Data Science Intern

Internal Job ID: 2025-02

GEMINI (www.geminimedicine.ca) is a unique big data platform in the Canadian healthcare landscape using advanced methods and analytics to extract and standardize data captured in hospital electronic health records. We are based out of Unity Health Toronto, St. Michael's Hospital and currently exist at 30+ hospitals in Ontario. GEMINI is a collaborative data and analytics platform for all Ontario hospitals to accelerate research and quality improvement, leading to excellent hospital care.

Position Overview:
GEMINI is seeking an undergraduate co-op student to create an automatic outlier detection tool for the various data quality control (QC) plots that are part of GEMINI's data pipeline. The goal of this project is to develop an algorithm to automate commonly performed data coverage checks, and evaluate the performance of the algorithm on various data elements.

The expected salary range for this position is $17.56 - $21.95 per hour.

Key Responsibilities:
- Participate in regular student meetings and one-on-one with your supervisor
- Learn about GEMINI's QC process and what QC plots are generated
- Conduct literature review on existing outlier detection algorithms in both fields of statistics and machine learning
- Develop the code for an algorithm in R/Python that can reliably detect gaps/drops in data coverage
- Establish meaningful performance criteria to evaluate algorithm performance (e.g., sensitivity/specificity with respect to known data coverage issues)
- Evaluate robustness and generalizability of the algorithm (e.g., through cross-validation)

Qualifications:
- Current undergraduate student in related fields such as data science, statistics, computer science, or engineering
- Strong interest in statistics, machine learning, and healthcare research
- A curious mind and a strong drive to learn
- Experience programming in at least one of R or Python is required.
- Basic knowledge in statistics and machine learning
- Knowledge, or demonstrated willingness to learn, about best practices for model evaluation (e.g., cross-validation of machine learning models)
- Eligible to work in Ontario, Canada (e.g. Canadian citizenship, permanent resident, study permit, etc.)

Time Commitment:
Full time contract (37.5 hours per week) for at least 10 weeks from May to August 2025. Student will be

expected to work in a hybrid environment with the ability to come into GEMINI's office (located at Li Ka Shing Knowledge Institute, Downtown Toronto) at least once a week.

How to Apply:
Please submit your **CV/resume, unofficial transcript**, and a **quick paragraph** on your interest in GEMINI (around 100 words) to gemini.data@unityhealth.to with your subject header "**<Your Name> Application for Position 2025-02".**