Ali Abbas Baygi 🌐 in

Data scientist

Experienced data scientist with expertise in LLMs, algorithm design, and machine learning software development. Holds a Master's degree from TU Berlin with 5+ years of hands-on experience in modeling intricate systems. Skilled in data analysis, forecasting, and visual representation of complex datasets. Self-taught background with a knack for creating and evaluating adaptive services to produce valuable deliverables. Passionate about deep work and mastering challenging tasks quickly. Proficient in addressing ethical issues and biases in NLP models.

Resident: Berlin (permanent residence)   **E-mail:** alibaygi@gmail.com   **Website:** https://alibaygi.github.io

WORK EXPERIENCE

Samrtblick (Berlin-Germany)

Data scientist
June 2019-Present

- **MarketMoverBot** (internal): Developed a cutting-edge chatbot using Langchain (llm=gpt 3.5) and RAG technique, optimized for marketing and sales teams. Supported by retriever and memory modules.
- **OrderAnalystBot:** Developed a PostgreSQL analyst chatbot using Langchain's agent to interact with the order databases and provide answers to users' general questions.
- **Competitors analysis:** Conducted the process of evaluating the strengths and weaknesses of competitors. Leveraged NLP techniques, including variants of RAG to assist in information extraction, summarization and sentiment analysis.
- **Lead Generation:** Utilized various web scraping methods to extract contact information, relevant data, and potential leads from websites, contributing to the sales team's impressive 37% win rate.
- **Customer Analytics:** Led a comprehensive customer analytics and STP (segmentation, targeting, and positioning) framework initiative to categorize customers into distinct groups, leading to an increase in 11% customer retention.
- **Pattern counting:** Collaborated as a key member of a team focused on developing a robust algorithm for pattern recognition and similarity detection using cross-correlation.
- **Machine failure detection**: Developed a machine learning model utilizing staircase-like sequences for detecting machine failures. The model, created using Scikit-learn, Tsfel and TensorFlow, achieved an accuracy of 89% in detecting machine failures.
- **Machine data analytics:** Proficient in extracting industrial machine speed and loads by analyzing current signals, employing advanced signal processing algorithms for denoising, and extracting meaningful patterns.
- **AB testing:** A/B testing initiative with focus on P-values to evaluate various versions of the landing page, employing statistical methods to analyze user interactions and preferences. The A/B testing approach enabled us to achieve a remarkable increase of 18% in conversion rate.

Fraunhofer-IWF, (Berlin-Germany)

Internship,

Feb 2019-May 2019

- Simulating heat transfer and the influence of cooling on the cutting process
- Image processing with digital microscope VHX-5000 Series
- 3D model development with NX
- FEM and CFD simulation of heat transfer between tool and chip with ANSYS Workbench

TECHNICAL SKILLS

Programming for ML

Python programming, WebScraping, NumPy, Pandas, matplotlib, Sklearn, Nltk, fastapi, streamlit.

Generative AI and LLMs

Transformers(Hugging Face), LangChain, Openai

Database/Server

Postgresql, Mongodb, Elasticsearch, chromaDB, myScale.

Deeplearning

PyTorch, Keras, Tensorflow

Other Tools

Web Service API, Git, Gitlab, CI/CD, Application containerization e.g Docker, Microsoft Azure.

LICENSES & CERTIFICATES

TensorFlow Developer Professional Certificate

Generative AI with Large Language Models

Natural Language Processing in TensorFlow

Convolutional Neural Networks in TensorFlow

Neural Networks and Deep Learning

Sequence Models

Machine Learning Data Lifecycle in Production

Introduction to Machine Learning in Production

NLP With Transformers in Python

Energy Industries Engineering & Design, (Tehran-Iran)

Project Engineer and QHSE coordinator,

Apr 2009-Sep 2015

- Data Analysis, monitoring and measurement using statistics methods such as Pareto chart and fishbone diagram,
- Monitored, analyzed & evaluated failures or qualitative problems of the project and presenting necessary mechanisms to the project management team
- Analyzed data related to the project customer satisfaction

EXPERTISE

Statistics

Confidence intervals, significance estimation, distributions, hypothesis testings, A/B testing and experimental design, p-value interpretation, principal component analysis (pca), kernel density estimation (kde), regression analysis (linear, logistic, etc.), time series analysis, bayesian statistics and inference.

Supervised / unsupervised ML & Deep Learning

Gaussian process, k-means, random forest, deep learning, feature importance, anomaly detection, dimensionality reduction, debiasing data, NN, and RNN.

Natural language processing and Generative AI with LLMs

Text preprocessing, named entity recognition (NER), sentiment analysis, POS tagging, word2vec, bag of words, TFIDF, transformers and attention mechanisms (e.g., self-attention, multi-head attention), LLMs (GPT-3, Bloom, BERT, RoBERTa, T5), fine-tuning and transfer learning with pre-trained models, RAG (retrieval-augmented generation).

Digital Signal processing

Signal analysis, filtering and convolution, discrete fourier transforms, spectral analysis convolution, cross-correlation, noise reduction, feature extraction.