

Avik Biswas

Principal Data Scientist

📍 Bengaluru, India | ☎ +91-8095706732 | ✉️ aviksagi@gmail.com | [in/avik-biswas-stats/](https://www.linkedin.com/in/avik-biswas-stats/)

Profile Summary

Over 15 years of experience in AI/ML, Data Science, and Advanced Analytics across Automotive, Manufacturing, and Retail, with specialization in IoT/Telematics, Predictive Modeling, Machine Learning, Deep Learning, and Statistical Analysis. Skilled in developing LLM-powered applications and GenAI solutions using LangChain, RAG architectures and PEFT-based fine-tuning (LoRA/QLoRA), along with knowledge of deploying scalable AI/ML models using MLOps practices and production monitoring. Proven ability to lead and mentor data scientists, and to collaborate effectively with data engineering and product teams to deliver end-to-end analytics solutions within enterprise platforms.

Technical Skills

- **Programming & Data:** Python, R, MongoDB, Pandas, Numpy, Scipy, Scikit-learn, Statsmodels, PySpark, Multiprocessing
- **Statistics & Analytics:** Predictive Modelling, Exploratory Data Analysis (EDA), Time Series Forecasting, Principal Component Analysis (PCA), Bayesian Estimation, Linear Programming/Optimization, Anomaly/Outlier Detection
- **Machine Learning:** Regression, Classification, Random Forest, SVM, XGBoost, Feature Engineering
- **Deep Learning:** Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM)
- **Natural Language Processing (NLP):** Text Classification, Sentiment Analysis, Topic Modelling (using LDA)
- **Generative AI / LLMs:** LLaMA 3.1, Gemma 3, LangChain, Retrieval-Augmented Generation (RAG), PEFT-based fine-tuning (LoRA/QLoRA), Embedding Models (BERT, Word2Vec)
- **Visualization & Reporting:** Matplotlib, Seaborn, Plotly, R Shiny, Python Dash, Interactive/Dynamic Dashboards
- **Systems & Deployment:** FastAPI with Uvicorn; knowledge on MLOps practices

Key Strengths & Impact

- Delivered ~\$4M (4–5%) annual fuel savings and significant CO₂ reduction through fleet replacement optimization.
- Improved telematics data accuracy by ~30% and strengthened customer confidence in platform outputs.
- Work closely with leadership, engineering and product team to align priorities, streamline decision-making & accelerate delivery.

Professional Experience

DARBY TELEMATICS | PRINCIPAL DATA SCIENTIST | APRIL 2024 – PRESENT

Developed analytics modules for **C6 Insights**, a climate tech and fleet optimization platform, that helps customers reduce fuel spend, measure and report carbon emissions, and evaluate ROI on clean technologies to achieve net-zero goals.

- Designed and developed **Linear Programming (MILP)** and **Probability** based vehicle replacement optimization model and roadmap, enabling fleets to maximize cost savings (~\$4M annually) while maintaining operational continuity.
- Led the development of an **LLM-driven summarization engine** using **transformer fine-tuning, RAG workflows, embeddings** for event clustering and **prompt engineering/few-shot learning** to convert high-frequency telematics data into weekly natural-language driver behavior insights—**reducing manual reporting by 70%** and saving ~4% fleet fuel savings.
- Developed a **Multiple Linear Regression** model for fuel consumption incorporating driver behaviour (idling, acceleration, braking) and telematics signals, improving customer fuel-efficiency insights and driving measurable savings across fleet.
- Built a **Polynomial Regression model** for ICE vehicle fuel economy using speed, behaviour, and terrain data; introduced a fleet-level scoring system that allowed customers to benchmark performance and take targeted improvement actions.
- Designed a **Nonlinear Depreciation model (sigmoid-based)** to estimate vehicle resale value, supporting data-driven decisions for value optimization and replacement planning.
- Created a **financial scenario modelling tool** (fuel switching, technology upgrades, vehicle replacements) using **ROI** and **Payback Period** analysis; integrated directly into the platform so customers could evaluate strategies in one place.
- Implemented **IQR based Anomaly Detection** to identify irregularities in telematics-reported data, improving the reliability of platform's (**improved data accuracy by ~30%**) analytics and strengthening customer confidence in platform outputs.

