

PARAS SIMKHADA

London, UK · +44 (0) 7352 130293 · simkhadaparash@gmail.com · [linkedin.com/in/paras-simkhada](https://www.linkedin.com/in/paras-simkhada)

PROFESSIONAL SUMMARY

Data Scientist with a strong statistical foundation, 2+ years' experience teaching Statistics to 1,000+ undergraduate and postgraduate students, and internship experience preparing share-market data at NIC Asia Capital. Currently completing an MSc in Data Science at the University of Greenwich. Builds end-to-end regression, classification, NLP and time-series pipelines in Python, R and SQL — recent work reached $R^2 = 0.98$ on lifespan prediction and 92.5% accuracy on defect classification. Translates analytical findings into decision-ready recommendations for technical and non-technical audiences. Seeking a graduate or junior Data Scientist role focused on predictive modelling and measurable business impact.

TECHNICAL SKILLS

Languages & Data	Python, R, SQL, SQLite, Neo4j (Cypher), Pandas, NumPy, data cleaning
Machine Learning	Scikit-learn, regression, classification, Random Forest, Gradient Boosting, Decision Trees, K-Means, Isolation Forest, SMOTE, feature selection, hyperparameter tuning, SHAP
Statistics & Time Series	Statistical modelling, predictive analytics, ARIMA, stationarity testing (KPSS, ACF/PACF), evaluation (RMSE, MAE, AIC/BIC, F1, AUC)
NLP	spaCy, Sentence-BERT, TF-IDF, semantic similarity, text classification
Visualisation	Matplotlib, Seaborn, Streamlit, graph/network analytics (NetworkX)

PROFESSIONAL EXPERIENCE

Intern — NIC Asia Capital May – Aug 2025
Kathmandu, Nepal

- Assisted with share-market data work, organising and preparing market information for analysis.
- Supported simple analysis and reporting for the team using the prepared data.

Statistics Educator 2+ years
Undergraduate & postgraduate programmes

- Taught Statistics to 1,000+ students across BSc, BBA, MBA, MPA and MBS programmes.
- Explained statistical concepts through practical, clear examples for learners from technical and non-technical backgrounds.

SELECTED PROJECTS

Anti-Money Laundering Analytics & Fraud Detection Platform

Python · Scikit-learn · Neo4j / Cypher · Isolation Forest · K-Means · SMOTE · Graph analytics

- Developed an AML-as-a-Service solution combining machine learning, graph analysis and privacy-aware system design.
- Implemented fraud detection with Decision Trees, Isolation Forest, K-Means and SMOTE to handle severe class imbalance.
- Used Neo4j and Cypher to surface suspicious transaction networks, shared identities and round-tripping behaviour.

Manufacturing Lifespan & Defect Prediction

Python · Scikit-learn · Gradient Boosting · Random Forest · SHAP

- Built regression and classification pipelines on 1,000 manufacturing records with preprocessing, feature selection and hyperparameter tuning.
- Achieved $R^2 = 0.98$ for lifespan prediction, and 92.5% accuracy with $F1 = 0.93$ and $AUC = 0.99$ for defect classification.
- Applied SHAP to explain the drivers of component failure to a non-technical audience.

UK Fire Safety Regulatory Guidance Comparison System

Python · Sentence-BERT · TF-IDF · SQLite · Streamlit

- Built an NLP system comparing 2,873 regulatory clauses across UK fire-safety guidance versions and jurisdictions.
- Combined Sentence-BERT semantic similarity with TF-IDF scoring to align and flag divergent clauses, storing results in SQLite for query and audit.
- Delivered a Streamlit interface letting non-technical users search clauses and view side-by-side differences.

S&P 500 Time-Series Forecasting

R · ARIMA · KPSS · ACF/PACF · Ljung-Box

- Analysed daily S&P 500 data from 2010–2025 with log transformation and stationarity testing.
- Compared manual and auto-selected ARIMA models on AIC, BIC, log-likelihood, RMSE and MAE across an 80/20 train-test split.

Venue Performance & Visitor Analytics

Python · Time-series analysis

- Analysed daily visitor and venue-performance data using segmentation, 7-day rolling averages, heatmaps and box plots.
- Identified visitor trends, anomalies and venue openings/closures, surfaced through interactive dashboards.

Algorithmic Impact Assessment — Police Risk-Scoring System

Responsible AI · Data ethics · Algorithmic fairness · Governance

- Evaluated a data-driven policing and risk-scoring system against fairness, transparency and accountability principles.
- Identified governance risks and proposed review, challenge and community-oversight mechanisms.

EDUCATION

MSc Data Science — University of Greenwich, London

Sept 2025 – Sept 2026

Predicted: First Class with Distinction

BSc Statistics — Tribhuvan University, Nepal

Graduated 2023

Distinction

CERTIFICATIONS

Data Analysis with Python · Machine Learning Fundamentals — Mindrisers Institute of Technology

REFERENCES

Dr Ram Prasad Bhatta, PhD — Postdoctoral Researcher, Department of Engineering Science, University of Oxford · ram.bhatta@eng.ox.ac.uk