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, PySpark, NumPy, Pandas, Keras, Scikit-learn, Matplotlib, ReactJS, Flask
Machine Learning:	TensorFlow, NLTK, Regression, Clustering, Feature Selection, Time Series Analysis, Decision Trees
Quantitative Analysis:	Risk Management, Forecasting, Statistical Analysis, Linear Programming, Optimization Models
Databases and Analysis:	SQL, Hadoop, Postgres, SAS, Apache Kafka, Tableau, PowerBI, MS suite (Excel, PowerPoint)
Cloud Services & Tools:	Amazon Web Services (S3, Lambda, RDS), Azure, ElasticSearch, Git, Jupyter, Figma

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.
- Automated content filtering with **Hate-Speech API**, resulting in **90%** improvement in moderation processes and **saving 200+ hours**.
- Collaborated with cross-functional teams to optimize data pipelines using **Azure Data Factory** and **Azure Databricks**, resulting in **30%** improvement in **data processing** efficiency, **20% cost reduction**, and minimizing production delays.

Data Glacier, Remote | Data Science Intern June 2022 - Aug 2022
- Developed a **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.
- 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%** by using **K-nearest neighbor (KNN)** algorithm.
- Collaborated with SDE team to seamlessly integrate ML model with **ReactJS**, enhancing user experience and engagement
- Devised **A/B experiments** with Product and Engineering teams for **data-driven decision-making** and **product testing**.

PROJECTS AND PUBLICATIONS

Optimization and Analytics - Supply Chain and Scheduling Optimization (MS Excel, Tableau) Jan 2023 - May 2023
- Revamped **deterministic model** in **MS Excel Solver** to streamline supply chain, minimizing lead times, and inventory overheads.
- Reduced projected inventory costs by **18%** using **Monte Carlo Simulation**, providing a balanced demand-supply relationship.
- Created **Tableau dashboards** highlighting supply chain optimizations, reduced lead times, and **KPI tracking** for decision-making.

Big Data - 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.

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**.
- **Attendee** at Grace Hopper Celebration'23 (GHC), Strange Loop'22 Conference (Scholar).