DANLAW INC. | PRINCIPAL DATA SCIENTIST | OCTOBER, 2018 – MARCH, 2024

Developed analytics modules for **DeepView Analytics**, a telematics platform, that processes high-frequency OBD-II data from vehicles to deliver insights into vehicle health (ICE and EV), driver behaviour, road conditions and fleet performance.

- Built a **Convolutional Neural Network-based model** using accelerometer and speed data, enabling the platform to detect rough terrain and improve road condition assessments for fleet operators.
- Developed **regression-based EV range prediction models** incorporating battery capacity, speed, terrain, and temperature, empowering customers to plan trips and reduce range-anxiety for electric fleets.
- Designed **algorithms leveraging State of Charge (SOC) data** to classify charging and discharging events, providing customers with actionable EV charging behavior insights within the platform.
- Created an **unsupervised EV battery health prediction model**, enabling proactive maintenance and minimizing downtime, which improved fleet reliability and reduced costs.
- Implemented a **Mahalanobis Distance-based Anomaly Detection model** for Vehicle Health Monitoring, identifying potential sub-component failures early in the platform's predictive maintenance capabilities.
- Developed **Python algorithm to detect fuel pilferage/theft in ICE vehicles**, reducing theft-related losses and providing a key differentiator for DeepView Analytics in cost-sensitive markets.

TATA CONSULTANCY SERVICES | SENIOR DATA SCIENTIST | MAY, 2010 – OCTOBER, 2018

- **Text Analytics on Net Promoter Score (Recommend & Improve) Comments by Customers:**
 - Developed a text classification model to analyse customer comments from Net Promoter Score (NPS) feedback, categorizing them into "Recommend" and "Improve" themes. Enabled efficient identification of customer sentiments and actionable insights. Built a dynamic dashboard in Excel, enabling stakeholders to interact with the data & view key trends.
- **Detecting Anomalous Machine Units Manufacturing Electrical Equipment:**
 - Developed a distribution-based solution using the **Multivariate Normal Distribution** to detect anomalies by analysing patterns between two selected device measures, identifying deviations in the production process.
- **Warranty Claims Prediction and Cost Prediction of Engines**
 - Developed a warranty predictor tool using the **Generalized Renewal Process** to estimate the expected number of warranty claims and associated costs over their warranty period, simplifying the complexity and improving prediction accuracy.
- **Root Cause Analysis of Accounts Receivable Disputes Data Using Text Mining Approach:**
 - Developed a text analytics model using **Regular Expressions, SVM, and Random Forest** to classify customer dispute notes into root cause categories such as invoice discrepancies and quantity mismatches.
- **Part Failure Prediction for Alternator in Refrigeration Cooling Units:**
 - Developed a **Machine Learning** models to predict alternator failures using historical operational, environmental, and maintenance data, enabling proactive servicing, reducing downtime, and improving overall equipment reliability.
- **GKPI Automation (Lean Six Sigma Green Belt Certification):**
 - Created an Excel Based automation template using Excel VBA.
 - **TAT Reduction: 92% & Accuracy Improvement: ~85%** (saved ~\$57,000, implemented across all MEAP countries).

Educational Qualification

- **M.Sc. Statistics** | 2008-2010 | University of Madras, Chennai
- **B.Sc. Statistics (Hons.)** | 2003-2006 | Calcutta University, Kolkata

Peer-Reviewed Publication

- Sabapathy, A., Biswas, A. (2023). **Road surface classification using accelerometer and speed data: evaluation of a convolutional neural network model**. Neural Computing and Applications 35 (19): 14183–14194.