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 (Optiver volatility forecasting: ARMAX, HAR-RV modelling, team coordination)
- 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 AI-driven approach to align product suggestions with real-time stock availability.

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*

- Engineered fraud features (per-card amount z-scores, trailing-window velocity counts, haversine distance) entirely in SQL as PostgreSQL materialized views across 1.3M transactions
- Selected XGBoost (PR-AUC 0.883) from a 5-model bake-off against LightGBM, Random Forest, Logistic Regression and Isolation Forest on a time-based train/test split, lifting precision from a 17% rule-engine baseline to 99.2% at equal recall
- Built an expected-loss ranking (P(fraud) × amount): top 1,000 of 2,730 alerts catch 75% of fraud and ~$529K of exposure; deployed dual dashboards (HTML/JS + Tableau)

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*

- Built frequentist (two-proportion z-test, confidence intervals, power/MDE) and Bayesian (Beta-Binomial posterior, expected loss) experimentation engines side by side
- Added Sample Ratio Mismatch health checks and a Decision Centre translating statistics into ship/no-ship calls with revenue-impact projections
- Validated on 294,000+ real e-commerce observations from the public Udacity dataset, correctly returning a do-not-ship recommendation; added automated PDF/Markdown reporting

NLP Resume Matching Engine (In Progress) – *Python, NLP, ESCO Taxonomy*

- Building an ATS-scoring and skills-extraction engine using the ESCO taxonomy and live Adzuna job data to match resumes against real job postings
- Completed live job-data integration and taxonomy loading; currently building document parsing, ATS scoring and named-entity extraction for skills and experience

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