Dwaipayan Roy

+1 (972) 897-3997 | dwaipayan.roy@utdallas.edu | www.linkedin.com/in/dwaipayan08 | https://github.com/RoyDwaipayan

EDUCATION

The University of Texas at Dallas, Richardson, Texas — Dec 2025
Master of Science, Business Analytics and Artificial Intelligence, Data Science Track — **GPA - 4.0**
Achievements: UT Dallas Dean's Excellence Scholarship
Relevant Coursework: Applied Machine Learning, Advanced Statistics, Large Language Models (LLM), Predictive Analytics, Prescriptive Analytics, Database Management, Natural Language Processing, Econometrics and Time Series Analysis

Jadavpur University, Kolkata, West Bengal — Jun 2020
Bachelor of Engineering, Mechanical Engineering — **GPA - 3.54**

SKILLS

Languages:	Python, SQL, R, Apex, Java, NoSQL
Tools:	Power BI, Tableau, Advanced Excel, Azure, GitHub, KNIME, Jira, Looker, Qlik Sense, BigQuery, Streamlit
ML & Analytics:	Machine Learning, Deep Learning, NLP, Time Series, Causal Inference, Prompt Engineering, Time Series Forecasting, LLM Fine Tuning, RAG, AI Agents, Data Analysis, Data Visualization, Data Modelling, Communication
Frameworks:	TensorFlow, PyTorch, Scikit-learn, statsmodels, XGBoost, Pandas, NumPy, PySpark, Openai, LangChain, Matplotlib
Cloud:	Microsoft Azure, Snowflake, Google Analytics
Certifications:	Udemy Certified Data Scientist, Data Dashboards in PowerBI, Google Analytics, KNIME Analytics L3, etc.

WORK EXPERIENCE

Fractal Analytics Pvt Ltd, Bengaluru, Karnataka — Jul 2022 – Aug 2024
Data Scientist

- Developed Causal Inference package using K-means with MDM and Double Machine Learning algorithms to assist a leading media entertainment client gauging the impact of Tournaments and media, showing 20% increase in ARPU and 35% increase in MAU.
- Developed automated pipelines for projects including cohort-based churn calculation, sales and engagement prediction, and KPI development using datasets of up to 3 billion records, assisting decision-making and contributing to a 14% revenue uplift.
- Developed an automated Fraud Detection pipeline using pattern analysis from telemetry data, identifying fraudulent players with 97% accuracy, ensuring fair play and saving 8% of revenue in gaming add-ons.
- Developed dynamic dashboards in MS Power BI and Qlik Sense for EDAs, KPIs and model outputs for stakeholders, and monitoring automation load, which saved $200k+ in resource utilization.
- Led client interactions, defined analytical needs and approaches, generated insights and recommendations from analysis, and communicated findings to non-technical audiences through comprehensive PPTs and executive-level storytelling.

PricewaterhouseCoopers LLP, Kolkata, West Bengal — Sept 2020 – Jul 2022
Full-stack Salesforce Developer

- Led development of full-stack Salesforce solutions for a leading client in Health Care to automate the routing and management of incoming emails, using Apex Object Oriented Programming, validation rules, etc. leading to a 65% reduction in manual labor.
- Predicted region wise vehicle financing demands to optimize marketing strategies and resource allocation, by integrating time series analysis using Einstein Analytics into Sales Cloud, aimed to reduce processing time by 25% during peak seasons.
- Led demos and training sessions for stakeholders and end-users, leading to improved adoption rates.

RELEVANT PROJECTS

Diabetes Prediction App and Recommendations using LLMs, UT Dallas | Link: https://tinyurl.com/2v3nr8n5

- Developed stacked ensembled models by training a Deep Neural Network meta-learner over XGBoost, LightGBM, Random Forest and Logistic Regression base models to predict risk of diabetes onset of an individual with 87% precision and 91% recall.
- Built a Streamlit-based UI integrated with Openai's Gpt-4o LLM for users to interact with the model and provide personalized recommendations on lifestyle changes based on user input and model output, to avoid diabetes in later life.
- Conducted a deep feature analysis to identify and interpret key factors correlating with the onset of diabetes.

Context extraction and context-based question answering with LLMs using RAG and AI Agents, UT Dallas

- Extracted specific date, jurisdiction, party and date information from large corpus of unstructured NDA documents and 10-K statements using meta-llama-70b and Openai's Gpt-4o LLMs with structured one-shot prompting.
- Developed a RAG pipeline to provide context to the LLM to answer specific questions regarding the 10-K statements.
- Implemented a ReAct chain (Reasoning + Action) using task specific agentic functions, enabling the LLM to perform advanced tasks such as complex calculations and web scraping.