Zohreh Nikkhah

M.Sc., Computer Science & Engineering (Applied AI)

LinkedIn: linkedin.com/in/zohreh-nikkhah

Email: zohreh.nikkhah1984@gmail.com

GitHub: github.com/Zohreh6384NKH

Tableau: Profile - zohreh.nikkhah | Tableau Public

Phone: +44 07951642050

Summary

Data Scientist with a solid background in machine learning. Skilled in developing AI and ML models, analyzing complex data sets, and building Agentic AI systems using AutoGen on top of world-renowned LLMs (GPT, Llama, DeepSeek, Mixtral)

Skills

Programming: Python, PyTorch, AutoGen, AutoGen Studio, Scikit-Learn, Numpy, Pandas, Scipy, OpenCV, Matplotlib, MySQL, Tableau, Power BI

Operating Systems: Linux (Ubuntu), Windows

Selected Courses: Machine Learning, Deep Learning, Data Mining, Database Systems, Linear Algebra, Statistics and Probability, Specialized Data Science, Advanced Python Programming Language, LLM Engineering

Misc.: Agentic AI, Contrastive Learning, CLIP, CLIP-Seg, Transformers, Large Language Models (LLM), Convolutional Neural Networks (CNNs), GANs, Self-Supervised Learning, SAM, YOLO, Random Forest, Logistic Regression, SVM, XGBoost, MLP, Object Detection, Semantic Segmentation, Classification

Experience

Minimus AI / Data Scientist

November 2024 - present, San Jose, CA, United States, Remote

Developed a **Multi-Agent** system to automate analyzing, categorizing and rewriting of any given **Github repository** using the **AutoGen** framework and **Llama3.2**

- Implemented an autonomous **multi-agent pipeline** that reduces manual code adaptation time by over 90%
- The system simplifies a complex codebase and makes it readable while **maintaining 100% functionality** proven by automated tests
- Built an **agentic RAG** application with database routing that uses multiple specialized **vector databases** (product info, customer support, financial data) with an agent-based router to direct queries to the most relevant database, using Langchain, LangGraph, Streamlit, GPT-4o

Pegah Software Company

Business Analyst (2019 – 2022)

- Developed **predictive** models for financial forecasting and equipment failure detection using Python libraries
- Implemented **machine learning algorithms** to optimize inventory management and reduce shortages
- Designed **automated data pipelines** (SQL, Pandas) for real-time analytics in finance and maintenance KPIs
- Built **interactive dashboards** (**Power BI**, **Tableau**) to support data-driven decision-making
- Applied **anomaly detection** techniques to identify irregularities in energy consumption and production data

Data Analyst & Project Control (2016 – 2019)

- Designed **interactive dashboards (Power BI, Tableau)** for tracking project KPIs across finance, operations, and supply chain.
- Built **ETL pipelines** (SQL, Pandas) to consolidate and transform project data for performance monitoring.
- Developed **predictive models** (Random Forest, Regression) to support **forecasting, risk analysis, and variance tracking** on key project metrics.
- Conducted **root cause analysis** on project delays and operational variances to recommend actionable improvements.
- Collaborated with cross-functional teams to ensure **data-driven project control and reporting** aligned with business goals.

—

Projects

Aston University / Applied AI Masters Student
September 2023 - September 2024, Birmingham, UK

Developed an end-to-end fashion item **classification** and **color extraction**

- Utilized **YOLOv8** for fashion item **classification**
- Utilized **SAMv2** for fashion item **segmentation**
- Applied **KMeans** for **color extraction**
- Tested the pipeline with a curated dataset from random images on the internet

Automated the labeling of new features on large-scale datasets using **CLIP**, **CLIP-Seg**, and **SAMv2**

- Utilized HuggingFace **CLIP** and **CLIP-Seg** models to detect clothing items and their colors in the COCO dataset
- Utilized **SAMv2** to cross-check the accuracy of segmentation masks
- Added the clothing items and colors to the captions of images and rephrased the captions
- Trained a model to detect and describe clothing items in images
- Tested the model against a curated test set

Built a **pipeline** for a dataset of **60+ million data points**

- **Data Cleaning** (missing value handling, outlier detection, balancing, and normalization)
- **Feature selection** (Correlation analysis and PCA)
- **Prediction** using Scikit-Learn Machine Learning models (Random Forest, Logistic Regression, XGBoost, SVM, and MLP)
- **Visualization and Presentation** of results using Confusion Matrix, Precision, Recall, Accuracy, and F1 Score

Developed an **ML** model to predict daily aggregate solar energy output from a photovoltaic (PV) system using months of weather and energy production data (**time-series**)

- Implemented baseline ML models (Linear Regression, Random Forest, XGBoost) and advanced deep learning architectures (LSTMs, Transformers), achieving a 24% accuracy boost with Transformer-based models

Education

Aston University

M.Sc., Computer Science & Engineering (Applied AI) - with Distinction

September 2023 - September 2024, Birmingham, UK
Thesis: End-to-End Fashion Item Classification and Color Extraction (Using YOLOv8, SAMv2, K Means Clustering)

Tehran Science and Research University
M.Sc., Industrial Engineering - with Distinction

Damghan University of Science
B.Sc., Applied Mathematics