

DEREK DUONG

DerekDuong20@gmail.com ◇ ◇ Toronto, ON

Software Engineering student with a solid understanding of software design, development, and testing. I aim to create solutions.

WORK EXPERIENCE

Machine Learning Data Analyst - Castle Ridge Asset Managment May 2025 – August 2025

- Worked on developing prompt compressor; leveraging spaCy's dependency parsing and named entity recognition compressed prompts by 35-45% while preserving semantic meaning
- Created technical documentation enabling 3 engineers to maintain prompt compressor system
- Utilized Python to bridge gap between theoretical research and practical applications of prompt compression and NLP

Machine Learning Data Analyst - Castle Ridge Asset Managment May 2024 – Dec 2025

- Finetuned & trained predictive models to facilitate internal information/document retrieval utilizing **Python, and libraries such as Numpy, Pandas, Tenserflow, HuggingFace**
- Developed and trained predictive models by leveraging key financial ratios (e.g., Quick Ratio, Debt-to-Equity Ratio, Return on Equity) to identify correlations with stock market performance.
- Analyzed the impact of financial metrics on stock performance, uncovering actionable insights to guide data-driven decision-making.

Student IT Technician - University of Guelph CCS Jan 2024 – April 2024

- Designed, implemented, and documented process improvements which resulted in increased work efficiency and results.
- Oversaw the diagnosis and replacement of failed hardware components, ensuring a tool's optimal performance through systematic testing and component evaluation, which minimized downtime and maintained high security standards.

PROJECTS

PhishGuard: Phishing Email Detection System - *Python, Scikit-learn, XGBoost, Pandas*

- Designed and built an end-to-end **machine learning pipeline** to classify phishing emails, aggregating **8 benchmark datasets** (SpamAssassin, Enron, TREC, CEAS-08, Zenodo) into a combined corpus of over **200,000 emails** for training and evaluation
- Conducted rigorous **cross-dataset generalization testing** and **chi-squared/mutual information feature analysis**, uncovering that high-capacity models were exploiting dataset-specific artifacts rather than genuine phishing signals
- Achieved **97% F1-score and 95% recall** on within-dataset test sets across all three models

SKILLS

Technical Skills	C, Java, Python, Javascript, Html, CSS, SQL
Management & Tools	Github, Trello, Slack, Git, Docker, Gradle, AWS

EDUCATION

University of Guelph
Honours Bachelors of Computing - Software Engineering (4.0 GPA)