# DISHA LAMBA

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

## EDUCATION

**New York University, NY, USA** Sept 2021 - May 2023
Master of Science in Computer Engineering
Relevant Coursework: Machine Learning, Deep Learning, Statistics, Big Data, Decision Optimization Models and Data Analytics
Academic Achievements: A+ in all Machine Learning courses, Tandon Summer Scholar'21, Academic Merit Scholarship
Course Assistant: Deep Learning, Computer Networking

**Guru Gobind Singh Indraprastha University, Delhi, India** Aug 2016 - Sept 2020
Bachelor of Technology in Information Technology

## TECHNICAL SKILLS

| | |
|---|---|
| **Programming and Libraries:** | Python, PySpark, NumPy, Pandas, Keras, Scikit-learn, Matplotlib, Seaborn, ReactJS, Flask |
| **Machine Learning:** | TensorFlow, NLTK, Regression, Classification, Clustering, Feature Selection, Decision Trees, PCA |
| **Databases and Analysis:** | SQL, Postgres, Hadoop, Apache Kafka, Tableau, PowerBI, MS Suite (Excel, PowerPoint) |
| **Cloud Services and Tools:** | Amazon Web Services, Weights and Biases (Wandb), Docker, Git, Linux, Jupyter |

## PROFESSIONAL EXPERIENCE

**Subconscious AI, New York City, NY** | Research Intern June 2023 - Present
- Transcribed **20+ Generative AI research papers** on human behavior, learning essential concepts for practical application.
- Utilized **Wandb, Python** to enhance MLOps, ensuring effective data visualization, analysis, and tracking of ML experiments.
- Automated content filtering with **Hate-Speech API**, resulting in **90%** improvement in moderation process and **saving 200+ hours**.

**Data Glacier, Remote** | Data Science Intern June 2022 - Aug 2022
- Developed **semantic text-ranking search engine**, leading to enhanced user experience and **45%** improvement in **overall efficiency**.
- Designed **ETL data pipelines** using **SQL, Pandas** for data cleaning, reducing analysis time by **30%** for **1M** unstructured dataset.
- Attained **K-S test score** of **48%** and captured **80%** of **escalations** by using **BERT** and **TF-IDF** ensemble model with **Python**.
- Optimized data storage & processing capabilities with **AWS S3, Lambda**, achieving **25%** data access improvement & cost reduction.

**Sapio Analytics, Mumbai, IN** | Machine Learning Engineer July 2020 – June 2021
- Spearheaded the development of **job-search portal** for the **Government of India**, generating **1.2M job opportunities** nationwide.
- Improved data quality and processing speed by **17%** using **SQL** Server Integration Services **(SSIS)** and automation.
- Achieved **85% accuracy** in user-job matching, boosting job placements by **50%** using **K-nearest neighbor** algorithm with **Python**.
- Collaborated with **SDE team** to seamlessly integrate ML model with **ReactJS**, enhancing user experience and engagement by **25%**.
- Devised **A/B experiments** with Product and Engineering teams for **data-driven decision-making** and **product testing**.

## PROJECTS AND PUBLICATIONS

**Machine Learning - Fake Review Detection System for Amazon Products (Python, SQL)** June 2023
- Conducted EDA with **Python, SQL** on 0.5M **Amazon Fine Foods dataset**, identifying patterns & anomalies to improve data quality.
- Built **Logistic Regression** predictive model, resulting in **90%** improvement in detection and elimination of fraudulent reviews.

**Deep Learning - Facial Editing System via Interactive Dialogue Interface (PyTorch, NLP)** Jan 2022 - May 2022
- Implemented facial editing system using 0.2M **CelebA-Dialog dataset** that allows users to get desired facial effects in natural language and interactively achieve their goals through dialogue interface, guided by **REINFORCE policy gradient algorithm**.
- Achieved **80% user satisfaction rate**, underscoring system's ability to interpret user needs and deliver precise facial editing results.

**Big Data - Movie Recommendation System (Python, PySpark, SQL)** Jan 2022 - May 2022
- Conducted in-depth **exploratory data analysis** on 27M-record **MovieLens dataset** using **PySpark**, **SQL, Pandas**, and **Seaborn**.
- Built a baseline predictive movie recommender system using **collaborative filtering ALS algorithm**. Compared baseline model performance against **Python LightFM** and **Annoy (ANN)** models, optimizing system to deliver top 3 movie recommendations.
- Achieved superior performance with the ANN model, with an **MSE score** of **0.77** and runtime of 3ms per loop.

**NLP - Integrated System for Occupational Category Classification of Resumes (Python, NLP) - Research Paper**
- Developed **NLP model** to automate resume screening, reducing recruitment time & categorizing candidates into occupational fields.
- Utilized **tokenization, segmentation techniques** to pre-process resume data, enabling essential skills extraction and matching.
- Achieved **accuracy rate of 83.5%**, co-authored the paper at the International Conference on Innovative Computing, May 2020.

## ACHIEVEMENTS AND LEADERSHIP
- Secured **3rd position** in Cybersecurity and Games Hack3D'21 competition.
- Won **Best Medical Hack** at MHacks 13 hackathon; created a JS web app to match COVID-19 patients with plasma donors.
- Former **Vice President**, **NYU Machine Learning Club**: Ignited machine learning interest in **30+ students**.
- **Conference attendee** at AWS re:Invent'23 (Scholar), Grace Hopper Celebration'23, Strange Loop'22 (Scholar).