NIHIRA SHARMA

Sydney, NSW | +61 492 968 975 | nihirasharma911@gmail.com | LinkedIn | GitHub | Portfolio

EDUCATION

Bachelor of Advanced Computing (Data Science & Business Analytics) – University of Sydney
Expected Graduation: June 2027
Relevant Coursework: Data Structures & Algorithms, Introduction to AI, Predictive Analytics, Business Forecasting, Statistical Modelling for Business, Scalable Data Management (PySpark/Databricks), Data & Information Management

- **Data Science Capstone – High Distinction:** tested whether bid-ask spread improves short-horizon volatility forecasts on 35M+ high-frequency Optiver order-book rows across 10 stocks and 5 models (HAR-RV, ARMA/ARMAX, LASSO, XGBoost, LightGBM). Found the signal Granger-causal but accuracy-degrading out-of-sample (confirmed via Diebold-Mariano tests), with baseline ARMA the best-performing model
- **Industry & Community Project – Tata India, "Accelerating AI Adoption for India's Largest Online Grocer" (BigBasket) – Distinction:** identified a disconnect between BigBasket's recommendation engine and inventory system and proposed an integrated AI decision layer connecting customer recommendations, inventory and frontline action; benchmarked against comparable retail AI deployments *(Woolworths: up to 40% food-waste reduction, 21% productivity gain)* to project meaningful waste reduction and margin recovery

TECHNICAL PROJECTS

Credit Risk Intelligence Platform – *R, PostgreSQL, Shiny*

- Built a logistic-regression credit scorecard (WOE/IV feature selection, PDO scaling) on 307,511 applicants aggregated from 58M+ rows across 7 relational tables in PostgreSQL
- Benchmarked against XGBoost and selected the logistic model for interpretability and regulatory transparency, after catching and dropping 6 collinear count features that aliased to NA coefficients
- Validated with ROC/calibration analysis (MSE 0.0001), train-test population stability (PSI ≈ 0) and gender/age fairness checks confirming no added disparity across groups
- Deployed a live 4-page Shiny dashboard (performance, scorecard explorer, interactive risk tool, insights) backed by 28 unit tests

Fraud Risk Intelligence Platform – *PostgreSQL, Python, XGBoost*

- Built an expected-loss ranking system (probability of fraud × transaction amount) that turned the model into a decision tool, not just a classifier: **the top 1,000 of 2,730 daily alerts caught 75% of fraud and ~$529K of exposure**
- Compared multiple machine learning models and selected XGBoost (PR-AUC 0.883), lifting precision from 53.5% to 99.5% at equal recall
- Engineered fraud signals – per-card spending z-scores, transaction velocity, and distance between transaction locations – directly in SQL as PostgreSQL materialized views across 1.3M transactions
- Deployed dual dashboards (HTML/JS and Tableau) for real-time monitoring

Stock Trend & Volatility Insights – *R, Shiny, rugarch, tidyquant*

- Built a GARCH family model suite (sGARCH/eGARCH/gjrGARCH) with BIC-based selection and a news-impact curve to capture the leverage effect in volatility shocks
- Computed VaR and CVaR with Kupiec proportion-of-failures backtesting and an ARCH-LM pre-test for volatility clustering; validated 10–60 day GARCH forecasts out-of-sample (80/20 split) against a constant-volatility benchmark
- Deployed a live 5-page dashboard (trends, risk, volatility, multi-ticker compares) covering US, index and ASX tickers, with two-layer caching for instant reloads

A/B Testing & Decision Analysis Dashboard – *Python, Streamlit, SciPy*

- Tested a real product decision – whether to ship a proposed e-commerce change – using frequentist and Bayesian A/B testing side by side: correctly returned a **do-not-ship recommendation** on 294,000+ real observations from the public Udacity dataset
- Built the underlying engines: two-proportion z-test, confidence intervals, power/MDE (frequentist) and Beta-Binomial posterior with expected loss (Bayesian)
- Added Sample Ratio Mismatch health checks and a Decision Centre translating statistics into ship/no-ship calls with revenue-impact projections, plus automated PDF/Markdown reporting

Bank Churn Intelligence Platform – *Python, XGBoost, SHAP, Streamlit*

- Built an XGBoost churn model (ROC-AUC 0.875, beating Random Forest at 0.865 and Logistic Regression at 0.778) on 10,000 customers, and explained individual and portfolio-wide churn drivers with SHAP waterfall and beeswarm plots
- Caught a data-leakage bug – a complaints field alone pushed ROC-AUC to ~1.0 – and excluded it, keeping the model honest
- Built a Revenue-at-Risk framework (~$105.9M exposure estimated) and K-Means customer segmentation (4 risk segments), deployed across a six-page Streamlit dashboard on Streamlit Cloud

NLP Resume Matching Engine (In Progress) – *Python, PostgreSQL + pgvector, spaCy, NLP, sentence-transformers, Reflex*

- Built an ATS-style resume-to-job matching engine that scores resume against live postings (Adzuna/Reed APIs) using a transparent weighted blend of keyword overlap (BM25), title, seniority and experience match and semantic similarity (sentence-transformer embeddings over pgvector) – explainable by design, not black-box
- Added an ATS-parsability check that flags formatting (tables, Columns, images, headers/footers) real ATS engines silently drop, plus spaCy NER skills/experience extraction against a combined ESCO/O*NET taxonomy
- Designed the Anthropic API integration for rewrite suggestions; validated end-to-end on real postings and resumes (not synthetic data) behind a multi-page Reflex UI

LEADERSHIP & ACTIVITIES

- Peer Mentor, University of Sydney – supported new students through academic transition and course navigation
- Team Lead across group assignments in data science and computing coursework; collaborated in interdisciplinary teams for the Capstone and Tata/BigBasket Industry Project

SKILLS

Languages & Tools: Python, R, SQL, PySpark, Streamlit, Shiny, Power BI, Tableau, Git
Machine Learning & Statistics: XGBoost, Random Forest, Logistic Regression, SHAP, GARCH, A/B Testing, Bayesian Inference
Domains: Credit risk, fraud detection, quantitative finance & volatility modelling, NLP