# **ГЛАВА 1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ**

# **Data Science и его инструменты**

Сегодня миром правят данные. Эта тенденция способствовала огромному спросу на специалистов по обработке данных. Специалисты по обработке данных, также data scientist-ы, помогают компаниям принимать решения, основанные на данных, для улучшения их бизнеса. Data scientist-ы являются представителя науки о данных.

Data Science (наука о данных) - это сочетание нескольких дисциплин, которые используют статистику, анализ данных и машинное обучение для анализа данных и извлечения из них знаний и идей. Это также поиск закономерностей в данных с помощью анализа и составление прогнозов на будущее.

Используя науку о данных, компании могут создавать:

* Лучшие решения (должны ли мы выбрать A или B)
* Прогнозирующая аналитика (что будет дальше?)
* Выявление закономерностей (поиск закономерностей или, возможно, скрытой информации в данных)

Примеры того, где наука о данных необходима:

* Для планирования маршрута: Для определения наилучших маршрутов доставки
* Для прогнозирования задержек рейсов /судов / поездов и т.д. (с помощью прогнозного анализа)
* Для создания рекламных предложений
* Для определения наиболее подходящего времени доставки товаров
* Для прогнозирования доходов компании на следующие годы
* Для анализа пользы обучения для здоровья
* Предсказать, кто победит на выборах
* Наука о данных может применяться практически в любой сфере бизнеса, где имеются доступные данные. Примерами являются:
* Потребительские товары
* Фондовые рынки
* Промышленность
* Политика
* Логистические компании
* Электронная коммерция

Существуют следующие data science инструменты, которыми специалисты по обработке данных пользуются в различных задачах:

* Языки программирования (Python, Java, Scala, R, Julia, Matlab);
* Среда разработки (Jupyter Notebook, JupyterLab, JetBrains PyCharm, Visual Studio Code);
* Версионирование кода и данных (Git, GitLab, DVC);
* Управление экспериментами (MLflow, ClearML);
* Машинное обучение (scikit-learn, CatBoost, XGBoost, LightGBM, TensorFlow, PyTorch);
* Библиотеки анализа данных (Matpotlib, Seaborn, Plotly, Numpy, Pandas).

Комбинация инструментов и их вариативность зависят от поставленной задачи.