DISHA LAMBA

(929) 444-1550 | dl4747@nyu.edu | Portfolio | GitHub | LinkedIn

EDUCATION

New York University, New York, NY **Sept 2021 - May 2023**
Master of Science in Computer Engineering
Coursework: Machine Learning, Deep Learning, Computer Vision, Big Data, Decision Optimization Models and Data Analytics

Bharati Vidyapeeth's College of Engineering, New Delhi, IN **Aug 2016 - Sept 2020**
Bachelor of Technology in Information Technology

TECHNICAL SKILLS

Languages and Software	Python, R, PySpark, SAS, Git, Linux, HTML/CSS
Packages and Libraries	Pandas, NumPy, Scikit-learn, Keras, Matplotlib, Seaborn, TensorFlow, NLTK
Machine Learning	Regression, Feature Selection, Clustering, Decision Trees, Dimensionality Reduction
Statistics and Analytics	Tableau, R Studio, MS Excel, Google Analytics
Databases and Cloud	MySQL, PostgreSQL, SSIS, Hadoop, Apache Spark, Kafka, AWS

PROFESSIONAL EXPERIENCE

Subconscious AI Remote
Research Intern June 2023 - Present
- Leveraging advanced language processing tools like **Wandb(Weights & Biases), OpenAI** to analyze and visualize data, collaboration on ML model training to enable Generative AI methods for casual experimentation.
- Incorporated **Hate-Speech Moderation** endpoint that follows OpenAI guidelines to check inappropriate experiment prompts.

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 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 Remote
Machine Learning Engineer 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 to improve data quality & processing time by **17%** using **SQL SSIS** to automate data loading process.
- 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%**.
- Devised **A/B experiments** with Product & Engineering teams to validate recommendations.

PROJECTS

Option Pricing Dashboard (MS Excel, Tableau) Jan 2023 - May 2023
- Developed interactive dashboard using **MS Excel**, **Tableau** to analyze data & visualize performance of **100+ option trades**.
- Leveraged **Monte Carlo Simulation** to model various market scenarios like risk assessment, and decision-making.

Movie Recommendation System (Python, PySpark) Jan 2022 - May 2022
- Performed **exploratory and statistical data analysis** on the **27M** MovieLens dataset using **Pandas**, and **PySpark.**
- Built a predictive recommender system using **collaborative filtering** and **ALS** method. Conducted comparative study with Python **LightFM** and **Annoy(ANN)**, fine-tuned the performance to provide top 3 movie recommendations.
- ANN outperformed the ALS method with an **MSE** score of **0.77** and a **runtime** of **3ms** per loop.

Plasma Desk web app (ExpressJs) 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.