

Валерия Хе

+7 (777) 230-10-69 | Telegram: @khevlr

ПРОЕКТЫ

Дообучение BERT моделей (https://github.com/khevlr/tags_classifier):

- парсинг 15,000 новостных статей с сайтов tengrinews.kz и nur.kz. Использование **BeautifulSoup**, **Selenium**, **requests**, **JSON**, **Kaggle**.
- очистка данных, визуализация и аналитика, ребалансировка количества статей по каждому тегу.
- дообучение BERT моделей для предсказания тегов для статей, с использованием: **Hugging Face Transformers**, **Trainer**, **NLP methods**. Сохранение дообученных моделей и токенайзеров.
- вывод интерфейса для инференса в **Streamlit**.

Хакатон. Система рекомендаций карьерного пути на основе знаний, интересов и резюме:

- Сбор данных по вакансиям. Составление матрицы компетенций. Работа с API LLM для экстракции списка навыков и требований в вакансиях.
- Создание опросника по компетенциям для определения уровня. Меппинг уровня компетенций человека с требованиями вакансий
- **Cosine similarity**, **Jaccard similarity** для дополнительного расчета релевантности резюме человека для вакансии
- Работа с LLM для аналитики слабых и сильных сторон. Составление роадмапа для дальнейшего обучения.
- вывод интерфейса для инференса в **Streamlit**.

ОПЫТ РАБОТЫ

Бизнес аналитик, Samsonite Kazakhstan

09/2024-04/2025

- Анализ данных о продажах, клиентах и товарных категориях с использованием **SQL** (запросы, объединения, агрегации) и **Excel** (сводные таблицы, формулы).
- Разработка автоматизированных отчетов о продажах и сотрудниках. Подключение к Dropbox по API и 1C.
- Визуализация данных с помощью **Seaborn**, **Pyplot**.
- Обработка и очистка данных с помощью **Python (pandas, numpy)**, включая обработку пропущенных значений и аномалий.
- Построение простых моделей прогнозирования с использованием **scikit-learn** (линейная регрессия, RandomForest) для оценки будущих продаж.

EDUCATION AND CERTIFICATES

UIB

Бизнес логистика

Outpeer Data Science Course

- **Темы:** Python, Scikit-Learn, NLP, A/B tests, PyTorch, API, Статистика

Технический стэк

- **AI/ML:** Machine Learning, Deep Learning, LLMs (GPT-3/4, BERT), RAG Pipelines
- **Data Analysis:** Clustering (KMeans), SQL, A/B Testing, Time-Series Forecasting
- **Programming Languages:** Python (Advanced), SQL (Intermediate)
- **Libraries/Frameworks:** PyTorch, Hugging Face Transformers, LangChain, Scikit-learn