

# Преподавание анализа данных и машинного обучения для студентов всех направлений

Евгений Соколов

Академический руководитель ПМИ

НИУ ВШЭ

Бакалавриат «Прикладная  
математика и информатика»

# Общая характеристика

- Программа реформирована в 2014 году
- Основные направления подготовки:
  - Машинное обучение и анализ данных
  - Распределённые системы
  - Теоретическая информатика
  - Прикладная математика

# Общая характеристика

- Программа реформирована в 2014 году
- Основные направления подготовки:
  - **Машинное обучение и анализ данных**
  - Распределённые системы
  - Теоретическая информатика
  - Прикладная математика

# Какая математика нужна?

- Математический анализ — оптимизация и анализ моделей
- Линейная алгебра — ускорение вычислений, анализ моделей, методы на основе матричных вычислений
- Теория вероятностей и математическая статистика — вероятностные методы, аналитика и эксперименты
- Дискретная математика — работа с графовыми данными

# Младшие курсы

- Математический анализ
- Линейная алгебра
- Теория вероятностей и математическая статистика
- Дискретная математика
- Python
- Алгоритмы и структуры данных

# Как понять, хочется ли этим заниматься?

- Лабораторные работы на линейной алгебре
- Курс по матричным вычислениям
- Проекты на втором курсе

# Программные проекты

- Выполняются на 2 курсе с ноября по май
- Есть много тем, связанных с машинным обучением
- Изучение некоторых разделов и использование стандартных библиотек
- Акцент на процессе разработки и инструментах



# Специализация «Машинное обучение и приложения»

6 обязательных дисциплин:

- Машинное обучение 1
- Прикладная статистика в машинном обучении
- Машинное обучение 2
- Методы оптимизации для машинного обучения
- Глубинное обучение
- Машинное обучение на больших данных

# Машинное обучение 1

- Длится один семестр
  - Около 8 очень объёмных практических ДЗ
  - Проверочные работы на семинарах (примерно каждые 2 недели)
  - Контрольная работа и экзамен
  - Бонусные ДЗ
- 
- [http://wiki.cs.hse.ru/Машинное обучение 1](http://wiki.cs.hse.ru/Машинное_обучение_1)

# Машинное обучение 1 — задания

- Имплементация методов своими руками
- Работа со сложными данными, поиск проблем в них
- Исследование данных
- Изучение ограничений методов

# Научно-исследовательский семинар

- Чтение статей и обсуждение на занятиях
- Обсуждение курсовых и ВКР
  
- Проверочные по докладам
- Кросс-рецензирование КР/ВКР

# Курсовые работы и ВКР

- Два формата: исследовательский и программный
- Возможны групповые работы

# Курсы по выбору

- Моделирование временных рядов
- Анализ данных в бизнесе
- Байесовские методы машинного обучения
- Глубинное обучение в обработке звука
- NLP
- Генеративные модели в машинном обучении
- Обучение с подкреплением
- И т.д.

