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, Statistics, 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, React.js, Flask, PySpark, NumPy, Pandas, Scikit-learn, Keras, Matplotlib
Machine Learning:	PyTorch, NLTK, Feature Selection, Regression, Clustering, Time Series Analysis, Decision Trees
Databases and Analysis:	MySQL, Hadoop, SAS, Splunk, Tableau, PowerBI, MS Suite (Excel, Powerpoint)
Cloud Services & Tools:	Airflow, Amazon Web Services (S3, Lambda, ElasticSearch), Docker, Git, Linux
Quantitative Analysis:	Risk Management, Portfolio Optimization, Statistical Analysis, Deterministic and Probabilistic Models

PROFESSIONAL EXPERIENCE

Subconscious AI, New York City, NY | *Research Intern* June 2023 - Present
- Utilized **WandB, Python** for data visualization and analysis, to track ML experiments in Generative AI studies on human behavior.
- Achieved **95%** effectiveness in filtering inappropriate experiment prompts using OpenAI **Hate-Speech API** & **pytest** for validation.

Data Glacier, Remote | *Data Science Intern* June 2022 - Aug 2022
- Developed a **search engine** that ranks unstructured text documents based on semantic and contextual relationships to user prompts.
- Designed **ETL pipelines using SQL** and **Pandas** to analyze a large dataset of **1M** unstructured text data.
- Attained **48% K-S** and captured **80% of escalations** by using **BERT** and **TF-IDF** ensemble model.
- Led collaborative efforts to enhance data storage and processing capabilities via **AWS S3** and **Lambda**.

Sapio Analytics, Mumbai, IN | *Machine Learning Engineer* July 2020 – June 2021
- Generated **1.2M job opportunities** nationwide by spearheading the development of **job-search portal** for the **Government of India**.
- 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%** by using **K-nearest neighbor algorithm**.
- Integrated ML model with **React.js**, resulting in **25% increase in web user engagement** and insights into behavioral analytics.
- Devised **A/B experiments** with Product and Engineering teams for **data-driven decision-making** and **product testing**.

PROJECTS AND PUBLICATIONS

Portfolio Optimization Suite (MS Excel, Tableau) Jan 2023 - May 2023
- Utilized **MS Excel Solver** to refine deterministic model for portfolio allocation, maximizing potential returns while minimizing risks.
- Enhanced projected return on investments by 15% using **Monte Carlo Simulation**, offering insights into optimal asset diversification.
- Created **Tableau dashboards** to visualize asset allocation, return rate predictions, and **KPI metrics** for informed investment decisions.

Movie Recommendation System (PySpark) Jan 2022 - May 2022
- Conducted in-depth **exploratory** and **statistical analysis** on a 27M-record MovieLens dataset, using **Pandas** and **PySpark**.
- Built a predictive movie recommender system using **collaborative filtering** and ALS algorithm. Compared performance against **Python LightFM and Annoy (ANN)** models, optimizing the 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.

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**: Catalyzed machine learning interest in **30+ students**.
- **Attendee** at Grace Hopper Celebration'23 (GHC), Strange Loop'22 Conference (Funded).
- **Active Member**: Girls Who Code (GWC), Rewriting The Code (RTC), Society of Women Engineers (SWE)