Luan Nguyen

778-668-3009 | luan_nguyen_3@sfu.ca | [Portfolio](#) | [LinkedIn](#) | [GitHub](#) |

TECHNICAL SKILLS

Languages: Python, C++, SQL, R, Bash, JS, HTML/CSS
ML/DS: supervised learning, regression/classification, clustering, time-series, feature engineering, model eval
Data Stack: Snowflake, MS SQL Server, PostgreSQL, dbt, Hevo, Denodo
Orchestration & Pipeline: Docker, Kubernetes, Git, GitOps CI/CD
Analytics & Viz: Tableau, Jupyter, Hex

EXPERIENCE

ML Engineer Intern | *K8s, PostgreSQL, Docker, GitOps CI/CD, FastAPI, AWS*   May. 2025 – Sep. 2025
Workstream   *BC, Canada*

- Designed and deployed a production-ready ML microservice for employee retention prediction using XGBoost, integrating data from Snowflake, PostgreSQL, and Google Maps API; automated pipelines with feature engineering, class imbalance handling, and SHAP-based model explainability.
- Architected multi-environment AWS infrastructure with Terraform, GitOps CI/CD, and Kubernetes orchestration; implemented containerized deployments, S3 model artifact versioning, and observability with Datadog for scalable, secure ML operations.
- Developed FastAPI-based inference service with dual-database architecture, asynchronous feature extraction, and 95%+ tested codebase; enforced modern Python standards, configuration management, and production monitoring with model performance tracking and audit logging.
- Built automated DBT documentation pipeline using Python and Azure OpenAI GPT-4 to generate detailed table/column metadata for 100+ Snowflake schemas, improving LLM text-to-SQL accuracy and query reliability.
- Designed multi-threaded data extraction system with Snowflake integration and YAML auto-generation, reducing manual documentation effort by 95% and enriching AI-powered SQL generation with structured schema context.

Data Scientist Intern | *Snowflake, dbt, Tableau, TF-IDF, Word2vec, Cluster*   Sep. 2024 – Apr. 2025
Workstream   *BC, Canada*

- Designed and implemented a complete Tableau-based product metric lookup system with Snowflake and dbt, standardizing metric definitions and enabling self-serve exploration for GTM and Product; reduced ad hoc request turnaround by 60% and increased dashboard adoption by 3x across teams.
- Built end-to-end ELT pipelines in Snowflake/dbt/Hevo to model customer, usage, and revenue tables with tested, documented transformations; improved query performance by 45% and cut pipeline failures by 30% through incremental models and data quality checks.
- Applied TF-IDF/Word2Vec embeddings with clustering to 50k+ sales notes, surfacing top churn/downsell themes; delivered taxonomy and playbook updates that reduced downsell rate by 8% QoQ in targeted segments.
- Designed A/B test frameworks and statistical analyses for hiring product experiments, isolating drivers of time-to-hire and candidate quality.

Data Advisor Intern | *Databrick, Azure, MS SQL Server, Tableau*   Jan. 2024 – May. 2024
Vancouver Coastal Healths, Medical Quality Practice Dept.   *BC, Canada*

- Designed and implemented ETL pipeline using PyODBC to connect MS SQL Server and Denodo, automating data extraction and transformation processes to ensure real-time updates
- Automated the encryption and distribution of quarterly performance reports for 500+ physicians using Python and Tableau API, reducing manual effort and improving reporting efficiency
- Collaborated with data scientists and engineers to develop predictive models for patient readmission, length of stay, and mortality rates, enhancing data-driven decision-making and patient care strategies

HIGHLIGHTED PROJECTS

Probabilistic Models | *TensorFlow, Probabilistic Model, Git* | [Source](#)   Aug. 2023
- Performed feature extraction and hypothesis tests using MANOVA, Chi-square test, identifying relevant attributes
- Conducted regression and time series analysis, investigate trends and patterns
- Developed binary classification models (KNN, Naïve Bayes, Logis. Regress., SVM, Neural Net.) using TensorFlow

EDUCATION/PROFESSIONAL DEVELOPMENT

BASc, 4th Year, Computer Science Major, Applied Statistics Minor   Sep. 2025
Simon Fraser University   *Burnaby, BC*
- GPA: 3.5, Dean's Honour Roll 2022-2023
- Concentration: Machine Learning
- Relevant courseworks: ML, DS, KDD (3rd year); Probabilistic Analysis (3rd year); DSA (3rd year)
- Data Analytic Certificate: Python/R data wrangling; Tableau data visualization
- Social Activities: SFU Canadian Cancer Society Chair (2022-2023)

Data Structure and Algorithm Cert.   Sep. 2023
University of California San Diego   *Coursera remote*
- Skills: Algorithms (Greedy, Divide-and-Conquer), Dynamic Programming, Debugging, Software Testing