

- Pytorch, TensorFlow, OpenCV, Docker, Git, EDA, Deep Learning, Classic ML, Hierarchical Time Series, CV, NLP, LLM, RAG, A/B, OCR, Elastic Search/Open Search, FastAPI, frontend reactpy, MultiAgent system, Generative Neural Networks, GAN, CNN

Проекты

Data Scientist Leroy Merlin/ Прогнозирование товарооборота по магазинам Леруа Мерлен/Лемана Про

- Прогнозирование метрик по магазинам, отделам, подотделам Леруа Мерлена
- Исследование различных моделей прогнозирования: LSTM, Градиентный бустинг, Arima
- Различные методы согласования иерархичных временных рядов
- Генерация фичей, исследование постановки целей и прогнозирования магазинов

Data Scientist Leroy Merlin CV/ Детекция, оптимизация груза

- Детекция груза с помощью Yolo, разметка, сбор и обработка данных
- Расчет объема груза по изображению и с помощью лазеров
- Оптимизация груза, калибровка камеры
- Визуализация и EDA анализ, расчет метрик
- Работа с базой данных, S3

Data Scientist Leroy Merlin CV/ Детекция запрещенной символики

- Детекция с помощью Yolo, разметка, сбор и обработка данных
- Реализация генерации обучающей и тестовой выборки
- Расчет метрик, составление документации
- Визуализация, клиент-серверное приложение на FastApi

Data Scientist Leroy Merlin NLP/ Поисковик для юридических документов (OCR + open search)

- OCR текста со сканов юридических документов
- Обработка изображений, авторотейт deskew
- Поисковик на основе SequenceMatcher и расстояния Леванштейна, Open Search
- Оптимизация поиска с помощью HNSW
- Клиент-серверное приложение FastApi
- frontend на reactpy

X5 Retail group DS/ Детекция дефектов паллеты (Сколы, отсутствующие детали)

- Сбор, разметка датасета
- Применения различных моделей (классификация, детекция, сегментация) и архитектур для решения задачи
- Использование исправление перспективы
- Использование симметрии для выявления аномалий

X5 Retail group DS/ OCR сканов паспортов

- Сбор и разметка датасета
- Применения различных библиотек для задачи извлечения текста
- Подсчет и написание кода для метрик WER, CER, IoU
- Веб приложение

TWN Hearts Стартап LLM/ RAG система анализа отношений

- Сбор и разметка датасета
- Обработка данных, различные методы чанкирования
- Классификация токсичности
- EDA анализ, Построение RAG, LangSmith, LangChain

Гагарин Хакатон NLP 13 место NLP/ Поиск новостей о компаниях для брокеров и банков

- Генерация данных, обработка данных
- Выделение сущностей Natasha
- tf-idf, FastText
- Оптимизация поиска

Образование