Nabankur Ray

Data Scientist | ML Engineer | AI Solutions Developer

ray.nabankur@gmail.com | +91 88373 49959 | Kolkata, India linkedin.com/in/nabankur-ray-876582181 | github.com/nabankur14

PROFESSIONAL SUMMARY

Results-driven Data Scientist with a Master's in Data Science (Deakin University, pursuing) and a PGP in Data Science & Business Analytics (UT Austin & Great Lakes, GPA 3.86/5). Proficient in end-to-end ML pipelines — from EDA and feature engineering to model deployment — across domains including FinTech, Healthcare, Retail, Sports, and Cloud AI. Former Production Planning & Control Engineer with 2+ years at Delta Electronics India, bringing a rare blend of manufacturing operations expertise and advanced data science capabilities. Demonstrated ability to translate complex data insights into actionable business strategies.

TECHNICAL SKILLS

Languages	Python (Advanced), R, SQL, PySpark, JavaScript (Basic)
ML / DL	Scikit-learn, XGBoost, LightGBM, Keras/TensorFlow, Random Forest, SVM, Ensemble Methods
NLP & Gen AI	Transformer Architecture, LLM Fine-tuning, Prompt Engineering, Azure OpenAI (GPT)
Cloud & Azure	Azure Cognitive Services, Azure ML, Azure Bot Service, Azure AI Search, Azure App Service
BI & Viz	Tableau, Power BI (DAX, Power Query), Matplotlib, Seaborn, Plotly
Data Eng.	Pandas, NumPy, PySpark, XML Parsing, REST APIs, Streamlit
Methodologies	EDA, A/B Testing, Hypothesis Testing, Time Series (ARIMA/SARIMAX), Clustering, RAG
Tools	Git/GitHub, SAP, Google Colab, Jupyter, KNIME, VS Code

WORK EXPERIENCE

Production Planning & Control Engineer

Delta Electronics India Pvt. Ltd. | Jul 2021 – Jun 2023 | Hosur, Tamil Nadu

- Led a team of 4, successfully executing a full factory Transfer Project while maintaining 100% on-time delivery commitments across Solar Inverters (PVI), VFD Drives, and Power Supplies.
- Managed customer forecast-to-Master Production Schedule workflow in SAP, improved production efficiency by 12% and reduced dispatch delays by 15% through data-driven scheduling.
- Controlled BOM engineering changes (ECN) through cross-functional team (CFT) meetings; maintained live inventory accuracy via MRP/MPS and real-time SAP monitoring.
- Coordinated purchasing from international supply base; sustained on-time delivery rate during rapid growth, covering 23 months of multi-product group responsibility.

Data Analytics & Visualization Virtual Internship

Accenture North America (via Forage Job Simulation) | November 2024

- Cleaned, modelled, and analysed multiple social media datasets to uncover content performance insights for a simulated client engagement.
- Designed stakeholder-facing dashboards and a client presentation translating data-driven findings into strategic business recommendations.

EDUCATION

Master of Data Science

Deakin University, Australia | Jul 2025 – Jul 2026 (Currently Pursuing)

Coursework: Machine Learning, Data Wrangling, Engineering AI Solutions, Operations Research

PGP — Data Science & Business Analytics

University of Texas at Austin & Great Lakes Institute of Management (Dual Degree) | May 2024 – May 2025 | GPA: 3.86 / 5.0

Coursework: Predictive Modelling, Unsupervised Learning, Time Series, Tableau, SQL, Market Analytics

B.E., Electrical & Electronics Engineering

Thiagarajar College of Engineering, Madurai | Jul 2017 – May 2021 | CGPA: 9.1 / 10

Relevant: Signals & Systems, Embedded Systems; Minor in IoT & Robotics

KEY PROJECTS

BCCI Cricket Win Prediction — Capstone (Score: 94/100)

Tools: Python, Scikit-learn, GridSearchCV, Logistic Regression, Decision Tree, KNN

- Built 4 supervised ML classifiers on 2,930 records; Tuned Decision Tree achieved 95.5% recall and 0.927 F1-score.
- Applied VIF-based multicollinearity removal, p-value elimination, and ROC threshold optimisation (0.847) for production-grade performance.
- Delivered 5 opponent-specific match-winning strategies for BCCI using feature importances across 3 upcoming series.

