

DISHA LAMBA

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

PROFESSIONAL SUMMARY

Results-driven Data Scientist with 2 years of experience building end-to-end machine learning pipelines and performing large-scale data analysis. Proven impact across Healthcare, EdTech, and Consulting industries, delivering insights that inform strategy and drive product growth. Skilled at aligning data solutions with stakeholder needs through clear, collaborative communication.

WORK EXPERIENCE

CVS Health

New York, NY

Data Scientist

Aug 2024 – Present

- Built an LLM-driven note summarization system to extract key insights from unstructured discharge notes, preserving critical medical information, enhancing care manager efficiency and patient outreach across a 1.2M-member population.
- Led end-to-end preparation of a 27M-record healthcare dataset, transforming raw clinical encounters into a structured note repository for downstream ML and analytics, while ensuring HIPAA and FHIR compliance.
- Boosted summary reliability and accuracy by implementing schema-guided prompting using Pydantic and NLP techniques.
- Designed an automated summary validation framework using ClinicalBERT + GlinNER, ensuring hallucination detection.
- Designed Tableau dashboards visualizing summary usage, feedback, engagement metrics – enabling future development.
- Collaborated with engineers, clinicians, & product teams to integrate the model into workflow and streamline operations.

Mindgrasp

Columbia, MD

Full-Stack Machine Learning Engineer

Jan 2024 – April 2024

- Designed an LLM-powered feature enabling users to solve complex math queries, increasing user retention by 18%.
- Architected a scalable solution using AWS Lambda for backend computation, AWS API Gateway for secure communication, and AWS S3 for result caching, seamlessly integrating with a ReactJS frontend.
- Ran A/B experiments using Mixpanel and PostHog to analyze user behavior and deliver insights that shaped product roadmap and feature iteration.

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.
- Added task assignment, editorial workflow tracking, and role-based access to streamline collaboration and publishing.
- Created responsive layouts to enhance user experience for content creation, task management, & publishing capabilities.

EDUCATION

New York University

New York, NY

Master of Science in Computer Engineering

May 2023

Coursework: Machine Learning, Deep Learning, Statistics, Big Data, Decision Optimization Models, Data Analytics

Guru Gobind Singh Indraprastha University

Delhi, India

Bachelor of Technology in Information Technology

September 2020

Coursework: Data Structures, Computer Architecture, Cloud and Mobile Computing, Software Engineering, DBMS

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.