



Чубов Илья Денисович

Мужчина, 21 год, родился 2 августа 2003

+7 (989) 6302782 — предпочитаемый способ связи • Tg: @younlean

Проживает: Москва

Гражданство: Россия, есть разрешение на работу: Россия

Не готов к переезду, не готов к командировкам

Желаемая должность и зарплата

Data Science

Специализации:

- VI-аналитик, аналитик данных
- Программист, разработчик

Занятость: частичная занятость, стажировка

График работы: гибкий график

Желательное время в пути до работы: не имеет значения

Опыт работы — 8 месяцев

Май 2024 —
настоящее время
8 месяцев

DCS

Москва, dc-s.ru/

ML-engineer

Основные обязанности и достижения:

Проводил эксперименты с моделями (CatBoost, LinearRegeression, RandomForest), настраивал и оптимизировал гиперпараметры. Логировал метрики, параметры и артефакты в MLflow и wandb. Работал с контейнерами, используя Docker, автоматизировал окружение. Занимался предобработкой данных: очистка, трансформация. Взаимодействовал с командой аналитиков для улучшения модели.

Технологии и инструменты:

Python (pandas, scikit-learn, XGBoost, RandomForest)
SQL, polars
Docker
MLflow для трекинга экспериментов и версионирования моделей
Git

Образование

Неоконченное высшее

2026

Национальный исследовательский ядерный университет "МИФИ", Москва

ИФТЭБ, Безопасность информационных технологий в правоохранительной сфере

Повышение квалификации, курсы

2024

Анализ данных и машинное обучение

НИЯУ МИФИ, Анализ данных и машинное обучение

2023

Искусственный интеллект

Samsung Innovation Campus, Data Science

2023

Deep Learning School

МФТИ, Обработка текстов и аудиосигналов

Тесты, экзамены

2024

Хакатон "Nuclear IT Hack"

НИЯУ МИФИ, Data Science

2024

Хакатон "IT Academy Hack"

Samsung Innovation Campus , Data Science

Навыки

Знание языков

Русский — Родной

Английский — B1 — Средний

Навыки

ML Engineer

NLP

CV

Python

Matplotlib

Seaborn

Scikit-learn

TensorFlow

Keras

OpenCV

Kaggle

GitHub

Теория вероятностей

Математическая статистика

pandas, numpy

SQL

Statsmodels

MLflow

Docker

MySQL

PostgreSQL

Power BI

Эконометрика

Методы оптимизации

Дополнительная информация

Обо мне

Много времени уделяю чтению научной литературы, постоянно развиваю профессиональные компетенции, интересуюсь новыми технологиями и готов изучить новые языки программирования. Участвовал в хакатонах «ITHack» и «Samsung Innovation Campus Hackathon», где разрабатывал решения в условиях ограниченного времени и взаимодействия в команде. В перспективе планирую углубить знания в области классического и глубокого обучения, а также исследовать их практическое применение в реальных проектах.

Хакатоны:

NUCLEAR IT HACK (Кейс от РЖД)

April 2024

▢ Работал над созданием и сбором датасета, включающего изображения курящих и некурящих людей. Проводил аугментацию данных. Разработал и дообучил модель YOLOv8 для классификации курильщиков, а также использовал YOLO-NAS Pose для детекции ключевых точек поз человека.

▢ Stack: Python, Firebase, Flask, Tutorial, Material Design, MLFlow

▢ Ссылка на проект: https://github.com/IlyaChubov/Detection_smokers

SamsungHack

May 2024

▢ Работал над созданием и обработкой датасета погодных

данных. Выполнял предобработку данных, включая группировку, заполнение пропусков и аномалий. Разработал и применил модель SARIMAX для анализа и предсказания временных рядов. Проводил сезонное декомпозирование.

□ Stack: Python, Statsmodels, sklearn

□ Ссылка на проект: <https://github.com/IlyaChubov/SamsungHack>

Собственный проект

□ На основе открытого датасета футбольного поля дообучил модель yolov11 для задачи детекции и трекинга мяча. Полученные результаты логировал, используя wandb. Работал над подготовкой датасета для обучения модели, использовал техники для борьбы с некорректными предсказаниями модели (выбросами), включая интерполяцию.

□ Stack: Python, YOLOv11, Wandb, Roboflow

□ Ссылка на проект: <https://github.com/IlyaChubov/Soccer-ball-detection-and-track>

Решал задачи по классификации, генерации текста, реферированию текста, question-answering задачам (BERT, RoBERTa, GPT, T5, GRU, LSTM).

Мои проекты включали:

- Классификация текста: Разрабатывал модели на основе трансформеров, таких как BERT и RoBERTa, для решения задач классификации текстов. Применял fine-tuning для повышения точности моделей.
- Генерация текста: Использовал GPT и T5 для решения задач автоматической генерации текстов, ответы на вопросы.
- Реферирование текста: Применял T5 и BART для задач сжато-реферированного текста, где целью было создать краткие версии больших текстов без потери основного смысла. Использовал стратегии attention для выделения ключевых фрагментов текста и упрощения для удобства восприятия.
- Question-Answering: Разрабатывал решения на основе BERT и RoBERTa для задач question-answering, где модель должна была находить точный ответ на заданный вопрос из текста. Использовал подходы с вниманием для эффективного извлечения релевантной информации.

Знаком с методами векторизации (tf-idf, FastText, GloVe, Word2Vec) и способами обработки текста перед подачей в модель.

В рамках предобработки данных использовал различные подходы, такие как:

- Токенизация, лемматизация, удаление стоп-слов для очистки данных.
- Применение векторных представлений слов с использованием моделей FastText, GloVe и Word2Vec для конструирования эмбедингов, что позволяло захватывать семантические зависимости и контексты.

Понимаю, как устроены модели глубокого обучения, основные метрики и как стабилизировать обучение и бороться с переобучением.

- При обучении использовал кросс-валидацию и регуляризацию (Dropout) для уменьшения переобучения. Также активно использовал методы настройки гиперпараметров для оптимизации производительности модели.
- Понимаю принципы работы трансформеров и использую их для улучшения качества предсказаний, включая работу с multi-head attention и позиционными эмбедингами.
- Для метрик применял accuracy, precision, recall, F1-score и AUC-ROC для оценки качества моделей классификации, а также BLEU и ROUGE для задач генерации и реферирования текста.

Дополнительно:

- Знаком с современными методами предобработки и обработки текста с помощью библиотеки spaCy и интеграцией с Hugging Face Transformers для работы с предобученными моделями.
 - Работал с гибкими архитектурами, включая GRU и LSTM для обработки последовательных данных, а также применял методы внимания в сочетании с RNN для решения задач временных рядов и обработки длинных последовательностей.
- В своей работе использую PyTorch, Transformers и Hugging Face.

Контакты:

GitHub: <https://github.com/IlyaChubov>

Tg: @younlean

Phone: +79896302782