DISHA LAMBA

(929) 444-1550 | dl4747@nyu.edu | https://dldisha.github.io | GitHub | LinkedIn

EDUCATION

New York University, New York, NY **Sept 2021 - May 2023**
Master of Science in Computer Engineering GPA: 3.4/4
Coursework: Machine Learning, Deep Learning, Computer Vision, Big Data, Decision Models, Data Analytics, Network Security

Bharati Vidyapeeth's College of Engineering, New Delhi, IN **Aug 2016 - Sept 2020**
Bachelor of Technology in Information Technology GPA: 8.78/10
Coursework: Data Structures and Algorithms, Operating Systems, Software Engineering, Web Development, Database Management

TECHNICAL SKILLS

Languages	Python, R, C/C++, JavaScript, Git, Linux, HTML/CSS
Packages and Libraries	NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, PyTorch, TensorFlow, Keras, NLTK
Tools & Frameworks	PySpark**,** Flask, React.js, Ruby on Rails, Tableau, R Studio, MS Office suites**,** AWS
Databases	SQL, SAS, PostgreSQL, Hadoop, Apache Spark, Kafka

PROFESSIONAL EXPERIENCE

New York University New York, NY
Course Assistant Sept 2022 - May 2023
- Designated as Graduate Teaching Assistant for the **Deep Learning** Fall'22 and **Computer Networking** Spring'23 courses.
- Assisted more than 100 students with lectures, assignment queries via weekly office hours, in addition, to grading assignments.

Data Glacier New York, NY (Remote)
Data Science Intern June 2022 - Aug 2022
- Developed a **search engine** that ranks unstructured text documents based on semantic & contextual relationships to user prompts.
- Implemented scalable ETL pipelines using **SQL** and **Pandas** to analyze **1M** of **unstructured text data**.
- Proposed and built an ensemble of **BERT** and **TF-IDF**, capturing **80% of escalations** with a **K-S of over 48%**.
- Collaborated with the cloud team to optimize data storage and processing capabilities using **AWS S3** and **Lambda**.

Sapio Analytics Maharashtra, India (Remote)
Software Engineer - Machine Learning July 2020 - June 2021
- Led development of **job-search portal** for the **Govt. of India**, resulting in **1.2M** blue-collar job opportunities for Indian laborers.
- Implemented ETL pipeline using **SSIS** to automate data loading process; improved data quality & processing time by **17%**.
- Utilized **KNN** algorithm to match jobs based on user skillset and location, resulting in **50% increase in job placements**. By Integrating the machine learning model with **ElasticSearch**, achieved job-matching accuracy of **85%**.
- Developed robust **API** services with **Flask** & collaborated closely with the frontend team to integrate ML model with **React**.
- Devised **A/B experiments** with Product & Engineering teams to validate recommendations.

PROJECTS

Finance Portfolio Management System (Flask, SQL) Sept 2022 - Dec 2022
- Implemented statistical analysis capabilities within a Flask-powered financial portfolio management system, empowering users to perform in-depth analysis of their financial activities such as expense tracking, budgeting, goal setting.
- Deployed **PlaidAPI** to fetch real-time financial data, enabling users to view account balances, transaction histories, etc

Movie Recommendation System (Python, PySpark) Jan 2022 - May 2022
- Performed **exploratory and statistical data analysis** on the **27M** MovieLens dataset using **Python, and PySpark.**
- Built a predictive recommender system using **collaborative filtering** and **Apache Spark's ALS**. Through a comparative study with Python **LightFM** and **Annoy** libraries, fine-tuned its performance to provide top 5 movie recommendations.
- LightFM outperformed the ALS method with an **MSE** score of **0.77** and a **runtime** of **95.12 seconds**.

Plasma Desk web app (ExpressJs, HTML/CSS, Firebase, AWS - ElasticSearch, Heroku) Aug 2020
- Developed JS web app, driven by doctors, connecting eligible plasma donors with Covid patients, based on compatibility.
- The project received **Best Medical Hack** and **Wolfram Award for Top 30 Hacks** in the *MHacks 13 Beta Hackathon 2020*.

PUBLICATION

- Proposed an integrated system for Occupational Category Classification based on Resume and Job Matching that achieved an **accuracy of 83.5%**, published in *International Conference on Innovative Computing & Communication*, May 2020.

LEADERSHIP

- **Vice President** at **NYU Machine Learning Club** Jan 2023