



Карпенко Сергей

Мужчина

+7 (962) 4494951

iskarp35@gmail.com — предпочитаемый способ связи

Другой сайт: <https://github.com/SergeyKarpenko1>

Проживает: Москва

Гражданство: Россия, есть разрешение на работу: Россия

Не готов к переезду, готов к командировкам

Желаемая должность и зарплата

Data Scientist (NLP)

Специализации:

— Дата-сайентист

Занятость: полная занятость

График работы: полный день, удаленная работа

Желательное время в пути до работы: не имеет значения

Опыт работы — 3 года 1 месяц

Сентябрь 2024 —
Февраль 2025
6 месяцев

FronTech AI

Москва

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения
- Системная интеграция, автоматизации технологических и бизнес-процессов предприятия, ИТ-консалтинг

Data Scientist (NLP/RAG)

Стек: Python, pdfplumber, PyTorch, Hugging Face, Transformers, OpenAI, FastAPI, LangChain, Chroma, pytesseract.

- * Спроектировал и разработал с нуля систему Retrieval-Augmented Generation (RAG) для оптимизации поиска и генерации релевантной информации на основе корпоративных данных, включая регламенты, закупочную документацию и прайс-листы.
- * Использовал библиотеки PyPDFLoader и pytesseract для обработки документов в различных форматах (PDF, изображения), обеспечив 95% точность извлечения текста.
- * Реализовал семантическое разделение документов на чанки, что улучшило структуризацию данных и ускорило обработку.
- * Настроил и оптимизировал базу данных Chroma для индексирования данных, что сократило время поиска релевантной информации.
- * Применил алгоритмы MMR (Maximal Marginal Relevance), что повысило точность поиска релевантной информации в векторной базе на 20%.
- * Использовал GPT-4o-mini для генерации текстов и решения задач обработки естественного языка, что позволило сократить время и улучшить качество генерации ответов.
- * Построил с использованием LangChain сложные цепочки обработки данных, автоматизировав ключевые сценарии работы с текстом.

Результат: Система успешно внедрена и обеспечивает быстрый и точный доступ к данным, что повышает эффективность сотрудников и снижает затраты на поиск и обработку информации.

Ноябрь 2023 —
Август 2024
10 месяцев

TWINO

Москва

Финансовый сектор

- Финансово-кредитное посредничество (биржа, брокерская деятельность, выпуск и обслуживание карт, оценка рисков, обменные пункты, агентства по кредитованию, инкассация, ломбард, платежные системы)
- Банк

Data scientist (ML)

Стек: Python, Pandas, Numpy, Matplotlib, Sklearn, CatBoost, Git, PostgreSQL

- * Участвовал в полном цикле разработки ML-моделей: от формулирования целей и задач до оценки экономического эффекта, включая сбор и подготовку данных, обучение, валидацию и тестирование моделей.
- * Разрабатывал предложения и инструменты для улучшения процессов разработки ML-моделей, что позволило ускорить разработку и внедрение решений.
- * Мониторил текущие ML-модели и оптимизировал их для повышения точности и эффективности.
- * Инициировал и тестировал новые источники данных, что привело к улучшению ключевых метрик моделей.
- * Рефакторил существующие R-скрипты и эндпоинты на Python для унификации и повышения производительности системы.

Сентябрь 2021 —
Май 2023
1 год 9 месяцев

XFIT

Ставрополь

Data scientist / ML engineer

Стек: Python, Pandas, Numpy, Matplotlib, Sklearn, Xgboost, CatBoost, LightGBM, Git

XFIT это международная компания, которая уже не первый год является признанным лидером рынка фитнес-услуг в России и активно развивается за рубежом.

Мои основные обязанности:

- * Выявлял проблемы в бизнесе, которые можно решить при помощи Data Science;
- * обрабатывал и анализировал данные с использованием статистических методов (python, pandas, numpy).
- * Создавал информативные графические представления данных и извлекал ключевую информацию, которая помогала в принятии решений (matplotlib, seaborn).
- * Строил ML модели для оптимизации работы фитнес клуба, включая прогнозирование факторов, таких как посещаемость, выручка и эффективность маркетинговых мероприятий (LogisticRegression).
- * Выявлял потребности клиентов в новых видах услуг при помощи анализа посещения различных зон клуба.
- * Решал задачи детекции при помощи CV (YOLO, detectron2).
- * Внедрял и оптимизировал модели, а также мониторил и улучшал их эффективность в рабочем процессе.

Достижения:

- * Используя базу данных переписок отдела продаж с клиентами, разработал телеграм-бот с применением глубокого обучения (DL) и Sentence Transformers (PyTorch), что помогло улучшить эффективность коммуникации отдела продаж с клиентами клуба.
- * Увеличил прибыль фитнес департамента на 9% г/г, в последний год работы на 14% г/г.

Образование

Высшее

2008

Ставропольский государственный университет, Ставрополь

Информатика и вычислительная техника, Автоматизированные системы обработки информации и управления

Навыки

Знание языков

Русский — Родной

Навыки

Python Big Data Machine Learning Jupiter NLP Computer Vision
Seaborn Matplotlib pandas Numpy Git Docker NN
Deep Learning PostgreSQL RAG BERT GPT HuggingFace FastAPI
CatBoost sklearn LangChain

Дополнительная информация

Обо мне

Я Data Scientist с 3-летним стажем работы в области machine learning и deep learning. Специализируюсь на разработке решений с применением нейронных сетей и трансформеров для задач обработки естественного языка (NLP), включая BERT, GPT, и адаптацию моделей с использованием PEFT. Имею опыт реализации Retrieval-Augmented Generation (RAG) систем для интеграции внешних данных в генерацию текстов. Моя экспертиза включает построение, обучение и оптимизацию моделей для классификации, предсказания последовательностей и генерации текстов.

Имею опыт полного цикла разработки ML-моделей: от формулирования целей до оценки экономического эффекта, включая сбор данных, обучение, тестирование и валидацию. Разрабатывал инструменты для оптимизации процессов разработки ML, что ускорило внедрение решений. Оптимизировал текущие ML-модели, улучшая их точность и эффективность.

Стек: Python, Pandas, Numpy, Matplotlib, Sklearn, PyTorch, Hugging Face, Transformers (BERT, GPT), PEFT (LoRA, prompt tuning), RAG , XGBoost, CatBoost, LightGBM, PostgreSQL, Git

Катаюсь на сноуборде, люблю спорт.
Телеграм: Karpenko_Sergey1