

# Введение в машинное обучение

# Программа

1. Введение и регрессия
2. Классификация
3. Валидация
4. Ансамбли
5. Кластеризация
6. Нейронные сети
7. Нейронные сети
8. Нейронные сети

# Оценивание

## Накопленная:

- домашка по каждой теме

## Экзамен:

- Inclass соревнование на каггле на 1 неделю

Оценка = **Накопленная** (60 %) + **Экзамен** (40 %)

# Преподаватели

Михаил Нефедов - Вышка, Сбербанк

[@Nefedov\\_m](#)

Татьяна Шаврина - Вышка, Сбербанк

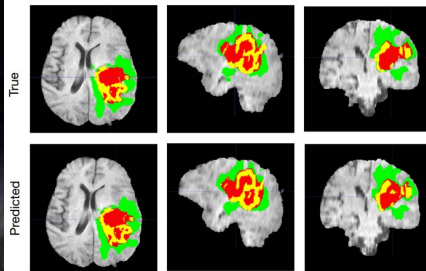
[@rybolos](#)

Мария Пономарёва - ТиПЛ, Вышка, ABBYY

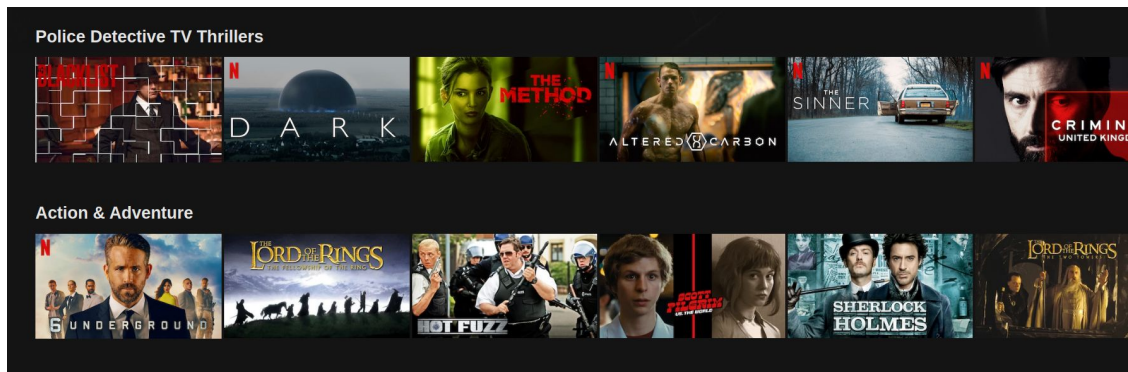
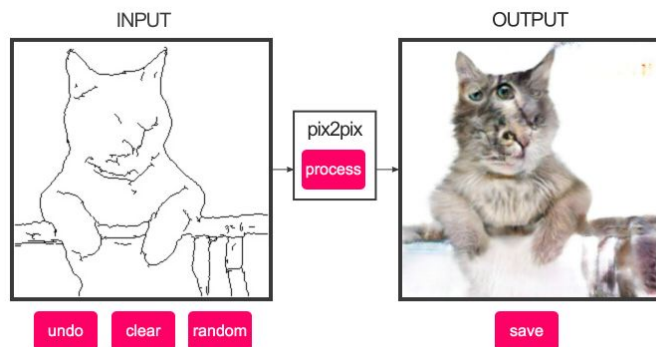
[@MashPo](#)

# ML Today

**This course on machine learning will be about** methods to extract feature-rich features from structured data, and how to combine these features with validation/classification/predictive algorithms. The goal of this course is to give an introduction to machine learning from a small, safe starting point. Students should be able to recognize and work with some machine learning techniques. Previous experience working with data will be helpful. Prerequisites: Basic programming skills

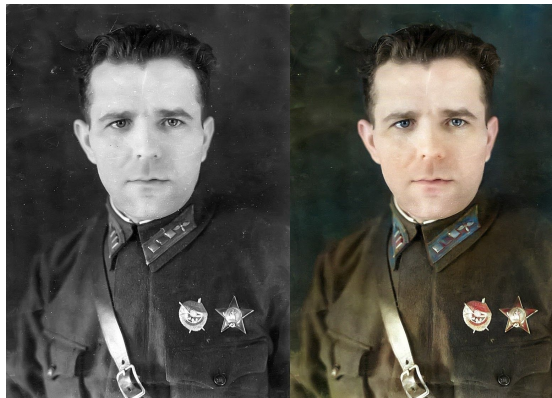


# ML Today



# ML Today

"Если кто-то сможет обеспечить монополию в сфере искусственного интеллекта, то последствия нам всем понятны - тот станет **властелином мира**"



Good News! The Elements of AI course will soon be available in all EU languages. [Read more here](#) →

## Welcome to the Elements of AI free online course!

Join over 320,000 other people learning about the  
basics of AI.

Select language  English ↓

[Start the course](#) →

THEY TOOK 'ER JOBS!



Привет, я Алиса



# Не МЛ или плохой МЛ



умею, и я расскажу подробнее.

Если что-то пойдет не так, скажите: «Хватит» или «Позови человека».

15:24

Одолжи сотку.

✓ 15:25



Не работает отпечаток пальца

✓ 11:20

Вам ответит Олег.



Олег

Пососите потом просите.

