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 Languages:	Python, R, NumPy, Pandas, Keras, Scikit-learn, Matplotlib, Seaborn, ReactJS, Flask, HTML, CSS
Machine Learning:	TensorFlow, NLTK, Regression, Classification, Clustering, Decision Trees, PCA, Neural Networks
Quantitative Analysis:	Risk Management, Forecasting, Statistical Analysis, Linear Programming, Optimization Models
Databases and Analytics:	SQL, Postgres, Hadoop, Spark, Kafka, Tableau, PowerBI, MS Suite (Excel, PowerPoint)
Cloud Services and Tools:	Amazon Web Services (AWS), ElasticSearch, Weights and Biases (Wandb), Git, Linux, Jupyter

PROFESSIONAL EXPERIENCE

Subconscious AI **July 2023 -Present**
Research Intern New York City, New York
- Collaborated with cross-functional teams to optimize **Python-based machine learning model** for market research, achieving **25%** performance increase and **30%** improvement in data quality through data pre-processing and feature engineering.
- Enhanced algorithm efficiency by **20%** through testing, contributing to more effective and ethical market research solutions.
- Automated content filtering with **Hate-Speech API**, resulting in **90%** improvement in moderation process and **saving 200+ hours**.

Data Glacier **June 2022 - Aug 2022**
Data Science Intern Remote
- 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 **80% escalation** detection accuracy with a streamlined **BERT** and **TF-IDF** ensemble model in **Python**.
- Optimized data storage & processing capabilities with **AWS S3, Lambda**, achieving **25%** data access improvement & cost reduction.

Sapio Analytics **July 2020 - June 2021**
Machine Learning Engineer Mumbai, India
- 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 with **K-nearest neighbor** algorithm in **Python**, tailored for user location and skillset.
- 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

Data Analytics - Portfolio Optimization (MS Excel, Tableau)
- Revamped **deterministic model** with **MS Excel Solver** for portfolio optimization, resulting in **18% decrease in projected risk**.
- Developed **Tableau dashboards** showcasing portfolio performance, **risk metrics, KPIs** for informed investment decision-making.

Machine Learning - Fake Review Detection System for Amazon Products (Python, SQL)
- 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.

Big Data - Movie Recommendation System (Python, PySpark, SQL)
- Conducted in-depth **exploratory data analysis** on 27M **MovieLens dataset** using **SQL, Pandas**, and **Seaborn**.
- Built a baseline recommender system using **collaborative filtering ALS algorithm in PySpark**. 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** on 0.4M **resume dataset** to automate resume screening and categorizing candidates into occupational fields.
- Achieved **accuracy rate of 83.5%**, co-authored the paper at the **International Conference on Innovative Computing**, May 2020.

ACHIEVEMENTS AND LEADERSHIP

- 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).