

Елисей Стаковский

NLP Engineer, Computational Linguist, Data Scientist

Опыт работы

Machine Learning Engineer (NLP), АО Эврика, Октябрь 2022 – По настоящее время:

- решение задачи классификации с помощью тематического моделирования (BigARTM)
- разметка данных, тренировка моделей NER
- написание pipeline для наполнения единой модели знания (EM3) с помощью моделей NER
- сбор данных, тренировка моделей для задачи Text Classification
- написание pipeline для решения задачи Text Classification с помощью NER моделей и дальнейшей кластеризации, используя кастомный Siamese Network Encoder
- сбор данных, тренировка моделей MLM для дальнейшего файнтюнинга под целевые задачи
- сбор данных, тренировка модели для задачи Relation Extraction для наполнения EM3
- менторство и тимлидинг младших сотрудников команды

Computational Linguist, Reverso, Сентябрь 2020 – Октябрь 2022:

- расширение списка глаголов немецкого для сервиса Reverso Conjugator с помощью парсинга Wiktionary: добавлено 80 глаголов
- написан инструмент на Python для предобработки текстовых данных, очистки и обработки с помощью API NeuroSpell: подготовлен чистый корпус на немецком (30 тыс. предложений) для дальнейшего файнтюнинга модели нейросети
- написан инструмент на Python с целью сбора актуальных текстовых данных с вебсайтов для испанского, французского и немецкого языков: собрано по 30 тысяч предложений для каждого языка
- сбор и маркировка данных для улучшения работы сервиса Ginger: размечено 20 тыс. предложений
- создание и оценка тестов для NMT (Нейронный машинный перевод)
- локализация и QA продуктов Reverso и Ginger

Образование

В.А Национальный исследовательский университет
«Высшая школа экономики»
(Филология, Лингвистика и Переводоведение),
Июнь 2020; Средний балл: 9.72/10



Контакты

eistakovskii@gmail.com

<https://www.linkedin.com/in/elisei-stakovskii-203b0419a>

<https://github.com/eistakovskii>

<https://www.kaggle.com/eliseistakovskii>

<https://huggingface.co/ElStakovskii>

Языки

Английский (C1, Advanced)

Французский (B2)

Немецкий (B2)

Сертификаты

GRE (Декабрь, 2019): 314/340

TOEFL (Октябрь, 2021): 112/120

Курсы

- Machine Learning by Stanford, Coursera
- Deep Learning Specialization by DeepLearning.AI, Coursera
- Natural Language Processing Specialization by DeepLearning.AI, Coursera
- Machine Learning Engineering for Production (MLOps) Specialization by DeepLearning.AI, Coursera (in progress)

Навыки

Теория: Статистика, Теория вероятности, Линейная алгебра, Математический анализ

NLP (Обработка естественного языка) задачи: sentiment analysis (logistic regression, naïve Bayes), POS tagging (Viterbi Algorithm), auto-complete (N-gram language model), auto-correct (minimum edit distance, dynamic programming), translation (locality-sensitive hashing, approximate k-nearest neighbor), Word2Vec (CBOW, Skip-gram), NER (LSTMs with linear layers, transformers), question duplicates identification ('Siamese' LSTM models), Transformer Summarizer, Question Answering, Chatbot (Reformer), Relation Extraction, Coreference Resolution, Text Generation

Библиотеки Python: NLTK, SpaCy, NumPy, Natasha, PyMorphy, Selenium, Pandas, Transformers, Scikit-learn, Pytorch, TraX, Matplotlib, OpenCV, FastAPI, Flask, Requests, BigArtm, fasttext, gensim;

Инфраструктура: Git, GitLab, Docker, SQL (MySQL), YouTrack, Jira, Linux (Ubuntu, Astra)

Machine Learning: Linear Regression, Logistic Regression, Neural Networks, Support Vector Machines, PCA, Random Forests, K-means

Deep Learning: RNNs (LSTM, GRU), GANs, Transformers (BERT, GPT, T5, Reformer; Finetuning, Adapters, Intermediary Pre-Training)