

Pragyan Jyoti Dutta

pragyan-portfolio.netlify.app | github.com/cyber-prags | linkedin.com/in/pragyan-jyoti-dutta
| duttapragyanjyoti@gmail.com | [+44-7407747069](tel:+44-7407747069)

ABOUT ME

Driven Data Scientist with a knack for unearthing insights from complex data. Adept at wrangling and analyzing datasets using Python, industry-standard tools (KNime, Mage, Snowflake, DBT), and cloud platforms (Azure, GCP, AWS). Experienced in leveraging Generative AI (GenAI) technologies to enhance data-driven solutions and drive innovation. Proven ability to identify key trends, build impactful solutions, and translate data into actionable visuals. I thrive on collaboration and possess a passion for innovation, eager to contribute my skills to a company at the forefront of data-driven progress.

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
<ul style="list-style-type: none">Assisting in 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.Supporting the analysis of data and aiding in the formulation of a global sales strategy, helping to identify key market opportunities and optimize the company's sales initiatives.	
Data Engineering Intern Developer As a Service(DaaS) - Team	Aug 2022 - Feb 2023
<ul style="list-style-type: none">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
<ul style="list-style-type: none">Training: Acquired in-depth knowledge in RNN, Neural Networks, TensorFlow, OpenCV, and Image Segmentation, enhancing my innovative capabilities in Machine Learning.Research: Optimized pulsar detection in imbalanced datasets using ML techniques, a critical and analytical contribution to the field.	
Research Internship Society for Space Education and Research	Nov 2021 - Dec 2021 Online, India
<ul style="list-style-type: none">Investigated "Indirect methods to determine the fundamental properties of a Stellar-mass black hole," demonstrating insightful and analytical research capabilities and precise attention to detail.	
Data Science Intern The Sparks Foundation	Sept 2021 - Oct 2021 Online, India
<ul style="list-style-type: none">Performed strategic Exploratory Data Analysis on diverse datasets, showcasing my resourceful and critical thinking skills.Gained profound knowledge in both Supervised and Unsupervised Learning, reinforcing my adaptive and analytical abilities in Data Science.	
Summer Research Internship Naxxatra Sciences and Collaborative Research	Aug 2021 - Oct 2021 Online, India
<ul style="list-style-type: none">Led a project on "Analyzing the forest cover of India and land usage pattern of the last decade using Python," demonstrating visionary and strategic thinking. This work empowered stakeholders to understand forest cover changes and adapt policies accordingly.	

PROJECTS

Uber Analytics ETL Project

Jan 2024

Personal Project

[Click to view](#)

- Utilized 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

Dec 2023

Personal Project

[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-driven recommendations.

Sentiment Prediction on Movie Reviews

Jun 2023 - Aug 2023

MLP Project , IIT Madras

[Click to view](#)

- Worked on 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

Jun 2023 - Aug 2023

MAD-1 Project , IIT Madras

[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.

Other Projects

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

TECHNICAL SKILLS

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

ETL Tools: Matillion DPC, Mage, Snowflake, Mage

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.