Проект Data Culture

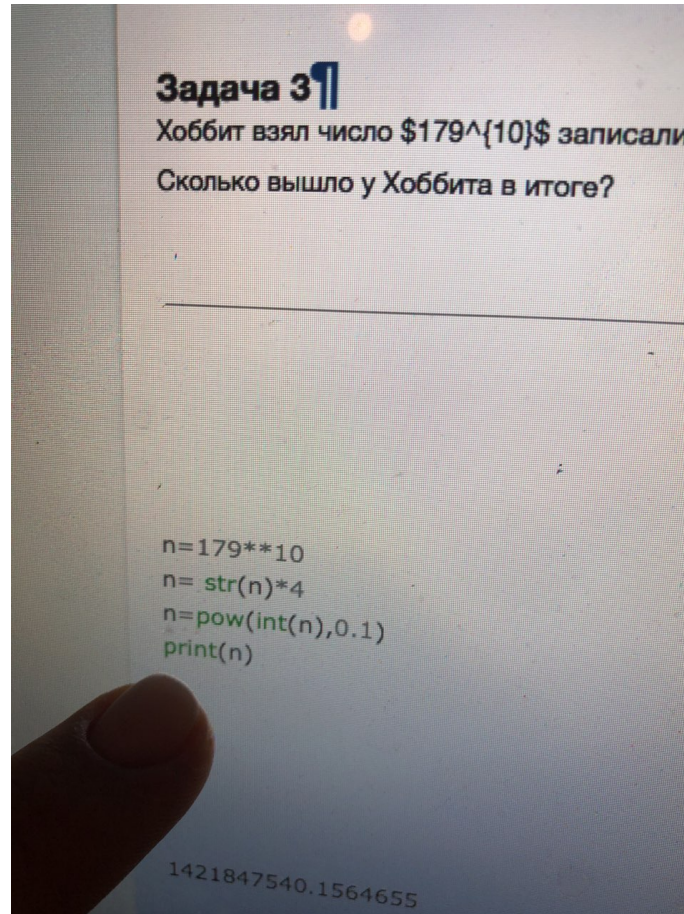
# Data Culture

- Общеуниверситетский проект с 2017 года
- Цель — развитие компетенций по программированию и работе с данными у всех студентов НИУ ВШЭ



Уровень 1: начальный

# Зачем?



# Цифровая грамотность

- Как устроен компьютер?
- Что такое большие данные?
- Основы компьютерной безопасности
- Основы медиаграмотности
- Библиография и работа с научной литературой
- Офисные технологии

# Модульность

- Основной блок — 9 недель
- Выборные блоки: статистика, корпусные технологии и т.д.
- Сейчас: образовательная программа выбирает состав курса
- Скоро: студент выбирает состав курса

# Чем плох онлайн-курс?

- Списывание
- Нет возможности развивать аналитические навыки
- Нет возможности оказать «психологическую поддержку» студентам

# Семинарская поддержка

- Проектная работа
- Разбор сложных тем
- Консультации

# Семинарская поддержка



Уровень 2: базовый



# Цели

- Обучить основам программирования
- Обрисовать возможности аналитики данных и машинного обучения, не прибегая к сложной математике

# Обучение Python

- Есть хорошие записи лекций и онлайн-курсы
- Сложно искать большое количество хороших семинаристов
- Важная часть обучения — большое количество практики

# Чем плохо онлайн-обучение?

- Списывание
- Сложно выйти за рамки курса
- Сложно поверить в себя

# Blended-обучение

- Очные консультации и контроль знаний
- На консультациях разбираются типовые задачи
- Регулярный контроль знаний стимулирует учиться

# Обучение анализу данных

- Источники данных и взаимодействие с ними
- Работа с табличными данными, их предобработка
- Визуализация данных
- Построение моделей готовыми инструментами
- Статистика и обработка экспериментальных данных

# Для кого?

- Менеджмент и маркетинг
- Международные отношения
- Экономика
- Политология
- Медиакоммуникации
- ...

Уровень 3: продвинутый

# Что изучать?

- Методы машинного и глубинного обучения
- Работа с данными сложной структуры
- Прикладная статистика в машинном обучении



# Майнор ИАД

- Отбор по рейтингу
- Студенты со всех образовательных программ
- Курсы:
  - Python
  - Основы машинного обучения
  - Основы глубинного обучения
  - Прикладные задачи анализа данных
- Слушателям открываются карьерные пути в Data Science

# Продвинутый уровень

- Физика: обработка данных физического эксперимента
- Политология: машинное обучение для анализа текстов и социальных сетей
- Маркетинг: машинное обучение для анализа клиентских данных
- ...

# Резюме

- Для профильных студентов важны полное погружение в методы и развитие исследовательских навыков
- Для студентов нетехнических направлений лучше начинать с освоения готовых инструментов без попыток понять математику за ними
- [esokolov@hse.ru](mailto:esokolov@hse.ru)