UTKARSH MATHUR

(425)-633-5220 • utkarsh.mathur@gmail.com • https://datamathur.github.io/ • https://www.linkedin.com/in/utkarshmathur1024

EDUCATION

MS: Engineering Science Data Science, University at Buffalo (SUNY) **01/2023 – 06/2024**
- Relevant Courses: Data Models Query Languages, Statistical Learning (2 Semesters), Introduction to Machine Learning, Algorithms Analysis and Design

B.Tech.: Polymer Science and Engineering, Indian Institute of Technology (IIT) Roorkee **07/2018 – 05/2022**
- Relevant Courses: Artificial Neural Networks, Database Management Systems, Mathematics (3 Semesters)

RELEVANT SKILLS

- **Programming Languages**: Python, SQL, R, MATLAB, C, C++, Perl, Java, JavaScript, Hadoop
- **Tools & Frameworks**: PyTorch, TensorFlow, Keras, Tableau, LangChain, Hive, Kafka, PySpark, Kubernetes
- **Skills & Knowledge**: Object Oriented Programming, Relational Databases, Distributed Systems, Machine Learning, Deep Learning, Data Structures and Algorithms, ETL Pipelines, CI/CD Pipelines, MapReduce Systems

WORK EXPERIENCE

Data Scientist, Quinbay, Bengaluru: **05/2022 – 10/2022**
- Deployed quantized version of pre-trained Object Detection model to process order returning requests 40% faster.
- Deployed features to identify counterfeit products of recognized brands which required analysis of product images and product details in English and Bahasa. The feature reduced human efforts to identify counterfeit products by 30%.

Machine Learning Intern, Hono, India: **07/2021 – 04/2022**
- Created automated pipelines for Machine Learning model training by leveraging Data Analytics, ETL pipelines, Machine Learning algorithms, and Time Series Analysis to add personalized features for clients.
- Developed employee attrition forecasting, promotion suggestion, hike estimation, attendance forecasting, and organization hierarchy features.

Data Science Intern, ImagoAI, India: **04/2021 – 05/2021**
- Built AI models for 4 Food Quality assessment services by training ML models on Hyperspectral Images.

ACADEMIC RESEARCH

Semantic Segmentations of Ocular Images: Dr. Mayank Goswami, IIT Roorkee
- Experimented on Deep Learning algorithms for semantic segmentation over tumor-infected ocular images of rats to estimate tumor growth and trends.
- Trained multiple versions of U-Net architecture based Deep Learning models over manually annotated OCT B-Scans.

Feature Selection with Grey Wolf Optimization: Dr. Kusum Deep, IIT Roorkee
- Developed programs for feature selection optimization techniques using Grey Wolf Optimization (GWO) techniques.
- Used datasets from multiple domains to test the optimized feature selection techniques which were based on Random Walk Grey Wolf Optimization (RW-GWO) technique.

ACADEMIC PROJECTS

Metaheuristic Optimization vs Backpropagation: Python (Computer Vision and Image Processing, UB)
- Created PyTorch model optimizers based on Particle Swarm Optimization (PSO) and Grey Wolf Optimization (GWO).
- Compared and analyzed the performances of PSO, GWO, and SGD (Stochastic Gradient Descent) in training computer vision models with CNN architectures to understand the importance of backpropagation.

Statistical Analysis Analysis: R (Statistical Learning and Data Mining I, UB)
- Demonstrated the role of dimensionality and volume of data in regression tasks by comparing 7 regression models trained with various optimization and regularization techniques on 2 different datasets.
- Compared performances 7 classification techniques in a binary classification task and analyzed the impact of each feature of the dataset in explanation of variance of target variable in each model to better understand the techniques.

Clustering and Time Series Analysis: R (Statistical Learning and Data Mining II, UB)
- Performed clustering analysis on Land Mine dataset to derive insights that might help in multiclass classification.
- Performed time series forecasting on Oil Sales data using Exponential Smoothing and ARIMA models and compared the results (RMSE score) of 100 days forecasts.

Top Spotify Tracks Database: Python, SQL, JavaScript (Data Model Query Language, UB)
- Implemented database system to store daily nationwide and worldwide Top 50 tracks metadata from Spotify Web API.
- Created a Web App (MERN Stack) dashboard and song recommender using Spotify Annoy from the database.