Hotel Booking Cancellation Prediction (Score: 60/60)

Tools: Python, Scikit-learn, Logistic Regression, KNN, Decision Tree, ROC-AUC

- Built cancellation prediction system on 36,275 INN Hotels records; Tuned Decision Tree achieved 85.4% accuracy and 82% recall (AUC = 0.86).
- Resolved multicollinearity via VIF analysis, iterative p-value elimination, and optimal threshold tuning (0.333).
- Recommended dynamic cancellation fee policies and targeted loyalty strategies to reduce last-minute revenue loss.

Time Series Wine Sales Forecasting (Score: 60/60)

Tools: Python, ARIMA, SARIMAX, Exponential Smoothing, statsmodels, pmdarima

- Engineered an end-to-end forecasting pipeline on 15 years of ABC Estate Wines sales data (Rose & Sparkling variants).
- Conducted seasonal decomposition, ADF stationarity tests, and RMSE-based model selection across ARIMA, SARIMAX, and ETS models.
- Delivered executive recommendations on peak-season inventory planning and targeted marketing ROI optimisation.

PySpark Business Review Analytics (Score: 88/100)

Tools: PySpark, NLP, ALS Collaborative Filtering, ARIMA, CountVectorizer

- Built scalable analytics pipeline on 500K+ Google Maps reviews; ALS recommendation system achieved RMSE of 0.9896.
- Applied NLP (CountVectorizer, WordCloud) to identify top 30 keywords and segment low-rating reviews for dissatisfaction root-cause analysis.
- Implemented ARIMA(1,0,1) time series forecasting with grid search across 27 parameter combinations, revealing COVID-19 impact on review activity.

Visa Approval Classification (Score: 51/60)

Tools: Python, XGBoost, Gradient Boosting, Random Forest, SMOTE, RandomizedSearchCV

- Developed a binary classification pipeline on 25,480 U.S. visa records; Tuned Gradient Boosting achieved recall 0.758 and precision 0.816.
- Addressed 67:33 class imbalance via SMOTE, undersampling, and recall-optimised hyperparameter tuning.

Azure CV Object Detection & Tracking — MDS

Tools: Azure Computer Vision API, Python, OpenCV, PASCAL VOC, IoU Tracking

- Built end-to-end cloud AI pipeline achieving F1 = 1.000 and Mean IoU 0.73–0.84 across object classes on real-world urban street scenes.
- Engineered IoU-based greedy tracking algorithm across 192 video frames (Track F1 = 0.821), reducing API consumption by 80% via frame subsampling.

ShopSmartBot — Multimodal AI E-Commerce Recommendation Bot — MDS

Tools: Azure OpenAI GPT, Azure Bot Framework SDK v4, RAG, Azure AI Search, Python aiohttp

- Architected 5-layer multimodal AI bot (text, voice, image) achieving sub-3s response latency and SUS usability score of 81.5.
- Implemented RAG pipeline with hybrid keyword + vector retrieval, grounding all LLM responses in live product catalogue data.

Finance & Retail Analytics — Credit Default & Market Risk (Score: 90/90)

Tools: Python, Logistic Regression, Random Forest, VIF, KNN Imputation, ROC-AUC

- Developed binary classification model on 50+ balance sheet features; Tuned Random Forest achieved 72% accuracy on imbalanced data.
- Quantified risk-return tradeoffs on 8-year weekly price data for 5 Indian equities (ITC, Airtel, Tata Motors, DLF, Yes Bank).

CERTIFICATIONS & PROFESSIONAL DEVELOPMENT

Generative AI Certification — *Great Learning, 2025*
Transformer architecture, LLM fine-tuning, multimodal AI, prompt engineering, responsible AI.

Data Analytics & Visualization Simulation — *Accenture North America (Forage), 2024*
Multi-dataset EDA, dashboard design, client-facing strategy report.

Digital Marketing Certification — *Internshala, 2020*
SEO, SEM, Google Analytics, Email Marketing, Social Media, Mobile Marketing.

Python for Data Science — *Cognitive Class (IBM), 2020*
Core Python, data manipulation, data visualisation fundamentals.

LANGUAGES

English (Professional) | Hindi (Professional) | Bengali (Native) | Assamese (Conversational)

GitHub: github.com/nabankur14 | Tableau Public: public.tableau.com/profile/nabankur.ray1543 | LinkedIn: linkedin.com/in/nabankur-ray-876582181