Biswajit Paul

biswajitpaul0678@gmail.com | linkedin.com/in/biswajit-paul-b55792356 |
github.com/BiswajitPaul0678
Registration Number: B2430043 | M.Sc. Big Data Analytics, RKMVERI

Technical Skills

Languages: Python, R, SQL

Data Tools: Excel, Pandas, NumPy, Dask

Visualization: Matplotlib, Seaborn

ML/AI: Scikit-learn, TensorFlow, PyTorch, BERT

Concepts: Regression, Classification, Clustering

Other: Git, LaTeX, Jupyter, U-Net, Grad-CAM

Education & Certifications

- **M.Sc. in Big Data Analytics (Pursuing)** – RKMVERI, West Bengal
- **B.Sc. in Statistics** – Bidhannagar College, West Bengal
- **Higher Secondary** / **Secondary** – Jadavpur Vidyapith, Kolkata

Projects

- **Weather Prediction with Dask** – Engineered a distributed pipeline using **Dask** to analyze and forecast weather patterns across large datasets with parallelism.
- **Fake News Detection with BERT** – Fine-tuned **BERT** transformer on the **WeLFake** dataset; achieved high F1-score by leveraging contextual embeddings and dropout regularization.
- **ChestX-ray8 Medical Imaging** – Built a **CNN-based classifier** for multi-label disease prediction; combined **Grad-CAM, bounding boxes, and U-Net** segmentation for explainability.
- **Housing Price Regression** – Applied **multiple linear regression** and residual diagnostics; implemented **feature selection, cross-validation**, and model tuning.
- **Customer Segmentation via RFM** – Used **Recency-Frequency-Monetary** model in Python for behavioral clustering; visualized segments for targeting using **Seaborn** and KPI metrics.

Academic Interests

- Deep Learning, Explainable AI (XAI), Data-driven Health Analytics, Distributed Computing with Dask/Spark, Statistical Learning

Highlights

- Strong foundation in **statistics**, **machine learning**, and **distributed systems**
- Proficient with **end-to-end ML pipelines**, EDA, data preprocessing, model building and evaluation
- Hands-on experience with real-world datasets: **ChestX-ray8, WeLFake, GFS weather models**