

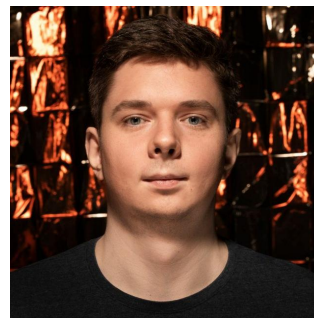
# Курсы по машинному обучению

от команды предиктивной аналитики myTracker

- 10 лекций (4 блока по 2-3 лекции)
- 2 домашних задания
- 10 практических занятий
- 2 классных лектора



Антон Лоскутов



Иван Мажаров

- **Telegram**  
Для основных новостей по курсу
- **Zoom**  
Исключительно для онлайн-лекций
- **Discord**  
Для вопросов по курсу, общения и обсуждения заданий



<https://t.me/joinchat/EK0aNkp08-imEpSTKM9ANg>

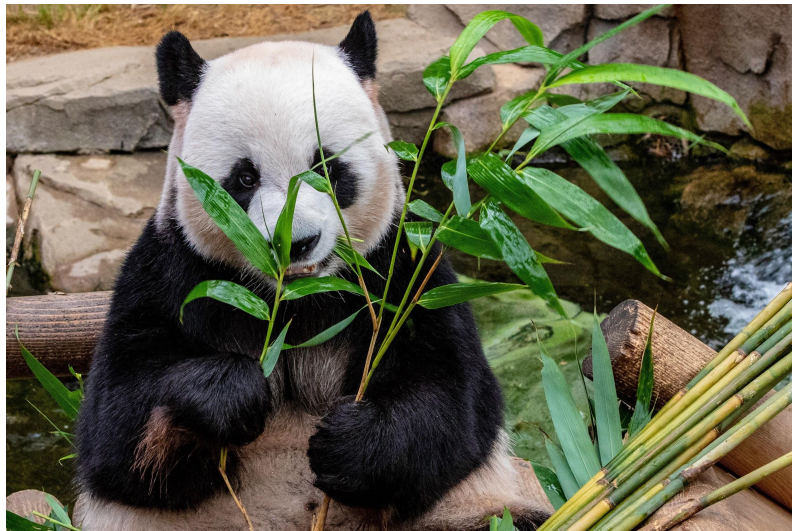


<https://discord.gg/KjFksXP>





- Базовые навыки работы с табличными данными
- Основные функции Pandas, примеры использования
- Практика на знакомство с данными



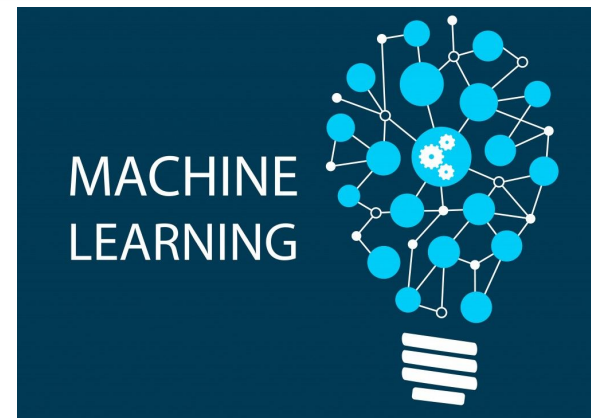
- Копирование в pandas, оптимальное использование функций, оптимизация памяти
- Работа с датами
- Высокопроизводительные типы данных



- Основные функции построения графиков в pandas, matplotlib, seaborn, plotly
- Визуальный анализ данных, объяснения графиков
- Практика на построение графиков и анализ результатов

