

Цепков Ярослав Алексеевич

ML/Data Engineer

Email: yaroslav.tsepkov@gmail.com | Phone: +79913961901

Github: [yaroslavtsepkov](#) | Telegram: [zireaelv](#)

Образование

Самарский университет

Магистр (незаконченное, не защищал диплом)

Август 2020 - Февраль 2023

Факультет информатики. Прикладная математика и информатика.

Разработка интеллектуальных систем.

Самарский университет

Бакалавриат

Август 2016 - Август 2020

Факультет математики. Прикладная математика и информатика.

Системный анализ.

Опыт работы

PepsiCo (проектная занятость)

Data Engineer

Декабрь 2022 - Июнь 2023

- Разработка ETL системы электронной комерции (ECOM)
- Проектирование озера данных на базе Azure Blob Storage, и базы данных MSSQL
- Разработка скриптов сборщиков данных Python+Pyspark
- Настройка конвейеров данных с использова Azure Data Factory
- Выявление потребностей бизнеса

Xim inc.

Data Engineer

Январь 2022 - по настоящее время

- Поддержка текущей ETL системы. Обновление API сборщиков данных, оптимизация процессов. Написание новых сборщиков данных по API, email, GCS, S3.
- Работа с BigQuery, ClickHouse. Написание моделей и макросов для DBT.
- Проектирование новой ETL системы.
- Работа с экосистемой Google Cloud Platform.
- Работа с различными API сторонних сервисов (MyTarget, Yandex, Google Ads, Yahoo, Apple Search, ...)
- Работа с данными медиаций и appsflyer.
- Работа с BI инструментами Grafana, Datastudio
- Подддержка текущей ETL системы и помощь в проектировние новой
- Разработка модулей по сбору данных различных нетворков используя API сервисов. Поддержка текущих модулей.
- Формирование витрин данных аналитикам для построения отчетов

СБЕР

Старший аудитор (Data Scientist)

Март 2021 - Август 2021

- Выгрузки данных из различных источников с использование различных диалектов SQL
- Очистка и маскирование данных с использованием regex и расстояния Ливенштейна
- Проведение анализа с использованием графовой теории. Построение графов связей и процессов с использованием Gephi, networkx, SberPM
- Исследовательские и продуктовые проекты в области машинного обучения: задачи классификации, кластеризации и поиска аномалий в данных, распознавание документов, детекция объектов
- Написание внутренних библиотек для будущих разработок

Интегра-С

Программист

Январь 2020 - Август 2020

- Выявление бизнес требований к приложениям.
- Помощь в разработке дизайна. Преимущественно работа с векторной графикой.
- Написание тонкого клиента для мобильных приложений iOS/Android.
- Работа с REST API, написание документации к ПО.
- Работа с потоковым видеосигналом RTSP.
- Поиск оптимальной модели для детекции объектов на видео для серверной части.

Персональные проекты

Умный дом

Участие в проекте умный дом в качестве мобильного разработчика. Разработка мобильного клиента на iOS, разработка архитектуры базы данных.

Высокопроизводительные вычисления

Решение задач линейной алгебры, обработки сигналов с использование технологии CUDA. Написание статьи на newtechaudit: [Использование технологии CUDA в решении прикладных задач](#).

Обработка изображений

Изучение библиотек для работы с изображениями и собственная реализация алгоритмов обработки изображений. Написание двух статей по этой теме на newtechaudit: [Анализ и обработка изображений с использованием операций математической морфологии](#) и [Разработка аналитических веб-приложений с использованием библиотеки Streamlit](#)

Навыки

- Языки программирования **Python, Java (EE 8 stack),Scala (only spark framework).**
- СУБД с которыми работал **BigQuery, ClickHouse, MSSQL, Oracle Database, Postgres.**
- Форматы с которыми работал **Parquet,Avro,ORC,JSON,YAML,MD,Protobuf.**
- Работа с данными **Pandas, Dask, OpenCV, Spark, PyArrow.**
- Машинное обучение и глубокое обучение **SciKit-learn, PyTorch, XGboost, Keras.**
- Высокопроизводительные вычисления **CUDA, Numba, Multiprocessing, Threads, Concurrent, Async.**
- Визуализация **Plotly, Dash, Altair, NetworkX, Gephi.**
- Облачные платформы **Google Cloud Platform (BigQuery,Cloud Storage), Azure (MSSQL, Blob Storage, Data Factory, Databricks, Key Vault), Amazon (S3 bucket)**
- Сетевой стек **socket.io, ssl, request, selenium, oauth2, rtsp.**
- Остальное **Linux, Bash, Git, Jira, Markdown, Latex, WolframMatematica, GNU Octave, HTML5, CSS3, Azure DevOps, BitBucket, Gitlab, Github.**