






Pragyan Jyoti Dutta

 pragyan-portfolio.netlify.app |  github.com/cyber-prags |  linkedin.com/in/pragyan-jyoti-dutta
|  pragyan.jyoti.dutta01@gmail.com |  +44-7407747069

ABOUT ME

A creative results-oriented Data Scientist with expertise in advanced analytics, machine learning, and AI-driven solutions. Proficient in Python, Generative AI (GenAI) applications including multi-agent systems, and building end-to-end data pipelines using KNIME, Mage, Snowflake, and DBT on cloud platforms like Azure, GCP, and AWS. Skilled at uncovering actionable insights, crafting data-focussed strategies, and delivering impactful solutions that drive business growth. A collaborative problem-solver with a passion for innovation and staying ahead of industry trends, I thrive in transforming data into strategic assets for forward-thinking organizations.

EDUCATION

University of Liverpool, United Kingdom
MSc. in Data Science and Artificial Intelligence

Sept 2023 - Aug 2025
GPA: 1st class(on course)

Indian Institute of Technology, Madras, India
Online Diploma in Data Science

Jan 2022 - Sept 2023
GPA: 7.75/10.0

Tezpur University, India
BSc. in Physics

Oct 2020 - Jun 2023
GPA: 8.06/10.0

PROFESSIONAL EXPERIENCE

Global Sales Planning Analytics Support Intern
Siemens Digital Industries Software

Jun 2024 - Jul 2025
London, UK

- Facilitated the development and integration of Generative AI (GenAI) technologies into existing applications, contributing to the enhancement of the current technological ecosystem and overall user experience.
- Spearheaded data-driven insights to architect a comprehensive global sales strategy, leveraging Tableau, PowerBI, KNime and Salesforce to deliver interactive dashboards and predictive analytics tools that improved sales team efficiency by 10% and empowered country-specific teams to optimize sales growth over a projected three-year horizon.

Data Engineering Intern
Developer As a Service(DaaS) - Team

Aug 2022 - Feb 2023
Atlanta, US(remote)

- Analysed and processed over 21,000 textual image data using OCRs, demonstrating advanced problem-solving abilities and contributing to a significantly lesser processing times.
- Engineered data using Regex and other methods on unstructured pixelated legal documents in a collaborative and cross-functional team environment.
- Partnered with team members to create a synergistic full-stack ecosystem for advanced legal research services.
- Employed articulate communication and strategic leadership skills, orchestrating team efforts that resulted in an impressive 80% increase in efficiency for legal document processing within the firm.

Machine Learning Research Intern
Spartificial

Jun 2022 - Sept 2022
Online, India

- Training: Acquired in-depth knowledge in RNN, Neural Networks, TensorFlow, OpenCV, and Image Segmentation, augmenting my innovative capabilities in Machine Learning.
- Research: Optimized pulsar detection in imbalanced datasets using ML techniques, a critical and analytical contribution to the field.

PROJECTS

Route Optimisation for Starbucks Delivery in London
Personal Project

Aug 2024
[Click to view](#)

- Designed and implemented a graph-based route optimization model using Dijkstra's algorithm to minimize delivery times and streamline logistics for Starbucks cafes in London. Achieved efficient pathfinding to ensure timely delivery and reduced operational costs, which is estimated to increase the delivery times by 37% and save 20% costs compared to unplanned deliveries.

Academic Research Paper Information Retrieval System
Personal Project

Oct 2024
[Click to view](#)

- Developed a system enabling researchers and students to upload and interact with academic PDFs using NLP techniques powered by Google PaLM2, enhancing information accessibility and estimated to enhance productivity in research by 43%.
- Utilized Google Vertex AI for advanced text extraction and implemented a conversational interface to allow question-based information retrieval from uploaded PDFs.
- Built with Streamlit, allowing seamless deployment, and integrated with Google Cloud for efficient data handling and processing.

Uber Analytics ETL Project

Personal Project

Jan 2024

[Click to view](#)

- Leveraged Google Cloud Storage and Big Query to manage and analyze data, demonstrating deep expertise in cloud-based solutions.
- Built 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.
- The project provides 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](#)

- Developed 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.
- Significantly aided Human Resources in strategizing employee retention, effectively 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](#)

- Implemented a Natural Language Processing project that aimed at predicting the sentiment of a movie based on its review by various reviewers using Machine Learning models.

Online Grocery Store WebApp- GROCIFY



MAD-1 Project , IIT Madras

Jun 2023 - Aug 2023

[Click to view](#)

- Created "Grocify", a community-centered e-commerce grocery web app utilizing Flask, HTML, CSS, and Bootstrap, innovatively merging traditional grocery shopping with online convenience.
- Drove personalized shopping experiences and precise promotions through advanced analytics, elevating user satisfaction and product excellence.

Dashboard Projects

- [Tableau Dashboards](#) 
- [PowerBI projects](#) 

TECHNICAL SKILLS

Programming languages: Python,R ,C++, Java

ETL Tools: Matillion, Mage, Snowflake, Mage, Knime

ML/AI: AutoML, Tensorflow, Pytorch, Numpy, Pandas, Matplotlib,Plot-ly, Scikit-learn, Keras, Tableau, PowerBI, LookerStudio, Jupyter Notebooks, Google Collab

Cloud Technologies: GCP, AWS, Azure

Web Technologies: HTML, CSS, JavaScript, Node.js, React, Django

Miscellaneous: Salesforce, SQL, Git, Shell, Latex, MS-Office Suite, Figma,BigQuery, Bash

PUBLICATIONS

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

April 2022

[Publication link](#)

- The main motive of this paper was to help the general people who are unaware of available technologies and to facilitate them to use it through the web app present online which has all the features to detect the disease at an early stage using ML techniques of Naive Bayes and RandomForest.

RELEVANT CERTIFICATIONS

Matillion Data Productivity Cloud, Foundations

[Certificate Link](#)

Matillion Academy

- Obtained foundational knowledge, practical skills, and a functional understanding of Data Engineering and various ETL and ELT tools available in the market.
- Gained a foundational knowledge of building data pipelines in Matillion Data Productivity cloud.

Generative AI with Large Language Models

[Certificate Link](#)

DeepLearning.AI

- Obtained foundational knowledge, practical skills, and a functional understanding of generative AI, gaining insights into the latest research and how companies leverage cutting-edge technology for value creation.
- Benefited from expert instruction by AWS AI practitioners actively involved in building and deploying AI for real-world business applications.