

AmirMasoud Azadfar

Data Scientist and AI Engineer

Thunder Bay, ON, Canada | amirmazadfar@gmail.com | +1 (343) 988-3995 | [LinkedIn](#) | [GitHub](#)

Tech Stack and Skills

Programming Languages	Python	C/C++	C#	SQL	Cypher	Bash (Unix Shell)	HTML/CSS	JavaScript		
Frameworks & Libraries	FastAPI	Flask	Django	MFC	.NET	Bootstrap	Matplotlib	Seaborn	Selenium	BeautifulSoup
Machine Learning	PyTorch	TensorFlow	Scikit-Learn	Hugging Face	LangChain	Pandas	NumPy	SpaCy	NLTK	
DevOps & Databases	Git	Docker	Uvicorn ASGI	Hetzner	AWS EC2	Neo4j	MongoDB	MariaDB MySQL	Redis	
Domains	AI Automation	Recommendation Systems	Data Engineering	Financial Modeling	Retrieval-Augmented Generation (RAG)					

Work Experience

Data Scientist and AI Engineer at CanApply | [Smart Platform](#) Mar 2023 - Jul 2024

- Web Scraping and NLP-Driven Data Pipeline:** Developed a data pipeline for 223 Canadian universities and colleges, collecting and processing over 13,000 degree programs. Used Selenium, BeautifulSoup, and proxy rotation for web scraping, and Pandas and RegEx for data cleaning. Fine-tuned a SpaCy NLP model for NER and relation extraction, built a Neo4j knowledge graph for knowledge retrieval, and used MariaDB for front-end data storage and retrieval.
- Degree Program Recommendation System:** Designed a content-filtering recommendation system to help students find suitable degree programs based on their preferences and academic background. Utilized text embeddings for semantic similarity, achieving nearly 100% accuracy in top-10 recommendations across 11 program categories. Maintained <800ms response time for recommendations and <500ms for filtration and pagination. Deployed using Docker containers on Hetzner VPS, and handled 100+ daily requests using FastAPI and Uvicorn ASGI.
- GraphRAG-Enabled Study Abroad AI Assistant:** Engineered 'Dana', an AI chatbot using retrieval-augmented generation (RAG) for personalized study abroad guidance based on real-time educational data. Employed multi-agent architecture and integrated OpenAI's GPT models API. Achieved 1.8-7.2 seconds RAG response latency on Hetzner production servers, handling 1000+ monthly conversations with FastAPI, Uvicorn ASGI, and Pusher WebSocket for real-time chat token streaming.
- Admission Chance Approximation Service:** Developed a scoring metric to approximate admission likelihood for Canadian universities, using 14 weighted features based on admission data, university requirements, and student profiles. Implemented a feature to generate an Instagram story image with the admission score, scoring between 300 to 1000, for a marketing campaign.
- Data-Driven Analysis and Business Intelligence:** Conducted analyses on the Canadian educational technology market, identifying key segments, growth areas, and potential partnerships. Implemented an ARIMA forecasting model to estimate future trends in international student enrollment, aiding in strategic decision-making for marketing and recruitment.

Data Engineer and Software Developer at Sepanta IT Co. Feb 2018 - Nov 2022

- Financial Data Pipeline Development:** Developed and updated data pipelines for TSE stocks and cryptocurrencies, collecting and processing technical, fundamental, and news data. Utilized Python, Selenium, Beautiful Soup, Pandas, and MongoDB for data retrieval and preprocessing. Achieved real-time data retrieval with sub-5000ms latency for TSE and sub-500ms latency for 1000+ cryptocurrency pairs using asynchronous API calls and WebSocket streams.
- Algorithmic Trading Strategy Development:** Created an option bonds pricing model using the Black-Scholes and Binomial Option Pricing Models. Designed strategies including Mean Reversion, Momentum Trading Strategies, Buyer/Seller Pressure, and Market Microstructure Analysis (Order Book Analysis). Developed a consensus clustering model for sector and industry analysis of stocks based on historical technical and fundamental data using K-Means and DBSCAN algorithms.
- NLP Model for TSE Stocks News Sentiment Analysis:** Built a sentiment analysis model for TSE stocks Farsi news and analyst reports. Manually labeled a dataset of over 15,000 Farsi news articles and reports. Used Hazm, a Persian NLP toolkit for sentence and word tokenization and preprocessing and implemented TF-IDF for vectorization. Trained a Multinomial Naive Bayes binary classifier for sentiment analysis and reached a macro average F1 score of 0.85.
- Market Analysis and Automatic Reporting Platform:** Designed and deployed a backend API for a Telegram-based platform providing market analysis and automatic reporting for TSE stocks using FastAPI and Uvicorn ASGI on Hetzner Cloud and MongoDB for date warehousing and user profiling, offering proprietary market analysis reports, news sentiment analysis, trading signals, and custom alerts.
- Cryptocurrency Triangular Arbitrage System:** Developed a real-time system for identifying triangular arbitrage opportunities on the Binance exchange in Python. Handled and scanned over 1000 trading pairs concurrently through asynchronous design and multithreading, with a latency of under 100ms, deployed on AWS EC2 Asia Pacific (Tokyo) zone for data feed latency reduction.

Projects

Automatic Job Hunting AI Assistant | [GitHub](#) 2024

- Developed an automated job hunting software that scrapes and interacts with major job listing sites, extracting job listings based on user preferences on roles and locations. Implemented a scoring system to match job descriptions with user resumes, automating the application process for high-scoring jobs, including resume customization and cover letter generation. Enabled batch processing to handle thousands of job descriptions, powered by OpenAI GPT-4 completions API for efficient decision-making and alignment analysis, streamlining the job hunting process.

RL-Driven Bidirectional Automatic Trading System | [GitHub](#) 2023 - Present

- Developing an RL-Driven Bidirectional Auto Trading System using a Deep Q-Network (DQN) to analyze historical and live trading data, technical indicators, and segmentation patterns for optimal trading decisions. Implemented a segmentation module to identify profitable trading zones via pivot points, guiding the DQN agent in making long, short, or hold decisions based on past market behaviors and technical indicators. Included real-time data fetching, order execution, and risk management modules to ensure continuous market updates, execute trades, and minimize losses while optimizing the portfolio.

GNN-Based Drug Interactions Recommendation System | [GitHub](#) 2023 - Present

- Developing a Graph Neural Network (GNN) model to predict drug-drug interactions based on chemical structures, side effects, and pharmacological properties. Utilizing a multi-layered graph architecture to represent drug compounds and their interactions, the model aims to provide accurate recommendations for potential drug interactions, contraindications, and adverse effects. Implementing a knowledge graph for drug compounds and their properties, the system will enable healthcare professionals to make informed decisions regarding drug prescriptions and patient safety.

Education

HBSc. in Computer Science, Lakehead University, Canada Sep 2023 - Present