|  |
| --- |
| Если у вас уже была/есть научная работа, расскажите о чем она |
|  |

Пока что вне рамок своих курсовых работ я не занимался научной работой. На первом курсе я писал о модальной логике и ее приложении к топологии (на модальном языке можно записывать определения и свойства топологических объектов). На втором курсе я исследовал некоторую подгруппу группы SL(2, Z), искал ее порядок, образующие, фундаментальную область, рассматривал ассоциированную с ней алгебру модулярных форм. Сейчас я пишу курсовую работу о подходах к построению рекомендательных систем на основе нейросетей.

|  |
| --- |
| Расскажите о своем опыте программирования |
|  |

Я окончил школу №179, в которой с 9 по 11 класс изучал язык программирования Python, который является моим основным языком. Также я дважды был в ЛКШ, в параллелях C (Python) и B (C++). Проходил на финал Всероссийской олимпиады школьников по программированию. Сейчас я прохожу специализацию Algorithms на курсере от профессора Тима Роугардена. Остальной мой опыт программирования связан с машинным обучением, о нем я написал с соответствующем разделе анкеты.

|  |
| --- |
| Откуда вы узнали про ШАД? |
|  |

Три моих однокурсника учатся в ШАД, я узнал от них.

|  |
| --- |
| Зачем вы поступаете в ШАД? Расскажите чему вы ожидаете научиться в ШАД, для чего вам эти знания, какое отделение привлекает больше всего. Чем честнее ответ - тем лучше. |
|  |

Я убежден, что знания и умения, которые я могу получить в ШАД, имеют очень большую ценность. Я планирую строить карьеру в сфере ML, Data Science и AI Safety. На основании уже имеющихся данных я считаю, что мне очень интересны эти области знаний и что мне удастся освоить программу ШАД. Я бы хотел учиться на направлении, в котором математике уделяется существенное внимание, поэтому для меня в приоритете Data Scientist. Сейчас меня очень интересуют случайные процессы и теория вероятностей, также я бы хотел углубить свои знания языка Python, теории игр и статистики, освоить RL, NLP, байесовские методы.

|  |
| --- |
| Изучали ли вы раньше машинное обучение/анализ данных? Каким образом? Какие навыки удалось приобрести, какие проекты сделать? |
|  |

Да. Во-первых, я прошел курс на курсере «Введение в машинное обучение» от НИУ ВШЭ и ШАДа. На нем я приобрел навыки работы с библиотеками pandas и numpy, узнал о базовых алгоритмах машинного обучения. Также на курсере я прошел специализацию Deep Learning от профессора Andrew Ng, благодаря чему знаком с основными архитектурами нейронных сетей и фреймворком keras.

В октябре-декабре 2018 года я освоил курс «Введение в машинное обучение», который провела компания Mail.Ru Group. Там я научился базовым навыкам решения задач с использованием регрессии, классификации, кластеризации, ансамблирования и обработки текстов. Во время прохождения курса я сделал проект по классификации статей на хабрахабре по темам, получил при этом навыки работы с word2vec, fasttext, wovpal wabbit.

Также я прошел курс по машинному обучению, который прошел в этом учебном году на Факультете Математики. Одним из домашних заданий было написать модель, предсказывающую таргет по искусственно созданным данным, и я много исследовал признаки на скоррелированность, а потом пробовал разные ансамблирующие библиотеки (catboost, xgboost, lightgbm).

Я участвовал в направлении «Искусственный Интеллект» олимпиады «Я - Профессонал» в этом году, являюсь призером. Также я принял участие в зимней школе по этому направлению. На хакатоне, который проводился во время школы, моя команда заняла первое место в своем направлении. Мы решали задачу от ИО РАН по детектированию мезоциклонов на спутниковых снимках с помощью архитектуры SegNet.

Сейчас я стажируюсь в Тинькофф Банке по специализации ML Engineer (начал 18 февраля, закончу 15 мая). Я разрабатываю рекомендательную систему. Сначала я сделал несколько моделей на основе Neural Matrix Factorization, но обучение было медленным и сеть тяжелой, поэтому команда решила экспериментировать с подходом DSSM. Пробовал много разных функций потерь, перебираю функции активации, число слоев, общую топологию сети. Данные «плохие», приходится тщательно продумывать, каким образом выкидывать аутлаеров, отбирать, конструировать и нормализовать признаки. Пока что не было достигнуто качества, при котором можно было бы выпускать модель в продакшн. Пишу на питоне с использованием keras.