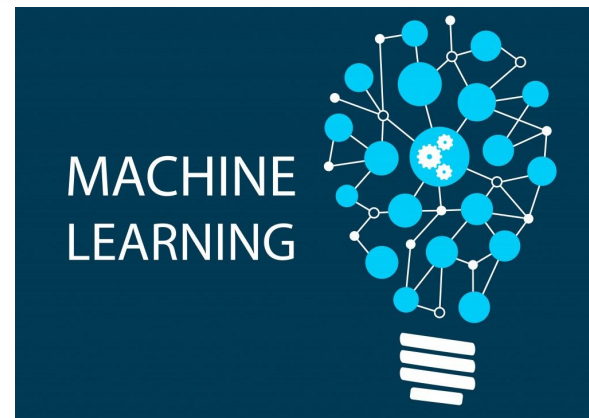
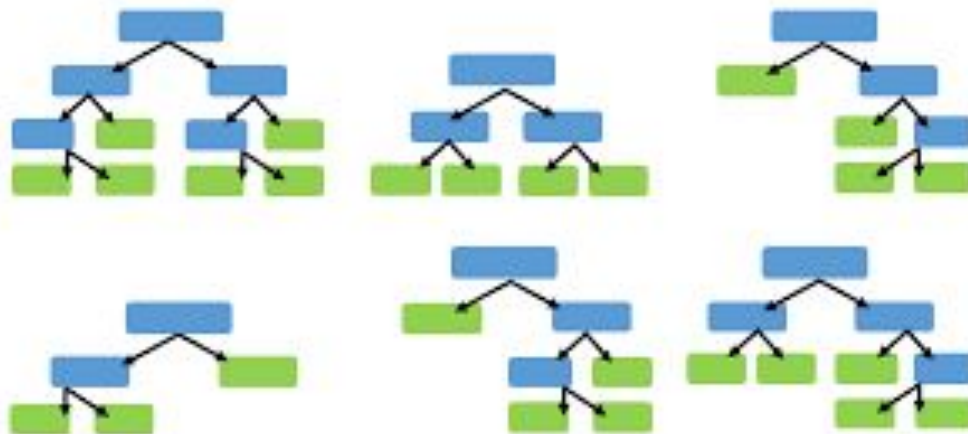


Соревнование на небольшой датасете для проверки ML навыков.

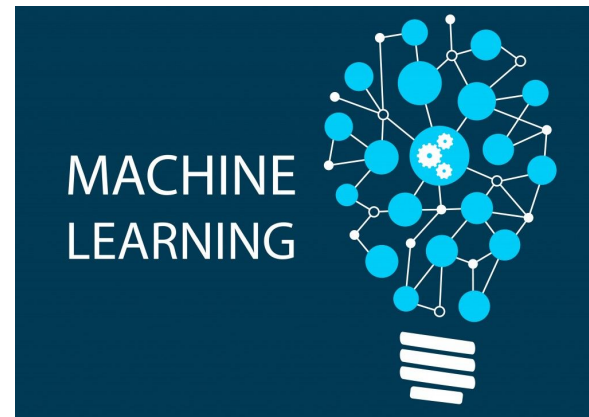


- Основы машинного обучения
- Деревья решений, энтропия
- Практика на знакомство с библиотекой Scikit-learn, знакомство с основными методами API библиотеки

