

Ali Naeimabadi

Edmonton, AB | (514) 690-1367 | naeimaba@ualberta.ca | [LinkedIn](#) | [GitHub](#) | [Personal Website](#)

Highlights of Skills

- Over 5 years of experience in designing and developing complex data science/deep learning projects such as enhanced Fact-based QA model using LLMs at University of Alberta
- Extensively experienced in NLP, LLMs and IR models (3+ years) for large structured and unstructured datasets through different projects: Information Retrieval-based QA, E-commerce Product Entity Matching, entity linkage, database integration, content-based recommendation systems, and fake disaster tweet classification by leveraging PyTorch, transformers, langchain, and scikit-learn libraries
- Enhanced the performance of LLM-based e-commerce record linkage model for Walmart-Amazon* dataset (F1: 87.0→90.56 (current state-of-the-art)), published in CIKM 2023 and WSDM 2024.
- Implemented LLM-based disaster tweet classification model and improved accuracy by 3% at Scotiabank
- Experienced in cloud computing (3+ years) using AWS, SageMaker, EC2, S3, GCP
- Experienced in large-scale Database Management Systems (DBMS) using PySpark, SparkML, Databricks, etc.
- 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

DATA SCIENCE RESEARCHER | Aug 2022 — Aug 2023

[University of Alberta](#), Edmonton, Canada

- Improved the accuracy of Information Retrieval-based QA model for Amazon-PQA dataset (2.5%) using LLMs (BERT, RoBERTa, ...) and textual/tabular product data
- Implemented few-shot e-commerce product entity matching models (TATEM) using LLMs (RoBERTa, GPT3), textual/tabular data, novel product attribute tables and novel table serialization technique
- Achieved new state-of-the-art results for Amazon*-Google (79.28 → 82.2) and Walmart-Amazon* (87.0 → 90.56) datasets by leveraging my TATEM model
- Created two e-commerce product entity matching datasets: Walmart*-Amazon* and Amazon*-Google. Added tabular information for 73% of products by collecting and scraping online product data (ATTARII scraper [\[GitHub Code\]](#))
- Presented the highlights and achievements of the project and public seminars (CIKM 2023, WSDM 2024)
- Collaborated with team members to share knowledge and make design decisions

DATA SCIENTIST | Sep 2020 — Apr 2022

[Scotiabank](#), Toronto, Canada

- Developed a comprehensive financial risk management model using multi-modal data (KPIs, KRIs, transaction history, entity relationships, financial news) to flag high-risk entities for investigators at Wealth department (GRM) / AML Department
- Improved the performance of the risk management model by 20% by leveraging XGBoost and embedded feature selection methods
- Improved LLM-based disaster tweet classification to recognize tweets about real events by 3% [\[GitHub Code\]](#)
- Implemented a database integration system to enrich internal client databases through a 2-stage entity matching methodology using PySpark, resulting in a 22% improvement in F1 score
- Implemented Scrum methodology, followed the TDD process and exploiting Jira and Confluence to ensure timely delivery of high-quality software
- Participated in daily and weekly sessions with the development team of 8 members to report progress and attended AML (Anti-Money Laundering) meetings with managers and financial investigators to gather requirements and deliver high-quality solution

SENIOR MACHINE LEARNING DEVELOPER | July 2018 — Sep 2020

[Farooob Zaman Inc.](#), Tehran, Iran

- Developed Time Series Forecasting (TSF) models for stock price and sale price prediction using ensembles and deep learning models [[GitHub Code](#)], improved performance by 5% by leveraging ensemble XGBoost+multi-variate-LSTM
- Transferred the knowledge, mentored two new junior members of the team, and provided technical leadership and supervision to them on maintaining the system
- Content-based recommendation systems for e-commerce products using textual and tabular information
- Engaged in meetings with stakeholders and potential clients to gather requirements, conceptualize software solutions, and establish project timelines

MACHINE LEARNING DEVELOPER | Sep 2016 — July 2018

[Faroob Zaman Inc.](#), Tehran, Iran

- Designed deep learning/machine learning for sentiment analysis on website reviews to evaluate customer satisfaction and achieved promising results through the LSTM model and word embeddings
- Developed an AI framework for microscope image processing to predict fracture strength of nanoclay/polyamide 12 nanocomposites using VGG16

Technical Skills

- **Programming Languages:** Fluent in Python | Fluent in SQL, C++ | R, Scala, Java, MATLAB, C#
- **Data Science:** Machine Learning, data mining, quantitative model, time series forecasting, decision tree, XGBoost
- **Deep Learning:** PyTorch PyTorch Lightning, TensorFlow, Keras
- **Parallel Processing:** Apache, Spark, Hadoop (HDFS), CUDA, PyTorch Lightning
- **Databases (SQL/NoSQL):** Oracle DB, MySQL, PostgreSQL, Casandra, Hive, Neo4j
- **NLP/IR:** LLMs, RoBERTa, GPT-4, Huggingface, transformer, Langchain, nltk, SciPy, Spacy, scikit-learn
- **Cloud Computing:** AWS, S3, EC2, SageMaker, GCP
- **Other:** GitHub, JIRA, Confluence, Software Development Processes (Agile, Waterfall, DevOps, TDD: Test-Driven Development, Spiral), object-oriented programming (OOP), Multithreading, Big Data

Education

MSc. IN COMPUTING SCIENCE | Sep 2020 — Aug 2023

University of Alberta, Edmonton, AB

- Thesis: E-commerce product entity matching by leveraging tabular data and Large Language Models
- GPA: 3.84

BSc. IN COMPUTING ENGINEERING | Sep 2011 — Aug 2016

Amirkabir University, Tehran, Iran

- Thesis: Microscope image processing to predict fracture strength of nanoclay/polyamide 12 nanocomposites using CNN
- GPA: 3.96

Selected Publications

1. **A. Naeim abadi**, T. Nayeem, D. Rafiei, TATTOO: Product Entity Matching as a Topology Construction, WSDM 2024 [under-review]
2. **A. Naeim abadi**, T. Nayeem, D. Rafiei, [Product Entity Matching via Tabular Data](#), CIKM 2023, Oct 21-23, 2023, Birmingham, United Kingdom
3. 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

Selected Achievements

- Received conference travel grant, Collision 2023 Conference
- Alberta Graduate Excellence Scholarship (AGES) (12,000 CAD)
- Awarded Alberta Innovates Graduate Student Scholarship (26,000 CAD)
- Awarded full scholarship (50,000 CAD) for MSc. program at University of Alberta

Selected Certificates

- AWS Machine Learning Specialty 2023
- Apache Spark with Scala - Hands On with Big Data!
- Spark and Python for Big Data with PySpark
- Docker Mastery: with Kubernetes +Swarm from a Docker Captain