Ali Naeimabadi

Database Lab, Computing Science Dept., Edmonton, AB | (438) 924-0342 | anaeim.gdrive@gmail.com | LinkedIn | GitHub | Personal Website

Highlights of Skills

- Over 5 years of experience in developing and maintaining complex ML/Deep Learning projects on a large scale, such as ML-based client risk management models for a major Canadian bank
- Strong knowledge of LLMs for NLP and Information Retrieval (IR) application through different projects
- Experienced in ETL development for large datasets
- Experienced in time-series forecasting using uni-variate, multi-variate, deep learning, and ensemble models
- Experienced in large-scale data base management systems (DBMS) using distributed computing platforms using Hadoop Distributed File System (HDFS), Apache Spark, etc, and SQL/NoSQL databases
- Demonstrated strong teamwork skills by being part of more than 4 teams of different sizes and collaborating with cross-functional teams over various large-scale projects

Related Experience

August 2022 - August 2023

Research Assistant | University of Alberta

Responsibilities:
- ETL (Extract, Transform, and Load) development for our datasets
- Creating two datasets: Walmart*-Amazon* and Amazon*-Google
- Implementing robust state-of-the-art e-commerce product entity matching models using and LLMs tabular data
- Designing unsupervised attribute ranking modules for tabular data using LLMs (RoBERTa, SBERT)
- Designing few-shot learning models for e-commerce product entity matching using LLMs (RoBERTa, GPT3)
- E-commerce data collection
- Web scraping on a large scale for e-commerce products
- Database extraction
- Dataset sanitization

September 2021 - April 2022

Research Assistant | Scotiabank

Responsibilities:
- Database integration using a 2-stage string matching methodology: blocking stage, machine learning-based filtering
- Knowledge Graph construction
- Building an internal search engine using Elasticsearch
- Implement the deep learning and machine learning solution into the pipeline
- Database extraction
- Database cleaning

September 2016 - June 2020

AI Scientist | Faroob Zaman Inc.

Responsibilities:
- Developing and implementing time series forecasting model for stock price prediction and sale price prediction
- Designing deep learning for sentiment analysis on website reviews to evaluate customer satisfaction
- Developing AI-assisted microscope image processing software for the thin-layer lab using deep learning (CNN) and machine learning (decision tree) models, University of Tehran
- Designing a deep learning model to predict non-linear chemical reactions to oil refinery yield and performing a global search for new lubricant products in gas and oil industries

Technical skills

- **Programming Languages:** Fluent in Python | Fluent in SQL, C++
- **Databases (SQL/NoSQL)**: Oracle DB, IBM DB2, MySQL, PostgreSQL, SQLite, MongoDB, Redis, Casandra, Neo4j
- **Deep Learning**: PyTorch, Hugging Face, TensorFlow, Keras
- **Parallel Processing**: Apache, Spark, Hadoop, CUDA, PyTorch Lighting, PyTorch DistributedDataParallel

- **Cloud Computing**: Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, IBM Cloud, Oracle Cloud
- **Other**: Docker, Git, Software Development Processes (Agile, Waterfall, DevOps, TDD: Test-Driven Development, Spiral), Data Structures and Algorithms, Clean Code, Code Refactor and Review, Multithreading, scikit-learn, SciPy, NumPy, Pandas, NTLK, GENSIM, Seaborn, Matplotlib, Ubuntu, Windows, MacOS

Education

September 2020 - August 2023

MSc. in Computing Science | University of Alberta

- Thesis: Product entity matching by leveraging tabular data and Large Langauge Models.
- Supervisor: Prof: Davood Rafiei.
- GPA: 3.84

September 2011 - August 2016

BSc. in Computing Engineering | Amirkabir University

- Thesis: Microscope images processing to predict fracture strength of nanoclay/polyamide 12 nanocomposites using CNN.
- Supervisor: Dr. Farkhindeh Hemmati.
- GPA: 3.96

Projects

Implementing a risk management model using multi-modal data

A robust deep learning model flags high-risk clients given multi-modal sources of data including transaction history, connections with other entities, news feeds on Twitter, and financial journals (e.g., Harvard Business Review, Wall Street Journal, etc.). To discover the connection between clients, a knowledge graph was constructed based on enriched client databases.

Skills: multi-modal data analysis, data fusion, data extraction, ETL, NLP, information retrieval, IR, cloud computing, parallel computing, knowledge graph construction, deep learning, LLM, DBSM, database management systems, HDFS, Hadoop, Apache, Spark, PyTorch, Scikit-learn, SQL, PostgreSQL, Python.

Time series forecasting models using deep multi-variate LSTM and ensemble models

A robust, high-performance time series forecasting model is implemented to predict the stock price. I designed ensembles by combining LSTM and decision-tress-based models. The ensemble model of XGBoost+multi-variate-LSTM outperformed other baselines for the AAPL stock price prediction task.

