

Irai Kumaran Sivanesan

Data Science Master's Student | former Software Engineer

Victoria, BC, Canada · (+1) 672-974-2998 · iraikumarans@uvic.ca · linkedin.com/in/iraikumaran

SUMMARY

Master of Applied Data Science student at the University of Victoria, with a year and a half of production software engineering behind it. I build machine learning systems end to end: recommender ladders evaluated against real baselines, distributed graph clustering in PySpark, cost-sensitive ensemble search, and NLP extraction pipelines over live labour-market and software-repository data. Comfortable owning a problem from raw data through evaluation to a written, defensible result, including the parts that did not work.

EDUCATION

University of Victoria, Victoria, BC

2026 – Present

Master of Applied Data Science (MADS)

- Coursework: Data Mining, Big Data and Distributed Computing, Data Management and Analytics, Empirical Software Engineering, Information Theory and AI.

SRM Easwari Engineering College (Anna University), Chennai, India

Aug 2020 – Apr 2024

B.Tech, Information Technology | CGPA 8.77 / 10

TECHNICAL SKILLS

Languages	Python, SQL, Cypher, Java, TypeScript, C++
ML and Statistics	scikit-learn, matrix factorisation, BPR, node2vec, random forests, gradient boosting, Bayesian belief networks, simulated annealing, cost-sensitive learning, SBERT, TF-IDF
Data at Scale	Apache Spark (PySpark), MapReduce, pandas, NumPy, Neo4j, MySQL, MongoDB
Analysis	Power BI, matplotlib, seaborn, experiment design, grouped cross-validation, effect-size reporting, fairness auditing
Engineering	Angular, Spring Boot, React, REST APIs, Docker, Git, Google Cloud Platform, AWS

DATA SCIENCE PROJECTS

SkillBridge: Canadian Labour-Market Skill Recommender

2026

Python, scikit-learn, BPR / matrix factorisation, node2vec, SBERT, pandas

- Merged five national data sources into a single modelling base: 900 occupation competency profiles (181 descriptors each), four months of Job Bank postings, LinkedIn Canada postings and skills, employment projections, and the NOC taxonomy; ETL removed 51,103 duplicate records.
- Reframed the task after finding the original link-prediction framing degenerate (162,899 of 162,900 cells observed, so every answer was yes) into ordinal matrix completion with a meaningful 14.1% positive-rate binary target.
- Built and benchmarked a five-model ladder against real baselines; BPR won at nDCG@5 0.617, MRR 0.821 and ROC-AUC 0.99 on the core-skill task.
- Shipped three further components: a labour shortage classifier (weighted F1 0.76), a salary and regional-demand regressor (MAE \$12,088, R2 0.64 against a 0.4 target), and a hybrid NLP extractor mapping free-text postings onto the competency taxonomy (micro-F1 0.37); audited fairness across TEER skill levels and documented every failure mode.

Merged Without Review: What AI Coding-Agent Acceptance Rates Measure

2026

Python, pandas, logistic regression, random forests, mining software repositories

- Analysed 33,596 pull requests authored by five AI coding agents across GitHub repositories, testing whether merge rate reflects code quality or the review culture of the host repository.
- Measured engagement by distinct human reviewers rather than comment counts, after showing that bots and agent self-comments inflate the latter.
- Found 82.6% of decided pull requests, and 81% of merged ones, were settled with no human review at all.
- Showed the tool-to-outcome association collapses from Cramer V 0.322 to 0.105 once restricted to reviewed PRs, and that agent identity adds only +0.053 ROC-AUC over repository context.
- Used repository-grouped cross-validation throughout and reported effect sizes rather than p-values, since at n = 31,679 every significance test is uninformative.

ThresholdX: Cost-Aware Ensemble Search for Customer Churn

2026

Python, simulated annealing, Bayesian belief networks, scikit-learn

- Formulated churn on 7,043 telecom customers under an explicit asymmetric business cost (a false negative penalised five times a false positive) instead of accuracy, which a do-nothing classifier already scores 73.5% on.
- Replaced the conventional stacking meta-classifier with simulated annealing searching a five-dimensional space of ensemble weights and the decision threshold jointly against that cost, reaching a test cost of 613 against 631 for a fairly configured baseline.
- Attributed 87% of the achievable gain to the decision threshold and only 13% to weighting via a controlled ablation, tracing the limit to 0.908 correlation between base learners, and reported the negative result that the system ties a tuned logistic regression under a corrected resampled t-test.

