Karim Nadim

Data Science | Machine Learning & Optimization

Montreal, Quebec
karim_ossama94@hotmail.com
+1 (438) 835-5250
LinkedIn: linkedin.com/in/karim-nadim
GitHub: github.com/karim-nadim
Website: karim-nadim.github.io

Professional Summary

Passionate about solving **business problems** using Data Science & Machine Learning. With a Ph.D. in industrial engineering, I creatively use my skillset to add **tangible value** to the team, the business, and the end-user. Strong background in statistical analysis, forecasting, and decision support using Python and SQL. I am constantly learning and always looking to improve.

SKILLS & Tools

- **Programming**: SQL (PostgreSQL), Python (Pandas, NumPy, Scikit-learn, Keras, PyTorch, TensorFlow)
- **Tools & Platforms**: Tableau, Github, AWS (S3, Lambda, IAM, EC2, SageMaker), MLflow, Docker
- **Statistics & Experimentation**: Statistical analysis, Hypothesis testing, A/B testing
- **Machine Learning**: Regression, Classification, Clustering, Time-Series Forecasting, Tree-Based Models (Random Forest, XGBoost), Causal Impact Analysis
- **Deep Learning**: Neural Networks, CNNs, LSTMs, Autoencoders
- **Generative AI**: RAG, LangChain, Prompt Engineering

WORK EXPERIENCE

Data Scientist | CanmetENERGY - Natural Resources Canada | Varennes, QC | Aug 2024 – Present

- To improve operational efficiency and reduce costs in industrial process control, I developed a recommendation system using deep learning (PyTorch), achieving a **15% reduction** in energy consumption.
- For accurate fault detection & identification in a large-scale industrial system, I built a fault diagnosis model using Decision Trees and Random Forest, achieving **93% accuracy** in identifying fault types and their root causes.
- For predictive maintenance and reducing unplanned downtime, I developed time-series forecasting models (LSTM, TensorFlow) to predict equipment degradation, reaching **R^2 up to 0.98** and enabling early intervention.
- Developed an indicator to track the health state of industrial equipment based on deep learning autoencoders **with a correlation coefficient of 0.91**, which led to enhanced equipment monitoring and anomaly detection.
- To optimize production planning for a paper machine, I developed a reinforcement learning (RL) scheduling agent (TensorFlow), delivering a **10% reduction** in operational costs.
- For enhancing information retrieval from technical industrial reports, I developed an agent using LangChain with RAG, reducing manual review time by **60%**.

Research Assistant | CanmetENERGY - Natural Resources Canada | Varennes, QC | Jan 2020 – Apr 2024

- Applied K-means clustering with Dynamic Time Warping (DTW) to distinguish normal and abnormal operating cycles in time-series data for an industrial concentrator enabling scalable anomaly identification.
- Developed an RL agent (TensorFlow) to optimize operational scheduling for bleach and pulp machines under inventory constraints, reducing electricity costs while minimizing inventory penalties.
- To detect abnormal recovery boiler operations from environmental and economic perspectives, I built an XGBoost classification model using sensor data, achieving **98% accuracy** in classifying normal vs. abnormal states.
- To understand relationships between high-dimensional process data, I applied dimensionality reduction techniques (PCA, t-SNE) to analyze links between process variables, economic, and operational KPIs.
- Published peer-reviewed research in high-impact journals, communicating applied machine learning and data-driven insights to both technical and non-technical audiences.

Projects

Measuring Campaign Effectiveness via Causal Inference
- Investigated the increase in sales after a customer subscription program by applying causal impact analysis to transactional data and constructing a counterfactual based on non-subscribing customers. The approach demonstrated a 41.1% statistically significant increase in post-campaign sales (95% confidence).

Visual Similarity Search for Product Discovery
- Built an image similarity search pipeline using deep learning-based feature embeddings and cosine similarity to retrieve visually similar products, addressing customer search ease and enabling data-driven product discovery.

LLM-Powered SQL Agent for Self-Service Business Analytics
- Designed and implemented a natural-language SQL agent using LangChain that translates plain-English business questions into validated PostgreSQL queries, facilitating self-serve analytics while enforcing safe query execution and SQL best practices.

EDUCATION
- **Doctorate in Industrial Engineering | Ecole Polytechnique de Montréal | Canada | Aug 2018 – Apr 2024**
- **Bachelor's in Mechatronics Engineering | German University in Cairo | Egypt | Jan 2012 – Jan 2017**