** — раздел алгебры, изучающий объекты линейной природы: векторные пространства, линейные отображения, системы линейных уравнений. Среди основных инструментов, используемых в линейной алгебре — определители, матрицы, сопряжение.

**Математический анализ** — совокупность разделов математики, соответствующих историческому разделу под наименованием «анализ бесконечно малых», объединяет дифференциальное и интегральное исчисления.

**Оптимизация** — задача машинного обучения по выбору набора оптимальных гиперпараметров для обучающего алгоритма.

**Статистика** — отрасль знаний, наука, в которой излагаются общие вопросы сбора, измерения, мониторинга, анализа массовых статистических данных и их сравнение; изучение количественной стороны массовых общественных явлений в числовой форме.

**Хакатон** (англ. hackathon от hacker «хакер» + marathon «марафон») — форум для разработчиков, во время которого специалисты из разных областей разработки программного обеспечения (программисты, дизайнеры, менеджеры) сообща решают какую-либо проблему на время.