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 Computer Engineering
Relevant Coursework*:* Machine Learning, Deep Learning, Computer Vision, Big Data, Decision Optimization Models & 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 & Libraries	Python, R, PySpark, NumPy, Pandas, Scikit-learn, Keras, Matplotlib, Seaborn, SAS, Git, Linux
Machine Learning	TensorFlow, NLTK, Regression, Feature Selection, Clustering, Decision Trees, Dimensionality Reduction
Databases and Tools	MySQL, PostgreSQL, Hadoop, AWS, Tableau, R Studio, MS Excel, Google Analytics

PROFESSIONAL EXPERIENCE

Subconscious AI, New York City | *Research Intern* June 2023 – Present
- Leveraging **Wandb** and **OpenAI** for data analysis; collaborating on ML model training to enable Generative AI experiments.
- Implemented **Hate-Speech Moderation** endpoint, successfully filtering out **95%** of inappropriate experiment prompts.

Data Glacier | *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, Mumbai | *Machine Learning Engineer* July 2020 – June 2021
- Collaborated in development of **job-search portal** for **Govt. of India**, aiding in **1.2 million** blue-collar job opportunities nationwide.
- Improved data quality & processing speed by **17%** by optimizing ETL pipeline with SQL Server Integration Services & automation.
- Boosted **job placements by 50%** and fine-tuned **KNN algorithm**, integrated with **ElasticSearch**, to achieve an **85% matching accuracy** based on user skillsets and location.
- Devised **A/B experiments** with Product and Engineering teams to validate data-driven recommendations.

PROJECTS AND PUBLICATIONS

Option Pricing Dashboard (MS Excel, Tableau) Jan 2023 - May 2023
- Created an interactive dashboard in MS Excel and Tableau, enabling real-time analysis and visualization for over **100 option trades**.
- Utilized **Monte Carlo Simulation** for comprehensive market risk assessment and strategic decision-making.

Dialogue-based Interactive Facial Editing System via Reinforcement Learning (PyTorch, NLP) Jan 2022 - May 2022
- Implemented a facial editing system that allows users to articulate desired facial modifications in natural language & interactively achieve their goals through a multimodal dialogue interface, guided by the **REINFORCE policy gradient algorithm**.
- Achieved **80% user satisfaction rate**, underscoring system's ability to interpret user needs & deliver precise facial editing results.

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

Integrated System for Occupational Category Classification based on Resume and Job Matching - 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 for accurate selection.
- Achieved an **accuracy rate of 83.5%**, co-authored the paper at the International Conference on Innovative Computing, May 2020.

ACHIEVEMENTS
- Secured **3rd position** in Cybersecurity and Games Hack3D'21 competition.
- Awarded with **Best Medical Hack** and **Wolfram Award for Top 30 Hacks** in the MHacks 13 Beta Hackathon,

LEADERSHIP AND ACTIVITIES
- Member of Rewriting The Code (RTC)
- Scheduled Attendee of Grace Hopper Celebration'23 (GHC)
- Former **Vice President**, **NYU Machine Learning Club**: Ignited passion for machine learning in over **30 students**.
- Fully Funded Attendee, Strange Loop'22 Conference (STL)