



Выпускной проект

Господинов Георгий



Программа курса

Дата	Занятие	Баллы
29.09	Введение в анализ данных	4
06.10	Задачи классификации и регрессии	10
13.10	Оценка качества и работа с признаками	7
20.10	Обучение без учителя	2
27.10	Ансамбли моделей	15
03.11	Анализ социальных сетей	7
10.11	Работа с текстовыми данными	7
17.11	Рекомендательные системы	7
24.11	Семинар	5
01.12	Нейронные сети. Основы	2
08.12	Сверточные нейронные сети. Часть 1	7
15.12	Сверточные нейронные сети. Часть 2	2
22.12	Итоговое занятие(проект)	25



Что включает проект

- Постановка задачи
- Сбор данных
 - Открытый датасет
 - API
 - Scraping
- Выбор метрики качества
- Описание данных (Exploratory Data Analysis)
- Создание baseline-модели \Rightarrow оценка качества
- Первая модель \Rightarrow оценка качества
- Анализ важности признаков и ошибок модели
- Улучшения
 - добавление новых признаков
 - более сложные алгоритмы
 - чистка данных
 - регуляризация
- Сравнение опробованных подходов, выводы

Регламент

- Команда 1-3 человека
- Презентация, ~7 слайдов
- 10 минут на доклад
- 5 минут на вопросы
- Каждый участник должен рассказать не менее одного слайда



Примерный план презентации

- Название проекта, состав команды, распределение задач
- Описание данных и постановка задачи
- Описание возникших проблем
- Итоговое решение
- Выводы

Проекты прошлого года

Тема проекта	Количество участников	Данные
Выделение сигнала флуоресценции углеродных нанокompозитов на фоне аутофлуоресценции урины	2	Датасет из 1460 спектров флуоресценции
Предсказание суммы покупки	1	https://www.kaggle.com/c/msusums
Рекомендация групп ВК	3	https://vk.com/overhear_msu
Прогнозирование вероятности возврата авиабилета клиентом	3	https://boosters.pro/championship/onetwotrip_challenge
Поиск аномальных объектов в звездном каталоге Caltech ZTF	3	https://irsa.ipac.caltech.edu/docs/program_interface/ztf_api.html
Рекомендательная система, основанная на изображениях товара	1	https://www.vsemayki.ru/
Предсказание зарплаты по тексту объявления	1	https://www.kaggle.com/c/cmc-msu-machinese-learning-spring-14-15-competition-1
Предсказание жеста(движение рукой) по видео	1	Собирался самостоятельно

Идеи проектов

Тема проекта	Данные
Эмоциональная окраска отзывов к фильмам / книгам / ...	Scrapping / готовый датасет
Классификация статей по жанрам	Scrapping / готовый датасет
Классификация музыки по жанрам	https://zenodo.org/record/1302992#.XdOaQNFRWCh
Рекомендательная система музыки	http://marsyas.info/downloads/datasets.html
Предсказание пола/возраста/... по фото	http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html
Соревнование Россельхозбанк (19 ноября — 11 декабря)	https://agro-code.ru/datascience/
Соревнование Альфабанк (12 декабря — 20 января)	https://alfabattle.ru/2
Соревнование Сибур (21 ноября — 19 декабря)	https://sc2020.ai-community.com/

<https://habr.com/ru/post/452392/>



План действий

- Выбрать проект
- Собрать команду
- Записаться в google-таблицу (ссылка появится в telegram-канале курса)
- Реализовать проект
- Выступить на финальном занятии курса

Спасибо за внимание!

Господинов Георгий

✉ g.gospodinov@corp.mail.ru

📩 georgygospodinov

Machine Learning Development Lifecycle

