DEEPAK UDAYAKUMAR

Boston, Massachusetts | +1 (857)230-2532 | deepakudayakumar07@gmail.com | LinkedIn | Portfolio | GitHub

EDUCATION

Northeastern University, *Boston, Massachusetts*   January 2023 – Present
Master of Science in Data Science (GPA: 4.0/4.0)   Expected Graduation: December 2024
Courses: Supervised Machine Learning, Data Mining, Data Processing and Management (in R), Deep Learning, Algorithms.
Nitte Meenakshi Institute of Technology, *Bangalore, India*   August 2017 - July 2021
Bachelor of Engineering in Electronics and Communication Engineering (GPA: 8.92/10)
Related Courses: Machine Learning, Linear Algebra, Pattern Recognition, Python Programming, Data Structures, Image Processing (using MATLAB).

SKILLS

- **Programming Languages**: **Python**, R, **SQL**, Java, MATLAB
- **Machine Learning and Statistical Techniques**: Regression, Classification, Ensemble Learning, Clustering, **Deep Learning**, Natural Language Processing, Generative AI, **Computer Vision, Image Analysis and Segmentation**.
- **Databases, Big Data, and Cloud Platforms**: **MySQL**, MongoDB, Microsoft SQL Server (**MS SQL**), Hadoop, **Spark**, Data Pipeline Development using **Airflow**, **Google Cloud Platform (GCP)**, AWS (Amazon EC2, Amazon Redshift).
- **Toolkits and Libraries**: **NumPy**, Pandas, Scikit-learn, **SciPy**, Keras, **TensorFlow**, Pytorch, NLTK, **OpenCV**, **Matplotlib**, Plotly, Selenium**, Docker**, Tableau, **PowerBI**, Excel, **Git**, Jira.
- **Soft Skills:** Intellectual Curiosity**,** Strong Communication Skills, Collaboration and Teamwork, Adaptability and Learning Agility, Problem-Solving Skills, Customer-Centric Mindset.

WORK EXPERIENCE

Data Engineer   August 2021 - December 2022
Mindtree Ltd, *Bangalore, India*
- Leveraged Python and Apache **Airflow** to architect and configured **ETL** data pipelines for Data manipulation, enabling **large-scale data analysis** through the daily ingestion and transformation of approximately **1,800 data files** into **Big Query**.
- Implemented strategic optimizations in the **data pipeline**, resolving bottlenecks and boosting **productivity by 35%**, reducing data latency, and lowering operational costs.
- Employed various **Google Cloud's** services like Big Query, Cloud Dataproc, and Cloud Storage to manage and track workflows for simultaneous processes in the cloud environment.
- Collaborated with clients to develop **100+ Big Query Views** to meet specific business requirements, improving data accuracy by 25% through data cleaning and transformations using **SQL**.
- Exhibited strong communication skills, fostering collaboration within cross-functional teams in a multi-cultural environment while maintaining a keen focus on detail and timelines.

ACADEMIC PROJECTS

Street Sense: Real-Time Object Detection on Urban Roads | Computer Vision / Model Optimization / YOLO
- Implemented an **object detection model** to detect and track vehicles, pedestrians, and other objects in busy streets, thereby assisting in the **optimization of traffic flow**, improvement of road safety, and enhancement of overall urban planning.
- Improved the model accuracy by filtering low-score detections and managing overlaps, yielding a more precise output.

Physical Activity Prediction and Real-time Deployment on AWS EC2| Logistic Regression/SVM/XGBoost/AWS
- Leveraged machine learning algorithms and deep learning (Neural Networks) to predict physical activity.
- Deployed the best-performing model (Random Forest) on **Amazon EC2 instance**, **implemented APIs** for real-time predictions, ensuring a robust and scalable solution for physical activity monitoring.

Facial Emotion Recognition using Deep Learning and Residual Networks | CNN/ResNet/ Image Analysis /TensorFlow
- Developed a deep learning model with **CNN** (Convolutional Neural Network) and **ResNet** to predict facial key points.
- Preprocessed and augmented image data trained it using **TensorFlow** and **Keras.** This project demonstrates proficiency in deep learning, computer vision, and **data analysis**, focusing on emotion detection and customer satisfaction analysis.

US Air Pollution Time Series Analysis | Statistical Analysis/Neural Networks/Time Series/SARIMAX/LSTM/Naïve
- Conducted extensive **inferential statistical analysis** on the US pollution dataset, focusing on significant pollutants.
- Applied **SARIMAX** and **LSTM** for forecasting post time-series decomposition and used **data visualization** to effectively illustrate California's future air pollution trends, thus providing strategic insights for environmental planning.

ACCOMPLISHMENTS

- **Publication: Graph Neural Network for SARS-CoV-2 Drug Repurposing**
- Led the University Cricket Team as captain, securing multiple victories.
- Ranked in the top 10% of my undergraduate class.