

Fateme Najafi

Data Scientist

+989190866073 • fatemenajafi135@gmail.com • linkedin/fatemenajafi135 • github/fatemenajafi135

WORK EXPERIENCE

Data Scientist (NLP Developer) | Part AI Research Center

Apr 2023 - Present

- **Danabot** (A smart chatbot for customer support): Developed a cutting-edge chatbot via Langchain (RAG), optimized for handling FAQs, assisted by the retriever and memory module.
- **Sahab Services** (A platform and marketplace for AI-related APIs)
 - Optimized the POS Tagger service by integrating a transformer model, achieving a remarkable F1 score of 97%, particularly robust in handling informal texts.
 - Implemented a transformer adapters for resource optimization in POS tagging, NER, and news classification tasks alongside Virapart, a service for three exclusive tasks in Persian: punctuations, space/half-space, and kasre ezafe.
- **Comment Visualization**: Conducted data storytelling for a comment dataset through sentiment analysis, Wordcloud generation, and the development of a Kibana dashboard.

Data Scientist (NLP Developer) | Dadmatech

Oct 2021 - Feb 2023

- **ABSA** (An Aspect-Based Sentiment Analysis service)
 - Gathered data in three different domains: Online-sale website comments (tagged with a semi-supervision method), tweets related to a political person, and a telecommunication company.
 - Trained Aspect Extraction and Aspect-based Sentiment Analysis models on various datasets using BIOESQ sequence labeling for the aspect extraction part and BERT for sentence pair classification for sentiment analysis.
- **Chatyaar** (An Intelligent chatbot for customer service)
 - Utilized data from two businesses to enhance the chatbot pipeline's intent detection, boosting F-score by 10-15%. Incorporated specific similarity models for chitchat or FAQ intents and default answers for others.
- **Khabarjooy** (Persian news crawler and analyzer)
 - Designed an automated data pipeline using Scrapy to crawl, preprocess, and index data from the top ten Persian news agencies into an Elastic database every hour, while also conducting comprehensive analysis on the crawled news, encompassing keyword extraction, categorization, similarity identification, and NER with entity linking.
- **Harassment Detection in Divar's chat**: Fine-tuned a transformer-based multimodal model to detect harassment in user conversations on Divar, Iran's leading classified ads and e-commerce platform, achieving an impressive F1 score of 95%.
- **Nevis** (A Persian Spelling Correction): Evaluated Dadmatech's introduced model and other released models for Persian spelling correction, considering variations in Persian writing styles, including the use of spaces and half-spaces.

Research Assistant | Amirkabir University of Technology

May 2020 - Oct 2022

- Engaged in Natural Language Processing research, exploring various NLP tasks, including Transformers-based Language Models, Statistical and RNN-based Language Models, Text Classification, Sentiment Analysis, and Question Answering.

EDUCATION

Amirkabir University of Technology

Sep 2019 - Oct 2022

Master of Science: Computer Science (Data Mining)

- **(Thesis)** Irony detection in the Persian language: Applied transformer-based language models to a crawled and automatically labeled dataset, achieving an F1 score of 85% in classifying Persian tweets into ironic and non-ironic categories.

Kharazmi University of Tehran

Sep 2014 - Sep 2019

Bachelor of Science: Computer Science

SKILLS

Statistics, Machine Learning, Deep Neural Networks, Natural Language Processing, Python programming, Linux, Web Scraping, Web Service API, Git, SQL, Docker, CI/CD, Sphinx, Kibana, NumPy, pandas, matplotlib, sklearn, PyTorch, transformers, adapter-transformers, nltk, hazm, flask, Keras, TensorFlow, LangChain, chromaDB