Manoj Chandra

+44 7944 470201 | manojchandra910@outlook.com | manojchandra10 | manojchandra10 | Nottingham, UK

Data Scientist, MSc Data Science, University of Nottingham
Predictive modelling in Python and R, data preparation in SQL, reporting in Power BI

Right to work: full-time work permitted now; Graduate visa application in November 2026 (two years).

Education

MSc Data Science, Merit (expected)
09/2025 – 09/2026

University of Nottingham

- **Dissertation**: An Empirical Study of Mondrian Forests and Online Tree Classification Methods under Sample Size, Label Noise and Feature-Space Outliers (report 80%; project plan 81%)
  Mondrian forests implemented from scratch in Python and benchmarked against CART and random forests in paired simulations; tuning the lifetime reduced excess error 7- to 23-fold
- **Modules**: Statistical Machine Learning (80%), Data Science with Machine Learning (78%), Time Series and Forecasting, Databases, Big Data Technologies, Interface and Software Design Principles
- **Scholarship**: South Asia Postgraduate Excellence Award

UPSC Civil Services Examination Preparation
06/2022 – 05/2025

Sriram's IAS, Delhi, India

- Undertook full-time preparation for India's Civil Services Examination, run by the Union Public Service Commission: of 992,599 applicants in 2024, 1,009 were recommended for appointment
- Maintained preparation across three annual cycles of 10 months each on a syllabus that follows current affairs: the economy, governance and public policy, for nine essay-type papers and an interview

BSc (Hons) Mathematics, First-Class Honours (CGPA 7.53/10)
06/2019 – 07/2022

University of Delhi, India

- **Modules**: Probability, Statistics, Mathematical Finance, Discrete Mathematics, Mathematical Modelling
- **Scholarship**: INSPIRE (Government of India), for a top 1% result in the Class XII board examinations

Projects

Customer Repurchase Prediction for a UK Online Retailer
2026

Portfolio project, SQL, Python (scikit-learn) and Power BI (DAX)

- Predicted which of [n] customers would buy again within 90 days with a ROC AUC of [0.xx], against [0.xx] for a recency rule, by training gradient-boosted trees in Python on features built in SQL
- Prevented look-ahead leakage by building every feature in SQL (CTEs, LAG, window aggregates) from purchases before a fixed cut-off date, and repeated the evaluation on an earlier cut-off (ROC AUC [0.xx])
- Captured [n]% of the customers who lapsed within the top 20% of risk scores, and delivered the ranked list in a Power BI report with DAX measures for revenue at risk by market

Road-Collision Severity Prediction
Spring 2026

Research paper, Data Science with Machine Learning, Python

- Identified random forest as the best of 4 classifiers on 29,628 UK police collision records: macro-F1 0.555 ± 0.014, significantly above the other three ($p = 0.002$; Wilcoxon tests, Bonferroni correction)
- Prevented data leakage by fitting preprocessing inside each fold of a 10×3 nested cross-validation
- Developed an analysis that Dr Jeremie Clos chose as a teaching example for the 2026–27 cohort

Contaminated Chemical Detection

Spring 2026

Statistical Machine Learning group project (Distinction), R

- Achieved first place among 3 groups on the prediction leaderboard by engineering an R pipeline that predicted the impurity type and percentage of 100 chemical samples
- Reduced the error score by 33% by replacing a random forest with per-type linear models after exploratory analysis showed a near-linear link to one spectral channel

Work Experience

Student Consultant (Data Analyst)

06/2026 – 07/2026

Hermai.ai, University of Nottingham micro-placement (Practera)

- Delivered the World Cup Sleep Index for Hermai.ai in 2 weeks: cleaned the 104-match fixture data, mapped all 48 fanbases to home time zones and launched fifaworldcupsleep.me
- Integrated the co-founder's direction into the build: a shareable live site in place of the scheme's market report, with one data source, fresh data, one personal number per fan and a launch post
- Coordinated a Hermai API feed of FIFA match data with the co-founder and implemented a 6-hourly GitHub Actions rebuild that halts publication if any of the 104 fixtures fails its checks
- Recommended reporting average sleep lost per match beside the total, since totals favoured teams that played more matches; the live site reports both

Data Science and Business Analytics Intern

05/2024 – 07/2024

The Sparks Foundation

- Analysed data using SQL JOINs, CTEs and aggregate functions to identify monthly sales trends and compare performance across stores and product categories.
- Cleaned and reconciled sales and product records in Excel using Power Query, XLOOKUP and PivotTables to resolve data discrepancies and prepare accurate datasets for reporting.
- Developed Power BI dashboards with DAX measures and interactive slicers to track sales performance, compare product categories and consolidate monthly sales report.

Private Mathematics Tutor

2019 – 2025

Self-employed, India

- Instructed Class 6–12 students one to one and in small groups, from algebra to calculus

Skills

Machine learning	Random forests, gradient boosting, logistic regression, SVM, neural networks
Evaluation	Nested cross-validation, ROC AUC, macro-F1, leakage control, simulation studies
Statistics	Paired t and Wilcoxon signed-rank tests, Holm and Bonferroni corrections, SARIMA
Programming	Python (pandas, NumPy, scikit-learn), R, SQL (SQL Server, MySQL), PySpark
BI and tools	Power BI (DAX, Power Query), Excel, Tableau, Git, GitHub Actions, Databricks

Activities and Interests

Membership: Student member, Institute of Mathematics and its Applications (2025–2026)
Hiking: solo and with friends, as well as regular Peak District walks and group route planning with the University of Nottingham Rambling and Hiking Society
Travel: solo trips to visit and explore new places