Bahar Emami Afshar

171 O'Conner, Ottawa 613-447-1656 baharafshar2079@gmail.com linkedin.com/in/bahar-afshar github.com/beafshar Google Scholar

Summary — Research-focused Machine Learning Engineer with strong software development experience and a solid academic foundation. Passionate about optimizing AI and big data systems, with hands-on experience in explainable AI, large language models, online/active/semi-supervised learning, anomaly detection, and distributed systems. Published three peer-reviewed papers, with one under submission and one patent under review. Skilled in C++, Python, object-oriented design, and system-level development using Docker, Kubernetes, and big data tools such as MongoDB and Kafka. Open to relocation in Ontario.

Education

University of Ottawa **09/2023 – 08/2025**
M.Sc. in Computer Science - AI Applied Concentration – GPA: A+ *Ottawa, Canada*
University of Tehran **09/2018 – 07/2023**
B.Sc. in Computer Engineering – GPA: A+ *Tehran, Iran*

Skills

Languages Python, C/C++, Java, Javascript, R, SQL, Verilog.
ML Frameworks Pandas, NumPy, TensorFlow, PyTorch, Scikit-Learn, Hugging Face
Development Tools Docker, Kubernetes, Git, Kafka, Azure
Databases : MySql, HibernateORM, Elastic Search, MongoDB, Neo4j
Web React, Spring, HTML/CSS, Js
ML Expertise Anomaly Detection, NLP, Deep Learning, Active Learning, Online Learning, Semi-Supervised Learning, Large Language Models, Data Analysis, Big Data, XAI, Association Rule Mining, and Meta Learning.

Work Experience

H3M Analytics Inc. **09/2023 – 08/2025**
Graduate Researcher (MITACS Internship) *Ottawa, Canada*

- Submitted one **patent** in collaboration with H3M Analytics.
- Developed **X-ITERADE**, an explainable unsupervised anomaly detection framework for fraud detection that identifies high-quality suspicious cases without requiring labeled data, and achieved a **15-times improvement** in the imbalance ratio while maintaining flexibility under labeling budget constraints.
- Designed **ALISA**, a novel pipeline that transforms unsupervised outputs into a label-efficient learning process using weighted active learning, semi-supervised learning, and data augmentation, and improved the **f1_score by 22%** even in challenging one-class scenarios.
- Worked with extreme imbalance datasets with an imbalance ratio of **0.077%**
- Designed and Implemented various **active learning** and **semi-supervised learning** algorithms in Python using Pandas, and Sickit-Learn
- Implemented an explainability module for group behavior analysis in the financial domain using **LLaMA**, **Mistral**, and **GPT** with **in-context learning** and **prompt tuning**, and achieved an **accuracy of 90%**.

Peppy Digger **05/2021 – 09/2021**
ML Engineer Intern *Tehran, Iran*

- Implemented a 3-class **sentiment analysis** on the data gathered from Persian Twitter and improved the accuracy by **10%**.
- Handled imbalance cases with different resampling strategies
- Tuned model hyperparameters

Publications

- Afshar, B. E. et al (2025). Is Expert-Labeled Data Worth the Cost? Exploring Active and Semi-supervised Learning Across Imbalance Scenarios in Financial Crime Detection. Foundations and Practice of Security. FPS 2024. link
- Afshar, B. E. et al. (2025). ITERADE - ITERative Anomaly Detection Ensemble for Credit Card Fraud Detection. Discovery Science. link
- Afshar, B. E. et al (2024). To Label or to Pseudo Label? Active Learning vs Semi-Supervised Learning for Windows Malware Prediction. Canadian AI. link.
- Afshar, B. E. et al (2025). ALISA: Adaptive Learning through Iterative Semi-supervised Active learning. Under Submission.
- **Patent Submitted** – A two-stage framework combining unsupervised anomaly detection (X-ITERADE) and label-efficient learning (ALISA) to iteratively expand anomalous cases, with an integrated explainability module using SHAP values, decision trees, association rule mining, and large language models.

Full list available on my Google Scholar.

Notable Projects

PriceFinder (Operating Systems Course Project, C++) **12/2020**

- Developed a multiprocessing system in C++ to search directories and determine the lowest prices across multiple shops using named and unnamed pipes for interprocess communication; achieved a 2.44× speedup over the serial version through performance evaluation.

PhonePriceClassifier (Operating Systems Course Project, C++) **11/2020**

- Implemented a model to classify phone price ranges using both multithreaded and serial designs in C++; achieved 91% classification accuracy, demonstrating effective parallelization and performance optimization.

Semi-Supervised Learning for Bank Marketing Dataset (Machine Learning Course Project, Python) **03/2024**

- Implemented and benchmarked multiple semi-supervised learning models with resampling strategies on datasets containing varying proportions of unlabeled data, **improving accuracy from 76% to 88%** by pseudo-labeling the most informative samples.

Online Learning for Intrusion Detection (AI for Cybersecurity Course Project, Python) **11/2023**

- Implemented a dynamic model to deal with streaming data and detect attacks in real-time with **87%** F1-score

IMDB Clone (IEMDB), (Internet Engineering Course Project, Java) **05/2022**

- Developed full-stack IMDB-like website from scratch using React (frontend), SpringBoot, Hibernate, MySQL (backend), and deployed via Docker & Kubernetes.

Lung Images COVID Detection, (Artificial Intelligence Course, Python) **11/2020**

- Built a deep learning classifier from scratch using NumPy; compared results with PyTorch and TensorFlow on augmented chest X-ray data, and achieved an accuracy of 94%.

Languages

English : Fluent, IELTS Overall: 7.5, (L:8, R: 8, S: 7, W: 6.5)
Persian : Native
French : Beginner - A1 Proficiency