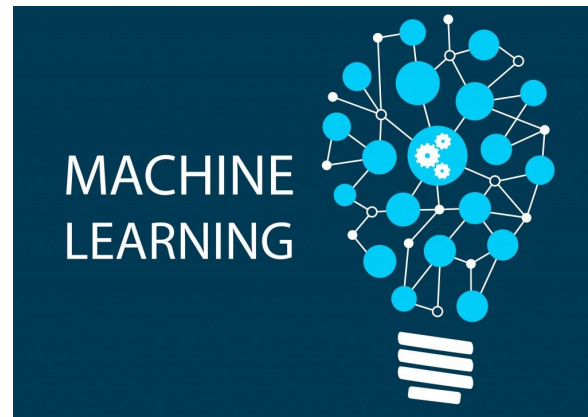




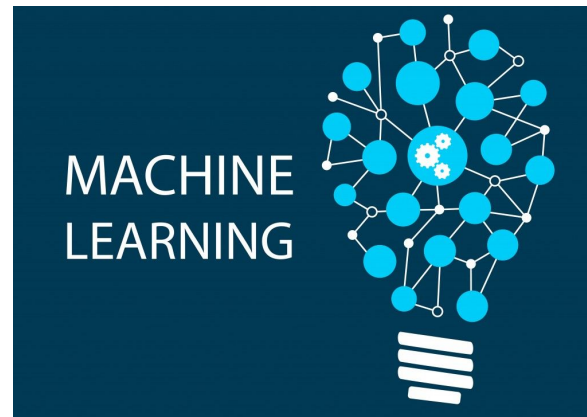
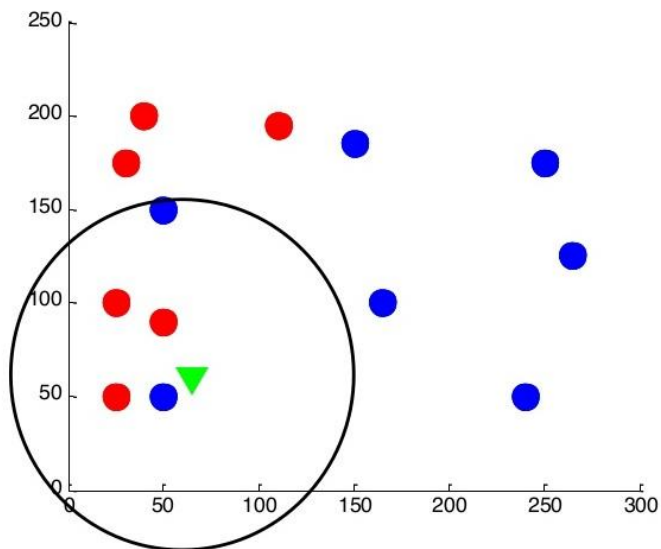
- Композиция алгоритмов, bagging, bootstrap
- Случайный лес
- Практика на применение случайного леса и оценку важности признаков



- Линейные модели классификации
- Кросс-валидация
- Практика на логистическую регрессию



- Обучение без учителя
- Principal Component Analysis
- Кластеризация



- Метод ближайших соседей
- Потенциальные методы
- Методы Парзенковского окна



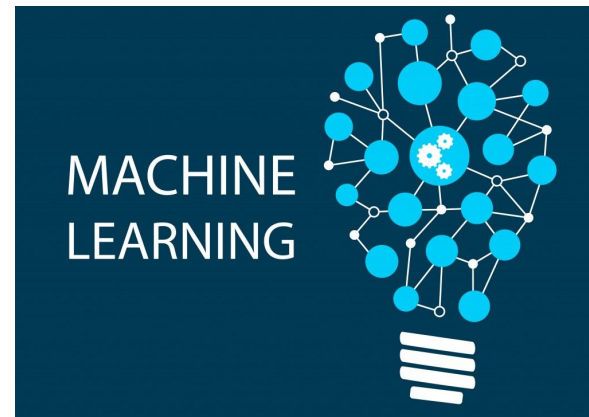
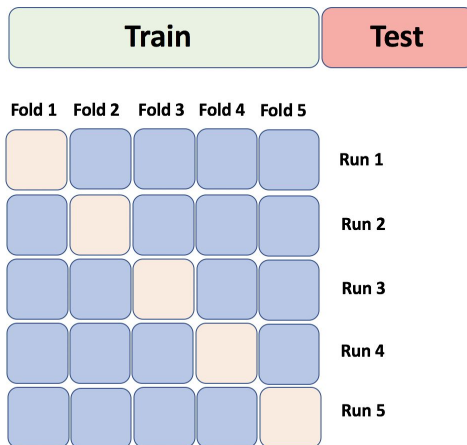
Более серьезное соревнование, в котором надо применить все накопленные навыки

# Feature encoding



Train

Validation



- ОН-кодирование
- Bag Of Words
- Kernel Trick



Yandex  
CatBoost

Microsoft  
LightGBM

*dmlc*  
***XGBoost***

- Современный взгляд на бустинг
- Теоретические основы
- Лучшие на сегодня реализации