Demographic Cybercrime Knowledge Graph

2026

MySQL, Neo4j, Cypher, Power BI, Python

- Designed a three-tier analytics pipeline over a 5,000-record victim dataset calibrated against eleven government and industry reports across nine countries.
- Built a normalised four-table star schema in MySQL, a Neo4j property graph of 5,040 nodes and 10,000 directed relationships, and a six-view Power BI dashboard for non-technical stakeholders.
- Surfaced cross-dimensional findings invisible to aggregate analysis: victims aged 60 and over absorb 36% of all financial loss while forming a quarter of the victim pool, and business email compromise drove \$121.9M of impact from only 122 cases.

Distributed Vulnerability Clustering with PSCAN in PySpark

2026

Apache PySpark, MapReduce, graph clustering

- Reframed victim profiling as graph clustering, building a weighted similarity graph over nine demographic and behavioural dimensions and running PSCAN.mr end to end in PySpark.
- Identified nine vulnerability clusters and classified all victims as cores (55.8%), hubs (21.0%) or outliers (23.2%), profiling each cluster by dominant demographics and average financial loss.
- Ran a systematic parameter sweep proving $\epsilon = 0.9$ is a real structural boundary and that demographic features alone collapse the clustering, validating the multi-dimensional framing.

Enhancing Healthcare and HCI with Conversational AI

Sept 2023

NLP, NLU, LSTM, ClinicalBERT, Hidden Markov Models

- Built a conversational system for hospital management, combining LSTM-based contextual understanding with Hidden Markov Model intent recognition and ClinicalBERT for clinical data retrieval.
- Integrated it with the hospital database for real-time access to patient records and scheduling, secured with role-based access control. Placed third at Project Day 2024 among 100+ projects.

EXPERIENCE

Laboratory Teaching Assistant, ECE 216 Electricity and Magnetism

May – Aug 2026

Dept. of Electrical and Computer Engineering, University of Victoria

- Teach the theory behind each weekly lab to a section of second-year engineering students, then hand over the bench so they run and record the experiment themselves.
- Grade submitted lab reports and write feedback that explains the mark, so the next report improves.
- Responsible for section lab safety under the ECE Laboratory Safety Guidelines.

Software Engineer, Tata Consultancy Services

June 2024 – 2025

Chennai, India | Banking and financial services client

- Rebuilt a legacy admin web application onto a modern Angular micro front-end architecture with dynamic forms, cutting page load time by 20%.
- Profiled and optimised the reporting query path, reducing database query execution time by 30%.
- Integrated cloud, AI and blockchain components with custom encryption, and built the Spring Boot services behind them, inside a full review and release process with cross-functional teams.

Web Development Intern, Engine Factory Avadi

Aug 2023 – Feb 2024

Chennai, India | Government defence manufacturing

- Delivered the L.T.C. Management System as part of the organisation's digital transformation.
- Built a Vaccination Management System used across the organisation during the COVID period.
- Stack: HTML, CSS, PHP, jQuery, IBM server infrastructure.

Freelance Front-End Developer

Jan – June 2023

Remote, various clients

- Built responsive Angular applications with RxJS-driven real-time updates, working to Agile delivery cycles and integrating directly with client back-end teams.

ACHIEVEMENTS AND CERTIFICATIONS

- **Third Place, Project Day 2024** (Feb 2024): conversational healthcare AI prototype, cash prize, 100+ competing projects.
- **Runner-up, Kavach Hackathon** (May 2023): fund trail analysis tool applying Markowitz portfolio optimisation to investment risk assessment.
- **Third Place, Debugging Event** (May 2022): placed against 400 students across multiple languages.
- **Organised technical events**, National-Level Symposium, Easwari Engineering College (Mar 2023): led hackathons and coding competitions for 50+ students.
- **Certifications:** Social Networks, NPTEL IIT Ropar (network analytics and social graph theory); Privacy and Security in Online Social Media, IIIT Hyderabad.