# Mohammed Farhan Baluch

Kitchener, ON N2H 6M4
226-961-0559 | farhanbaluch1301@gmail.com | linkedin.com/in/farhan1301 | https://github.com/farhan1301

## Education

**University of Windsor** — Windsor, ON
*Master of Science in Computer Science, Artificial Intelligence Stream* — *Sep. 2022 – Apr. 2024*

- Cumulative GPA: **93.4** / 100
- Courses: Intro to Artificial Intelligence, Machine Learning & Pattern Recognition, Neural Networks & Deep Learning

**Vit Bhopal University** — Bhopal, IN
*Bachelor of Technology in Computer Science* — *Jul. 2018 – Jul. 2022*

- Cumulative GPA: **8.95** / 10.0
- Courses: Computer Vision, Advanced Data Analytics, Probability & Statistics, Data Management & Visualization

## Technical Skills

**Languages/Databases**: Python, PySpark, C, C++, Java, MongoDB, MySQL, NoSQL
**Machine Learning**: Scikit-learn, Pandas, NumPy, spaCy, XGBoost, LGBM, DQN
**Data Analysis**: Power BI, Microsoft Excel, Tableau, Looker, Seaborn, Matplotlib
**Deep Learning**: PyTorch, TensorFlow, Keras, BERT, GPT, LSTM, GNN, Transformer
**Data Science**: Prediction, Timeseries, Quantile Regression, Outlier Detection, Hypothesis Testing, Statistical modeling, Conversational Agents, Prompt Engineering
**Model Optimization & Explainability**: Hyperopt, Optuna, GANs, VAE, Lime, SHAP, Tensorboard
**MLOps**: Docker, Streamlit, Flask, A/B Testing, CI/CD pipelines
**Platforms**: Git, Unix, OpenAI API, Azure (DevOps, Databricks, ML), AWS (S3, EC2, SageMaker), Apache (Hadoop, Spark, Airflow)

## Awards

- Departmental, University of Windsor - $4,000 — Jan 2023
- **Provincial, Ontario Graduate Scholarship** - $15,000 — Sep 2022

## Certifications

- **Microsoft Certified: Power BI Data Analyst Associate (PL-300)** — Apr 2024
- NASSCOM Gold Certified Associate in Data Analytics — Jun 2021

## Experience

**Artificial Intelligence Developer (Co-op)** — May 2023 – Aug 2023
*Agri-Foods Canada* — *Harrow, ON*

- Spearheaded the development of a high-precision **plant disease identification model**, achieving **98.3%** accuracy for powdery mildew detection.
- Co-engineered an advanced Agri-Foods chatbot utilizing the Generative AI **GPT-4 architecture**, integrating **Azure ML** for enhanced customer interaction and support.
- Innovated an internal PDF Parser tool for the HR department, enabling efficient **extraction and dashboard visualization** of key document data with **Spacy and PowerBI**.
- Optimized model hyper-parameters using **Bayesian optimization techniques** - random search & gridsearch.
- Fostered collaborative efforts - **Standups, Git, and Azure Boards** with cross-functional teams within an agile project environment.

**Teaching Assistant (Fastlane - DaRMoD)** — Jan 2024 – Apr 2024
*Vector* — *Remote, ON*

- Developed and **delivered comprehensive tutorials** in data readiness, model development, and deployment
- Helped startups understand and implement **complex AI and machine learning concepts**.

- Provided constructive feedback and **evaluations on company projects**, leveraging expertise in machine learning to enhance project outcomes.

**Graduate Teaching Assistant** Sep 2022 – Apr 2024
*University of Windsor* *Windsor, ON*

- Facilitated weekly lab sessions for **Computer Architecture** (COMP-2560) and **Systems Programming** (COMP-2660) courses.
- Responsible for 40+ students lab work and assignments.
- Assessed and **graded academic work**, offering personalized feedback to promote understanding.

## Projects

**Fine Tuning Llama-2** | *Python, PyTorch, Streamlit* Dec 2023 – Feb 2024

- Employed **Supervised Fine-Tuning (SFT) to adapt Llama 2** for enhanced dialogue generation
- Leveraged a meticulously curated **custom mini-Platypus dataset** for instruction-based training.
- Applied near-deduplication techniques **using Sentence Transformer embeddings** and FAISS to maintain dataset uniqueness.
- Significantly reduced redundancy without compromising data integrity.
- Pioneered the application of **4-bit precision fine-tuning using QLoRA**, significantly minimizing VRAM usage while maintaining model performance.
- Developed a **text generation pipeline** that incorporates the refined model.
- Demonstrated the practical application of fine-tuning through enhanced text generation capabilities.

**Stock Portfolio Optimization (Thesis project)** | *Python, OpenAI Gym* Jan 2023 – Dec 2023

- Pioneered a cutting-edge interpretable stock prediction model.
- **Integrated Graph Neural Networks (GNN) with Reinforcement Learning (RL)**, enhancing market trend analysis and investment strategy formulation.
- Engineered predictive algorithms by processing vast datasets with advanced techniques like **Node2Vec for embedding generation**.
- Utilized **TD-3** as core component for reinforcement learning.
- **Enhanced model transparency and interpretability** by incorporating SHAP, LIME, attention networks, and counterfactual explanations.

**Predictive Analytics For Market Spend** | *Python, Snowpark, Streamlit* Aug 2023 – Nov 2023

- Led a data engineering and machine learning project to develop a Linear Regression **model predicting advertising budget ROI** across multiple channels (search, video, social media, email).
- **Utilized Snowpark for Python** to establish secure connections to Snowflake, enabling efficient data load, analysis, and preparation from Snowflake tables into Snowpark DataFrames.
- Orchestrated the **end-to-end machine learning pipeline in Snowflake**leveraging Snowpark's ML capabilities.
- Created **Python Scalar and Vectorized User-Defined Functions (UDF)** in Snowflake for model inference,
- Facilitated real-time predictions and analyses integrated within the **Streamlit application**.

**Diabetes Readmission Analysis** | *Python, R, SQL, PySpark, MLlib* Apr 2023 – July 2023

- Performed **analysis of more than 100,000 Clinical Database** Patient Records to understand factors responsible for early readmission of patients given their clinical information.
- Implemented Logistic Regression, Lasso Regression, and Random Forest models **using PySpark and MLlib**.
- Utilized feature selection methods like **Chi-Square and AIC** (Akaike Information Criterion).
- Engineered a data processing **pipeline with PySpark**, including **VectorAssembler** for feature vectorization.
- Achieved a **ROC score of 0.64** with Random Forest, indicating strong predictive capability compared to baseline.

## Publications

- Desai, N. P., **Baluch, M. F.**, & Aziz, R. M. (2023). Computer vision model with novel cuckoo search based deep learning approach for classification of fish image. *Multimedia Tools and Applications*, 1-20.
- **Baluch, M. F.**, Patel, S., Aziz, R. M., & Ganie, A. H. (2022). LGBM: a machine learning approach for Ethereum fraud detection. *International Journal of Information Technology*, 1-11.