DISHA LAMBA

+1 (929) 444-1550 | jb.dishalamba@gmail.com | dldisha.github.io | GitHub | LinkedIn

PROFESSIONAL SUMMARY

Experienced Data Scientist with 3 years of experience building end-to-end machine learning pipelines and performing large-scale data analysis. Proficient in crafting success metrics, running power analyses, determining statistical significance, and presenting findings with clear recommendations. Skilled in synthesizing insights from data using Python, SQL, Tableau, and LLM techniques.

WORK EXPERIENCE

CVS Health   New York, NY
Data Scientist   Aug 2024 – Present

- Achieved **80% reduction in clinicians' patient-note review time** by architecting an **end-to-end LLM system** that extracts key insights from unstructured discharge notes for **1.2M-member population**, preserving critical medical information through **schema-guided prompting with Pydantic** and **ClinicalBERT-based hallucination checks** to ensure clinical accuracy.
- Developed LLM-based model that **analyzes claims data** to identify members at risk for specific conditions, collaborating closely with clinicians to validate detection logic, supporting a **projected $100M impact** through earlier targeting and interventions.
- Led a **CVS Speciality Pharmacy project**, benchmarking 36 pharmacies, analyzing data-sharing patterns **to guide CVS's strategy**.
- **Engineered 20M dataset** using **SQL** by unifying multiple sources into a structured repository, **streamlining team access to data**.
- **Built 20+ SQL-based utilisation metrics** to track user behaviour, engagement patterns, and efficiency trends, integrating them into **Tableau dashboards** to give stakeholders real-time visibility and support data-informed decisions.
- **Mentored 3 junior associates** on **ML best practices**, fostering technical growth and strengthening team capabilities.
- **Presented insights** to **C-level** and cross-functional stakeholders, enabling adoption of **data-driven decisions across workflows**.

Mindgrasp   Columbia, MD
Machine Learning Engineer   Jan 2024 – April 2024

- Designed a **math solver feature**, enabling students to solve complex queries within platform, **increasing user retention by 18%**
- Architected a scalable serverless solution using **AWS Lambda** for computation, **API Gateway** for secure communication, and **S3** for result caching, integrating with the **ReactJS frontend** to deliver real-time solutions **with minimal latency**.
- Ran **A/B experiments** with **Mixpanel** to **analyze user behaviour** and feature adoption, driving data-informed iterations.

SKILLS

Languages & Packages: Python, SQL, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Keras, TensorFlow, HTML/CSS, ReactJS
Machine Learning: NLP, Regression, Classification, Clustering, Statistical Analysis, Time-Series Analysis, Neural Networks
Database & Management: Data Warehousing, Data Modelling, MySQL, PostgreSQL, Redshift, BigQuery, Spark, Airflow
Cloud & Other Tools: Amazon Web Services, Google Cloud Services, Git, Tableau, Mixpanel, Jira, Confluence, MS Office Suite

PROJECTS

Platypus – Content Management System (Ruby on Rails, PostgreSQL) (**Blog**)

- Designed and implemented frontend and backend architecture of a Content Management System using Ruby on Rails, featuring WYSIWYG editing, drag-and-drop blocks, and hierarchical page structure with version control.
- Created responsive layouts to enhance user experience for content creation, task management, & publishing capabilities.

EDUCATION

New York University   New York, USA
Master of Science in Computer Engineering   May 2023
Coursework: Machine Learning, Deep Learning, Probability & Statistics, Big Data, Decision Optimisation Models, Data Analytics

Guru Gobind Singh Indraprastha University   Delhi, India
Bachelor of Technology in Information Technology   September 2020

CERTIFICATIONS & ACHIEVEMENTS

- **Speaker**, AICamp NYC 2024, Data Summit Boston 2025
- **IBM Data Science Professional Certificate**, Coursera – June'24
- **Best Medical Hack** at **MHacks 13 hackathon**; developed ReactJS app to match COVID-19 patients with plasma donors.