Pragyan Jyoti Dutta

London, United Kingdom

🌐 pragyan-dutta.netlify.app | ⓞ github.com/cyber-prags | 🔗 linkedin.com/in/pragyan-jyoti-dutta
| ✈ pragyan.jyoti.dutta01@gmail.com | 📞 +44-7407747069

About Me

Proactive and detail-oriented **Data Scientist** with expertise in statistical modeling, machine learning, large-scale data analytics, and dashboard development (Tableau, PowerBI). Proficient in **Python, SQL**, and **distributed data frameworks (GCP, AWS, Azure)**, specializing in building optimized **ETL pipelines**, executing rigorous **A/B tests, and developing scalable data-driven solutions**. Skilled at collaborating cross-functionally to deliver end-to-end projects, translate complex data into actionable visual insights, and streamline workflows to enhance operational efficiency. Passionate about leveraging data science and machine learning to drive strategic decision-making, user growth, and global business impact in dynamic, innovation-driven environments.

Professional Experience

Global Sales Planning Analytics Intern — Jun 2024 - Jul 2025
Siemens Digital Industries Software — London, UK

- Spearheaded the design, development & deployment of critical Tableau dashboards, including:
  - **Consensus Dashboard:** monitored Product demo performances across various industries and portfolios.
  - **ICP Country Performance Dashboard:** provided insights into the country's planned sales plans vs actual performance for enhanced sales planning.
  - **Account Development Team Dashboard:** Monitored pipeline progress and lead closed won and closed lost conversion rates, facilitating data-driven decision-making.
- Led the **end-to-end project lifecycle**, collaborating with **Directors, VPs, and Portfolio Developers** to gather requirements, align with business objectives, and drive data-driven decision-making.
- Designed and deployed **optimized ETL pipelines** integrating data from **Salesforce, Qlik, and SharePoint**, ensuring seamless production deployment and global accessibility.
- Facilitated the Global **Integrated Country Planning (ICP) program**, enhancing sales planning for countries through strategic data insights.
- Designed a **Time Series Outlier Detection and Forecasting Model**, using **SARIMA** & **Bi-LSTM** to identify anomalies and improving sales projections for countries.

Data Engineering Intern — Aug 2022 - Feb 2023
Developer As a Service(DaaS) — Atlanta, US(remote)

- Architected an **end-to-end ETL pipeline** for processing over **21,000** textual image datasets using OCR and NLP techniques, **reducing processing time by 80%.**
- Optimized SQL queries and indexing strategies in MongoDB, improving database read/write **efficiency by 45%.**
- Built an **AI-powered case builder** that helped California lawyers draft legal cases efficiently, integrating NLP models and automation workflows to reduce manual workload.

Machine Learning Research Intern — Apr 2022 - Sept 2022
Spartificial — Dehradun, India

- Worked on a high-noise Pulsar Star dataset to optimize Pulsar Star Detection in imbalanced datasets by applying classification algorithms achieving a **model accuracy of 93.5%** and analyzed the impact of Dimensionality Reduction and Resampling application order on model accuracy.

Data Science Intern — Sept 2021 - Oct 2021
The Sparks Foundation — India

- Conducted in-depth **Exploratory Data Analysis** on diverse datasets, uncovering key insights to drive data-driven decision-making while mastering **Supervised Learning** and **Unsupervised Learning** techniques to solve complex problems and enhance analytical capabilities.

Education

University of Liverpool, United Kingdom — Sept 2023 - Aug 2025
MSc. in Data Science and Artificial Intelligence — GPA: 1st class(on course)

Indian Institute of Technology, Madras, India — Jan 2022 - Sept 2023
Online Diploma in Data Science — GPA: 7.75/10.0

Tezpur University, India — Oct 2020 - Jun 2023
BSc. in Physics — GPA: 8.06/10.0

PROJECTS

Multi Agent AI Hedge Fund
Personal Project

Mar 2025
Click to view

- Engineered an end-to-end solution deployed in Streamlit using **8 advanced LLM agents** powered by **Perplexity Sonar API** & **LangChain** that aggregates financial data, sentiment, technical signals, and risk metrics to drive data-backed investment decisions that reduced analysis turnaround time by **50%**, delivering comprehensive market insights in **under 1 minutes per ticker** with over **85% predictive accuracy**.

Route Optimisation for Starbucks Delivery in London
Personal Project

Nov 2024
Click to view

- Innovated and implemented a graph-based route optimization model using **Dijkstra's algorithm** to minimize delivery times and streamline logistics for Starbucks cafes in London. Optimized pathfinding to ensure timely deliveries, resulting in a **37% improvement** in delivery speed and a **20% reduction** in operational costs compared to unplanned deliveries.

Academic Research Paper Information Retrieval System
Personal Project

Oct 2024
Click to view

- Implemented a **Retrieval Augmented Generation (RAG) system** empowering researchers and students to upload and interact with academic papers, enhancing information accessibility and estimated to enhance productivity in research by 43%.

Uber Analytics ETL Project
Personal Project

Jan 2024
Click to view

- Leveraged **Google Cloud Storage** and **Big Query** to manage and analyze data.
- Established a robust ETL pipeline with **Mage Data Tool**, improving data processing speed by **40%**, enhancing overall workflow efficiency.
- Created interactive dashboards in **Google Looker Studio**, facilitating a significant improvement in strategic decision-making efficiency through clear data visualizations providing valuable insights into the working of Uber, which can be leveraged for better customer service and increasing revenue.

Churn Model Predictor
Personal Project

Dec 2023
Click to view

- Defined a sophisticated **Churn Model Predictor** using **Google BigQuery**, integrating **Google LookerStudio** for advanced dashboarding and analytics.
- Leveraged **AutoML** for efficient and accurate predictive modeling in forecasting employee turnover that aided Human Resources in strategizing employee retention, reducing churn rates across departments through data-powered recommendations.

Sentiment Prediction on Movie Reviews
MLP Project , IIT Madras

Jun 2023 - Aug 2023
Click to view

- Built an **NLP-based sentiment analysis** model using **Naive Bayes** and **LSTMs**, achieving **85% accuracy** in predicting movie sentiments.

TECHNICAL SKILLS

Programming Languages: Python, R, C++, Java, Scala	**Machine Learning:** scikit-learn, XGBoost, TensorFlow, PyTorch, Keras, statsmodels, Prophet, QuantLib, KDB+
Data Wrangling: NumPy, Pandas, SciPy, Apache Spark, Hadoop	**Databases & Warehousing:** SQL, PostgreSQL, MySQL, NoSQL (MongoDB, Cassandra), Snowflake
Data Visualization: Plotly, Tableau, Power BI	**Cloud & MLOps:** AWS, Azure, GCP, Kubernetes, MLflow

PUBLICATIONS

IEEE Xplore: Health Risk Detection through Web App using Machine Learning

April 2022
Publication Link

- Developed an AI based disease risk detection system using **Naive Bayes** and **Random Forests**, enabling **early diagnosis** through automated analysis of initial reports.

RELEVANT CERTIFICATIONS

Amazon Web Services Certified AI Practitioner
Amazon Web Services

Feb 2025
Certificate Link