Evaluating Rust Integration for Longitudinal HbA1c Data Processing in the All of Us Research Program

Teotzin E. Pruneda
Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract

This study will evaluate whether integrating Rust into an optimized Python workflow improves the efficiency of processing longitudinal hemoglobin A1c (HbA1c) measurements from the All of Us Research Program. HbA1c reflects average blood glucose levels over approximately three months, making repeated measurements useful for examining changes over time. The public Data Browser reports approximately 212,720 participants associated with the main HbA1c measurement concept; the eligible longitudinal cohort and total measurement count remain to be determined. The proposed cohort will require at least two valid HbA1c measurements on distinct dates per participant. Records must contain a participant identifier, a valid date, a numeric result, and recognized units that can be standardized; duplicate records and records with missing required fields or incompatible units will be excluded. An optimized pandas baseline will group and chronologically sort measurements, calculate counts and mean HbA1c, and compute changes between consecutive observations. Profiling will determine whether these operations are meaningful computational bottlenecks; only identified bottlenecks will be considered for implementation in Rust. PyO3 will make Rust functions callable from Python, and maturin will build and package the Python extension, retaining the Python analysis workflow. Both implementations will process identical datasets on the same machine, with consistent CPU and memory resources, fixed thread settings, and documented software versions. Repeated benchmarks across increasing dataset sizes will measure execution time and peak memory usage, account for Python–Rust data-transfer overhead, and verify output agreement within predefined numerical tolerances. The study will determine whether any observed performance gain justifies the integration effort.