# Amirali D Shiraz

Website —Email — LinkedIn — GitHub

## Summary

Data engineer and machine learning practitioner with experience in Python, SQL, and large-scale data processing. Skilled in building ETL pipelines, feature engineering, and developing ML models for time-series and system-level data. Background in electrical engineering with hands-on experience in power systems and telemetry data. Fluent in English and French.

## Work Experience

**Bright Bee Technology Inc.** *03/2024 – Present*
*Data Scientist* Toronto, Ontario, Canada

- Built end-to-end data pipelines in Python (Pandas, NumPy) and SQL to ingest, clean, and transform structured and unstructured electrical engineering data, converting raw technical content into analytics-ready datasets.
- Designed scalable ETL/ELT workflows using PySpark, Databricks, and AWS (S3, Glue), improving processing efficiency and reducing manual data preparation by 30%.
- Developed optimized data models and complex queries in SQL, working with large datasets to create reliable, high-performance tables for analytics and reporting layers (e.g., Tableau / Power BI).
- Worked closely with data scientists to perform feature engineering, data validation, and preprocessing, using scikit-learn and Python-based workflows to improve model performance and consistency.
- Designed and supported ML models on healthcare data (e.g., blood pressure, glucose levels) using scikit-learn / TensorFlow, enabling anomaly detection and fatigue prediction, contributing to early identification of potential risk conditions.

**Polytechnique Montréal** *Jan 2021 - Sep 2024*
*Research Assistant* Montreal, Quebec, Canada

- Processed and modeled data from over 140,000 electric water heater sensors, transforming raw electrical telemetry into clean, ML-ready time-series datasets using Python and SQL.
- Developed data-driven droop control strategies for grid frequency stabilization, combining MATLAB/Simulink simulations with Python-based statistical and ML methods.
- Applied stochastic modeling (Kolmogorov PDEs) to simulate large-scale grid behavior under uncertainty, improving system-level analysis and robustness.
- Integrated power monitoring devices with telemetry systems aligned with SCADA / EPMS environments, enabling reliable data collection and validation.

**Rahbord Hooshmand** *Jan 2018 – Dec 2020*
*Intern*

- Performed statistical analysis on electrical and operational data using regression and Bayesian methods to support system performance improvements.
- Developed C++ components for telemetry systems in smart grid applications, improving modularity and system reliability.
- Optimized SQL queries and ETL pipelines for large-scale grid data, reducing processing latency by approximately 20%.
- Built dashboards using Power BI and Tableau to visualize system performance and support data-driven decision-making.

## Education

**Polytechnique Montréal** *Sep 2021 – Sep 2024*
*Master of Applied Science in Electrical Engineering, Montreal, QC, Canada*
**University of Tehran** *Sep 2016 – Jun 2021*
*Bachelor of Science in Electrical Engineering, Tehran, Iran*

## Projects

**DMFlix – Dagster + Snowflake + ML Embeddings**
Automated an ML pipeline with **Dagster** and **Snowflake** to decode embeddings, run t-SNE, and generate artifacts for

large-scale movie datasets. Built configurable assets with robust decoding, data cleaning, and feature augmentation, while hardening pipelines with parameterized SQL, resource-injected Snowflake connections.

**Temporal Difference Learning for Real-Time Control**
Implemented **SARSA** and **Q-learning** algorithms for stochastic control in unknown environments; trained RL agents with dynamic reward structures and benchmarked performance using Python and NumPy.

**Multi-GPU Stable Diffusion Pipeline for Visual Ad Generation**
Built and synchronized a custom GPU cluster for **Stable Diffusion 1.5** and **Stable Diffusion XL**, enabling parallelized image generation for advertising. Handled full system setup—GPU rack assembly, kernel tuning, NCCL sync—and optimized **CUDA pipelines** for batch performance. Engineered prompts and features to control output quality and style across multiple model checkpoints.

**Real-Time Disaster Sentiment Analysis on AWS**
Deployed a real-time **NLP pipeline** using **AWS Kinesis**, **Lambda**, and **Comprehend** to monitor social media sentiment during disasters; built a **QuickSight** dashboard to support live crisis response.

**Bayesian and Distributed Multi-Agent Estimation**
Developed a decentralized estimation system using **Bayesian modeling** and **distributed Kalman filters**; implemented real-time prediction across agents using Python, optimizing accuracy and communication in uncertain environments.

**Git System Development in C**
Designed a custom version control system in **C** with core Git functionalities (commit, log, checkout); built efficient data structures for snapshotting and delta comparison across file states.

**Super Mario Game Recreation in C++**
Rebuilt core mechanics of **Super Mario** in **C++** using object-oriented principles; implemented character physics, level progression, and collision detection to simulate 2D gameplay in a console environment.

## Technical Skills

**Programming & Data:** Python (Pandas, NumPy), SQL, PySpark, Time-Series Data, Data Structures & Algorithms
**Data Engineering & Cloud:** ETL/ELT Pipelines, Databricks, AWS (S3, Glue, Kinesis, SageMaker), Data Modeling, Large-Scale Data Processing
**Machine Learning & AI:** Scikit-learn, TensorFlow, PyTorch, Time-Series Forecasting, Anomaly Detection, Reinforcement Learning
**Electrical Systems:** Power Systems, Grid Frequency Control, Droop Control, Demand Response, Sensors & Telemetry, SCADA/EPMS
**Modeling & Visualization:** MATLAB, Simulink, Statistical Modeling, Power BI, Matplotlib, Feature Engineering
**Version Control & Workflow:** Git, GitHub, CI/CD

## Certifications

**Data Analytics Specialization** *Google, Nov 2024*
Covered exploratory data analysis, dashboards, SQL queries, and predictive modeling using Google tools and open-source libraries.

**Neural Networks and Deep Learning** *DeepLearning.AI, Nov 2024*
Completed practical training on forward/backward propagation, activation functions, gradient descent, and tuning multi-layer neural networks.

**Machine Learning Pipelines With Azure ML Studio** *Coursera Project Network, Sep 2023*
Designed ML pipelines using Azure's visual tools, including preprocessing, model selection, deployment, and performance tracking.

## Awards & Scholarships

**Ranked Top 0.17% in National Entrance Exam (Iran)**
Placed 250$^{\text{th}}$ out of over 150,000 participants nationwide, gaining admission to Iran's top-ranked engineering programs.

**Full Graduate Funding – Polytechnique Montréal**
Awarded full research funding to pursue M.A.Sc. at **Polytechnique Montréal**, ranked among the top 4 Electrical Engineering schools in Canada.

**Project Day Award – University of Tehran**
Recognized for technical innovation in undergraduate final project on simulation and modeling of residential energy consumption.