

# Машинное обучение

Практика 1: Обзор курса и введение

# Контакты

Фомченко Александр Валерьевич

- Email: [alexander.mirea@mail.ru](mailto:alexander.mirea@mail.ru)
- <https://github.com/alexander-mirea/ML2025>

# Описание раздела

1. Обзор курса
2. Административные вопросы
3. Обзор технологий, которые будут использоваться в курсе

# О курсе

- Есть лекции и семинары
- У нас упор в практику (как ни странно)
  - Анализ данных, препроцессинг, извлечение знаний
  - Теория есть, но поверхностно
- Нет нейросетей – обойдёмся без них и на практике

# Формат практик

- В этом семестре N практических заданий, их сдача упрощает сдачу зачета
- Зачет – теоретический вопрос (в формате билета)
- Сдача практических заданий осуществляется в компьютерном классе, я могу дать дополнительное задание или задать вопрос по коду

# Обзор технологий курса

- Python 3
- Jupyter Notebook (Google Colab)
- Kaggle.com

Рекомендую вспомнить почту gmail и посмотреть как выглядит kaggle.

# Классическое Обучение



# Основные виды машинного обучения

Простые данные  
Понятные признаки

**Классическое  
обучение**

Когда качество  
правда проблема

**Ансамбли**

Сложные данные  
Непонятно где  
признаки  
Есть вера в чудо

**Нейросети  
и глубокое  
обучение**

Данных нет,  
но есть среда,  
с которой можно  
взаимодействовать

**Обучение  
с подкреплением**

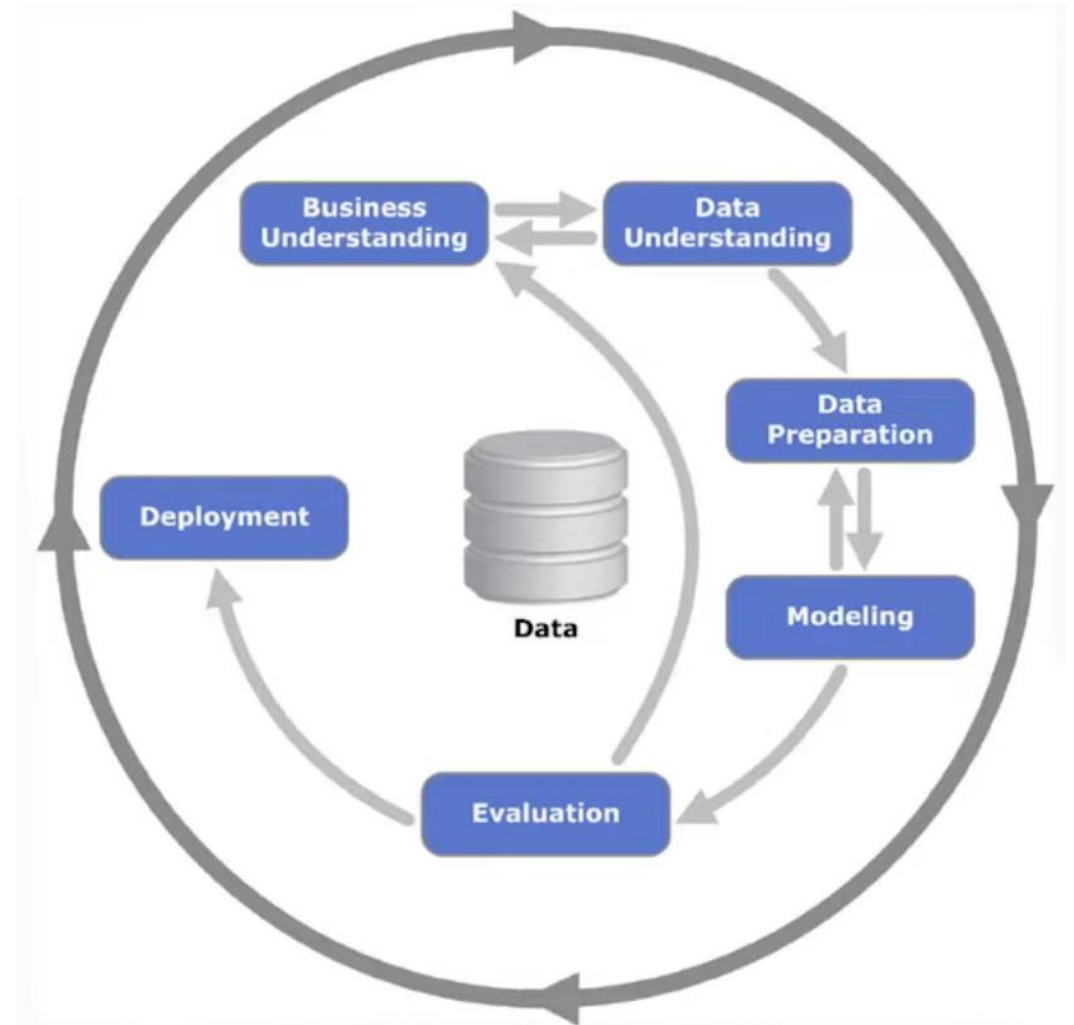
Вечные конкуренты



# Обзор процесса машинного обучения

- Постановка задачи
- Предобработка данных
- Моделирование
- Оценка модели
- Внедрение и эксплуатация

Справа модель CRISP-DM 1999 года.



# Постановка задачи

- «Задача поставлена, если у неё есть ДНК (Дано, Найти, Критерий)»  
– © Воронцов К.В.
- Делят на бизнес-анализ и анализ задачи в терминах машинного обучения
- Пытаемся выразить задачу, понять, как можно ее решить, и определить, поможет ли нам в этом машинное обучение
- Уже здесь надо думать, как это решение потом внедрять и использовать (и можно ли)

# Предобработка данных

- Сбор данных
- Анализ данных
- Предобработка данных
  - Масштабирование (нормализация)
  - Устранение пропусков/дубликатов
  - Устранение выбросов
  - Преобразование признаков
  - Снижение размерности
- Конструирование признаков (вручную или deep learning)

Успех вашей работы зависит от данных не меньше, а то и больше, чем от кода.

# Построение модели и оценка

- Выбор модели обучения
- Настройка гиперпараметров
- Сравнение, выбор лучшей



# Внедрение и эксплуатация

- Тестирование
- Масштабирование
- Мониторинг модели
  - Доработка в случае ухудшения показателей