Skills: time-series forecasting, deep time series forecasting, quant, quantitative model development, multi-variate LSTM, Prophet model, Neural Prophet model, decision tree, XGBoost, Light GBM, ensemble models, ensemble XGBoost+multi-variate-LSTM, ensemble XGBoost+LightGBM, ensemble LightGBM+multi-variate-LSTM, ensemble XGBoost+LightGBM+ multi-variate-LSTM, python, tensorflow, sci-kit learn, APPLE stock price prediction, TESLA stock price prediction.

E-commerce product matching model using Large Language Models (LLM) and tabular data: TATEM

A fast, highly informed, and effective record-matching model is designed to find the matching pairs among online e-commerce platforms: Google-Amazon and Walmart-Amazon. For the first time, the model benefits from product titles and other complementary textual and tabular information | University of Alberta.

Skills: entity matching, record matching, Large Language Models, table serialization, tabular data encoding, high-dimensional similarity search, web tabular data extraction, ETL development, Extract, Transform, and Load development, machine learning, GPT3, BERT, RoBERTa, SBERT, Universal Sentence Embedding, deep learning, PyTorch, SQL, Hadoop, Spark, business intelligence, BI, AWS, GCP, Cloud computing, Python.

Introduced two product entity matching datasets using complementary product-specific tabular data: Amazon*-Google, Walmart*-Amazon*

For the first time, Online product entity matching datasets include product-specific tabular data as complementary domain knowledge | University of Alberta.

Skills: data collection, data quality, data governance, HTML, Web Scraping, tabular data, entity matching, online product matching, business intelligence, BI, Amazon*-Google dataset, Walmart*-Amazon* dataset, ETL development, Extract, Transform, and Load, Beautiful Soup, Selenium, Requests, SQL, Python.

Fast and robust domain adaptation for automated essay scoring

The data-efficient cross-domain model was designed to adapt knowledge from source domains to unseen target domains in low-resource scenarios using meta-learning | University of Alberta.

Skills: meta-learning, few-shot learning, deep learning, LSTM, RNN, CNN, 2-way 5-shot text classification, prototypical networks, induction networks, educational assessment, Automated essay scoring, Glove-word embedding, neural networks, AWS, Cloud computing, Python.

Few-shot entity linking in knowledge graphs using limited resources

The model encoded the entity descriptions by BERT-based transformers and executed a high-dimensional similarity search to find promising candidates for an unseen knowledge graph in a zero-shot setting | University of Alberta.

Skills: knowledge graphs, databases, BLINK, similarity search in high-dimensional space, FAISS indexer, Wikipedia, Wikidata, deep learning, transformers, Large Language Models, BERT, PyTorch, AWS, GCP, Cloud computing, Python.

Automated essay scoring model based on cohesion, coherence and text diversity linguistic features

A deep neural networks model was designed to predict the score of essays using deep hand-crafted features | University of Alberta.
Skills: educational assessment, automated essay scoring, deep learning, machine learning, LSTM, RNN, CNN, PyTorch, cohesion, coherence, text diversity, AWS, GCP, Cloud computing, Python.

Database integration between two large-scale databases

Database integration was applied through a 2-stage string matching method: LSH and a Snorkel rule-based classifier | Scotiabank.
Skill: databases, database integration, pyspark, Hadoop, SQL, PostgreSQL, machine learning, locality sensitive hashing (LSH), Cloud computing, snorkel.

Enhanced deep learning models for automated essay scoring

The project studied the effect of type of word vector representation on the accuracy of automated essay scoring systems | University of Alberta.
Skills: NLP, language model, automated essay scoring, LSTM, RNN, CNN, deep learning, machine learning, TensorFlow, keras, BERT, Word2Vec, GloVe, pre-trained GloVe, pre-trained word embedding, nltk, gensim, Python.

Publications

- **A. Naeim abadi**, T. Nayeem, D. Rafiei, TATTOO: Product Entity Matching as a Topology Construction, WSDM 2024 [submitted].
- **A. Naeim abadi**, T. Nayeem, D. Rafiei, Product Entity Matching via Tabular Data, CIKM 2023, Oct 21-23, 2023, Birmingham, United Kingdom.
- T. Firoozi, O. Bulut, C. Demmanse Epp, **A. Naeim abadi**, D. Barbosa, The Effect of Fine-tuned Word Embedding Techniques on the Accuracy of Automated Essay Scoring Systems Using Neural Networks, NCME 2022, April 21-24, 2022, San Diego, USA.
- **A. Naeim abadi**, An Introduction to the Radio Communication Systems, 1st Pulse Magazine on Electrical Engineering, Mar. 9, 2016, Ferdowsi University of Mashhad, Mashhad, Iran.
- **A. Naeim abadi**, F. Hemmati, A new model for rubber-toughened polyamide 12 nanocomposite using convolutional neural network methodology, Biomedical image processing J, [Under revision].

Activities and Interests

Theater, environmental conservation, art, hiking, biking, motor cruising, skiing, and travel