

DISHA LAMBA

+1 (929) 444-1550 | dl4747@nyu.edu | dldisha.github.io/ | [GitHub](#) | [LinkedIn](#)

EDUCATION

New York University

Master of Science in Computer Engineering

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

NY, USA

May 2023

Guru Gobind Singh Indraprastha University

Bachelor of Technology in Information Technology

Coursework: Data Structures and Algorithms, Object-Oriented Programming, Software Design, Database Management Systems

Delhi, India

September 2020

SKILLS

Technical: Python, R, SQL, SAS, Postgres, Hadoop, Spark, Kafka, PowerBI, Tableau, Google Analytics, Advanced Excel (Pivot Table, VLOOKUP, Macros), SciPy, TensorFlow, NumPy, Pandas, Scikit-learn, Matplotlib, ggplot, HTML/CSS, ReactJS

Cloud Services and Tools: Amazon Web Services (AWS), Google Cloud Services (GCP), Terraform, Git, Linux, Docker

Others: Data Mining, Regression, Clustering, Statistical Modeling, Quantitative Analysis, Linear Programming, A/B Testing

WORK EXPERIENCE

Quantum Risk Analytics

Machine Learning Developer Intern

New York City, NY

Dec 2023 – Present

- Applied Logistic Regression to predict disease risk and global spread using hospitalization data, achieving **90%** accuracy.
- Generated insightful data visualizations using Tableau for presentation to senior leadership and epidemiologists.

Subconscious AI

Research Intern

New York City, NY

July 2023 – Present

- Optimized Casual Market Research product using Python, SQL, through feature engineering and data pre-processing methods, achieving **25%** performance boost and **30%** improvement in data quality.
- Implemented content filtering using Hate-Speech API, achieving **90%** moderation improvement, **saving 200+ hours**.
- Streamlined project timelines and resource utilization in a Git and Agile environment, enhancing the development lifecycle.

Data Glacier

Data Science Intern

New York City, NY

June 2022 – Aug 2022

- Developed semantic text-ranking search engine, enhancing user experience and **45%** improvement in operational efficiency.
- Implemented ETL with SQL, Pandas, transforming and cleaning unstructured data (~1M) to reduce analysis time by **30%**.
- Attained **85%** accuracy using Python-based **BERT and TF-IDF** ensemble model for effective natural language processing.
- Optimized data storage and processing capabilities with **AWS S3, and Lambda**, achieving **25%** data access improvement and cost reduction.

Sapio Analytics

Machine Learning Developer

Mumbai, India

July 2020 – June 2021

- Spearheaded development of job-search portal for the Government of India, generating **1.2M job opportunities** nationwide.
- Achieved **85%** accuracy in user-job matching using scikit-learn's K-nearest neighbor algo, tailored for user location, skillset.
- Collaborated with SDE team to seamlessly integrate ML model with ReactJS, enhancing user engagement by **25%**.
- Optimized test, build, and deployment processes using CI/CD and Docker, resulting in **30%** decrease in integration issues.
- Devised **A/B experiments** with Product and Engineering teams for data-driven decision-making and product testing.

PROJECTS

Big Data - Movie Recommendation System (Python, Spark, SQL) [[GitHub](#)]

- Utilized SQL, Pandas, and Matplotlib for EDA, extracting valuable insights from a large 27M MovieLens dataset.
- Built a baseline recommender system using **collaborative filtering ALS algorithm in PySpark**. Compared its performance against **Python LightFM, Annoy** models, achieving an accuracy of **92%** to deliver top 3 recommendations.

NLP - Integrated System for Occupational Category Classification of Resumes (Python) [[Publication](#)]

- Automated resume screening and categorization into occupational fields using **NLP-based parser** on 0.5M dataset.
- Achieved **84%** accuracy, co-authored the paper at the International Conference on Innovative Computing, May 2020.

ACHIEVEMENTS AND LEADERSHIP

- Best Medical Hack** at [MHacks 13 hackathon](#); developed ReactJS app to match COVID-19 patients with plasma donors.
- Conference attendee** at AWS re:Invent'23 (Scholar), Grace Hopper Celebration'23, Strange Loop'22 (Scholar).
- Former **Vice President**, NYU Machine Learning Club: Ignited machine learning interest in 30+ students.