1 Укажите тему вашей дипломной или курсовой работы и фамилию научного руководителя (если таких работ несколько, выберите наиболее интересную, если их нет — укажите это).

Ответ Сейчас, на третьем курсе, я пишу курсовую под руководством Дмитрия Игоревича Игнатова на тему «Модели глубинного обучения для рекомендательных систем».

2 Чему посвящены ваши исследования и учебные проекты? Расскажите о проекте, в котором вы участвовали. Нам хотелось бы узнать о технологиях, объёме проекта, вашей роли в нём и самой увлекательной задаче, которую вы решили

Ответ 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 языке Python с использованием фреймворка keras.

Во время прохождения курса «Введение в Машинное Обучение» от компании Mail.Ru Group я написал классификатор статей на хабрахабре. Итоговая модель была ансамблем из fasttext, word2vec и классификатора wowpal wabbit по подходу BOW.

3 Расскажите о своих учебных интересах. Какие из прослушанных курсов вам больше всего понравились?

Ответ Мне интересна теория вероятностей (в институте я уже прослушал два качественных курса на эту тему – «Введение в теория вероятностей» и «Математическая статистика», сейчас заканчиваю еще два – «Случайные процессы» и «Финансовая математика»). Я понимаю, что многие процессы в природе, хотя по сути случайными не являются, но хорошо моделируются случайными величинами. Иными словами, теория вероятностей – хороший инструмент для описания мира, поэтому я хочу ей заниматься.

Также я изучаю машинное обучение. 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).

4 Опишите, какими задачами вам было бы интересно заниматься во время стажировки.

Я бы хотел заняться NLP или Computer Vision. У меня есть знания о работе основных подходов к NLP из курса от Mail.Ru Group (большая часть времени была посвящена именно этой теме), а также я знаком с работой CNN и RNN из специализации от Andrew Ng на курсере.

5 Перечислите другие языки программирования и технологии, которые вы используете (в порядке предпочтения). Оцените уровень владения ими.

Ответ Я пишу на языке Python. Свободно использую библиотеки numpy, pandas, sklearn, среду jupyter-notebook. Для работы с нейросетями применяю фреймворк keras. Базово владею языком C++.

6 Расскажите о своих достижениях — медалях, дипломах, сертификатах, об участии в олимпиадах, конкурсах, хакатонах, соревнованиях ACM и так далее.

Ответ В этом году я стал призером олимпиады «Я - Профессионал» по направлению «Искусственный Интеллект». Также в этом году моя команда заняла первое место на хакатоне в Зимней Школе по Искусственному Интеллекту, мы лучше всех решили задачу по детектированию мезоциклонов на спутниковых снимках с помощью архитектуры SegNet.

В 10 классе (2014-2015 уч.г.) я участвовал в финальном этапе ВсОШ по информатике.

7 Приведите пример вашего кода — лучше всего в виде ссылки на аккаунт на GitHub, Bitbucket или на другом подобном сайте.

Ответ Ссылка на проект по «SNA Hackathon – Коллаборативная система» от платформы MLBootCamp.

<https://drive.google.com/open?id=1fDgtsYfyVgF7fOIRQdQuULUm0jixtwfg>

Ссылка на гитхаб <https://github.com/mishazybin>

8 Почему вас заинтересовала стажировка в Яндексе и чего вы от неё ждёте?

Ответ Я хочу строить карьеру в сфере Machine Learning и AI Safety. Я считаю, что стажировка в Яндекс позволит мне приобрести релевантный карьерный капитал. Я хочу порешать интересные задачи, поприменять на деле изученные мной алгоритмы.

9 Что-то ещё? Не стесняйтесь рассказать о себе.

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.

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.