Subrojyoti Paul

6003397131 | subrojyotipaul@gmail.com | linkedin.com/in/subrojyoti-paul | github.com/Subrojyoti

EDUCATION

Tezpur University — Tezpur, Assam
Master of Computer Application (MCA) — *2022 – 2025*
(CGPA: 9.01/10)

Darrang College — Tezpur, Assam
Bachelor of Science, Mathematics (Honours), Minor in Statistics — *2019 – 2022*
(CGPA: 8.55/10)

EXPERIENCE

AI Labeller — Mar 2025 – May 2025
Borde Inc.

- Annotated walnut images with high accuracy, categorizing them into respective classes to support **YOLO-based image classification models** for nut quality assessment.
- Monitored and analyzed model inference results by documenting false positives, false negatives, and other classification abnormalities to **identify model weaknesses**.
- Collaborated with cross-functional teams to **provide actionable feedback** for model retraining and enhancement based on annotation insights.

PROJECTS

Research Rover | *Python, Flask, React, Git* — Feb 2025 – Present

- Developed a full-stack web app for semantic research paper discovery and analysis.
- Integrated **CORE API** for large-scale research paper retrieval and **Gemini API** for advanced language understanding and Q&A.
- Implemented semantic search and context-aware recommendations using **SentenceTransformers, FAISS, and a Retrieval-Augmented Generation (RAG) pipeline**.
- Automated metadata extraction, batch PDF management, and CSV export with **real-time progress tracking**.
- Built with Python (Flask) serving as REST API, React, and Material-UI for scalable, user-friendly research workflows.

Diversified News Recommendation System | *Python, Flask, Git* — Aug 2024 – Dec 2024

- Developed a diversified news recommendation system in a team of 2 to address the **challenges of cold start and echo chamber** using the MIND dataset, which includes user behavior and news data
- Implemented three approaches: **(1) Collaborative filtering**, **(2) Content-Based filtering**, and **(3) A Hybrid model** combining the two approaches with a diversity re-ranking function to balance relevancy and diversity effectively.
- Achieved significant results with a balanced recommendation system, achieving **precision@10 of 74%, recall@10 of 88%, NDCG@10 of 34%, MRR of 19%, and ILS of 71%**, successfully solving both cold start and echo chamber challenges.

Fitbit Data Analysis | *R, Tableau* — May 2021 – Jun 2021

- Conducted comprehensive analysis of 30 Fitbit user data, **identifying trends** in physical activity, heart rate, and sleep patterns, which informed targeted marketing strategies and product development initiatives.
- Developed **interactive data visualization using Tableau** to present findings to stakeholders, resulting in an increase in engagement with data-driven insights and aiding decision-making processes.

TECHNICAL SKILLS

Languages: Python, C/C++, SQL
Databases: MySQL, MongoDB, Weaviate
Frameworks: React, Flask, FastAPI
Developer Tools: Git, Docker, Google Cloud Platform, VS Code
Libraries: Pandas, NumPy, Matplotlib, NLTK
Relevant Knowledge: Data Structures, Algorithms, NLP

ADDITIONAL COURSES AND CERTIFICATIONS

Courses: Machine Learning Algorithms (Supervised and Unsupervised Learning), Basics of Software Design
Certifications: Google Data Analytics, Python for Data Science