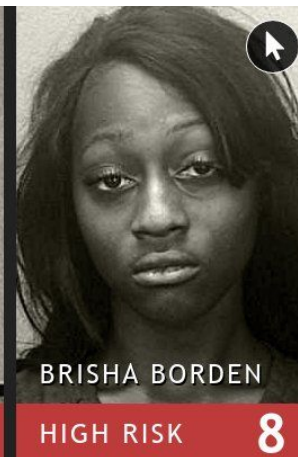
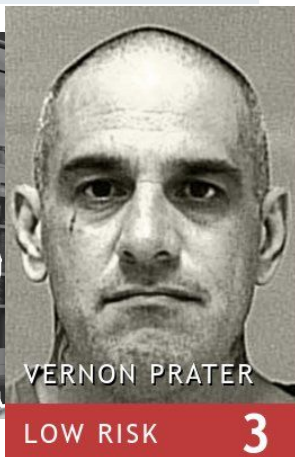
15:26



Олег

Пальцы бы вам отрезать.

11:21





# Основные понятия

Объекты

Признаки - формальное описание объектов

Целевая переменная - то, что мы хотим предсказать

Обучающая/тестовая выборка - матрица объекты-признаки и целевая переменная

Модель/Алгоритм

Функция потерь - критерий качества, использующийся при обучении

Метрика - критерий качества, использующийся для оценки

# Два типа целевой переменной

Класс или численное значение

1 - Классификация

2 - Регрессия

Практические задачи приводятся к этим двум типам

# Предсказание рака

Объекты - люди

Признаки - возраст, курение, алкоголизм, боль, результаты анализов, мрт

Целевая переменная - наличие/отсутствие рака

Обучающая выборка - результаты диспансеризации граждан

Метрика - процент правильных диагнозов

**Это задача классификации**

# Предсказание прибыльности при открытии кафе

**Объекты** - такие же (если это сеть) или похожие кафе

**Признаки** - рядом с метро/офисами/парком/трассой/конкурентом

**Целевая переменная** - прибыль по итогам первого года работы

**Метрика** - средняя абсолютная ошибка

Это задача регрессии

# Типы признаков

Численные - температура, скорость, задолженность

Бинарные - наличие собаки, присутствие на паре, наличие долга

Категориальные - пол, цвет глаз, уровень образования

Порядковые - оценка, звание

\**Картинки* - фото, рентген

\**Текст* - описания, твиты, названия товаров, имена

Для обучения все признаки приводятся к численным

# Функция потерь

Чаще всего используется средняя квадратичная функция потерь

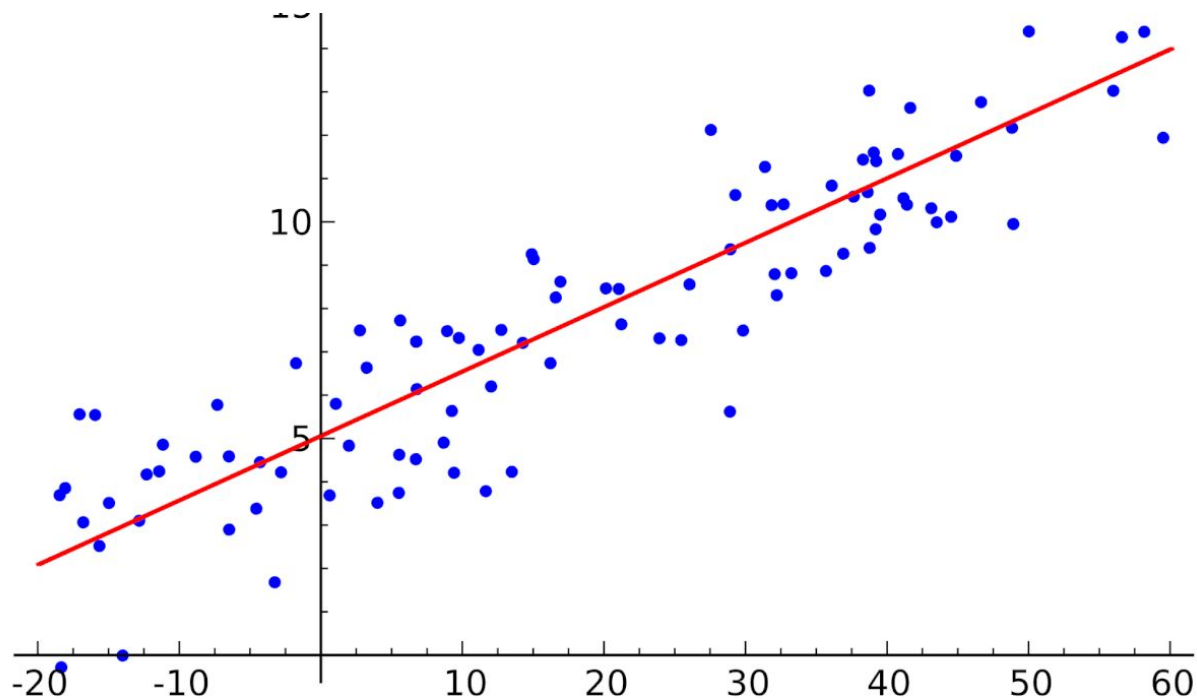
$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Если все предсказания совпадают, mse - ноль

Квадрат нужен, чтобы не учитывать знак (модуль тоже используется)



# Линейные алгоритмы



Линейная регрессия

SVM

Lasso

Ridge

ElasticNet

# Метрики регрессии

**RMSE** - квадратный корень от средней квадратичной ошибки

**MAE** - средняя абсолютная ошибка

**MAPE** - Средний процент отклонения

**R2** - [Коэффициент детерминации](#)