

Paras Varshney

Boston, MA • (857) 395 8732 • varshney.p@northeastern.edu

LinkedIn/blurred-machine

GitHub/blurred-machine

parasvarshney.ml

EDUCATION

Northeastern University, Boston, MA

Candidate for Master of Science (M.S.) in **Data Analytics Engineering** (GPA: 3.86)

Apr. 2024

Indian Institute of Information Technology, Kurnool, AP

Bachelor of Technology (B.Tech.) in **Computer Engineering** (GPA: 4.0)

Jun. 2020

WORK EXPERIENCE

LogicAI (KaggleDays), Warsaw, Poland

Data Scientist

Sep. 2021 – Aug. 2022

- Orchestrated initiation of "KaggleDays x Z by HP Global Championship", resulting in **13** successful data science meetups world wide building a community of **12 million** data scientists competing for the title. This resulted in improved participation by **80%**, titled **top Kaggle community globally**.
- Spearheaded automation of cutting-edge data processing pipelines, increasing efficiency by **1.4x** on **Kaggle**. Mentored event back-end, resulting in **95%** participant satisfaction catering to top grandmasters'.

India Urban Data Exchange (IISc. Bangalore), Bangalore, India

Data Scientist

Sep. 2020 – Sep. 2021

- Constructed an end-to-end **open-source** ITMS-focused data ingestion pipeline using **Apache Flink** and **Spark** to pre-process data before warehousing. The pipeline processed **900k** data points per minute from **5** major smart cities and communicated insights through an analytical dashboard leveraging Superset for effective decision-making. Also developed a **PyIUDX SDK** in Python with **CLI** and **GUI** Interactive capabilities.
- Predicted public transportation Estimated Time of Arrival (ETA) with deep neural network models, achieving an **R2 score** of **0.93** on unstructured data. The pipeline was optimized for real-time processing of **high-throughput** messages.

RESEARCH & PROJECTS

DeepFace & FaceNet ensemble Smart AI Attendance Engine

Feb. 2020 – Jun. 2020

- Designed a high-accuracy Smart AI Attendance Engine employing an ensemble of **DeepFace** and **FaceNet** for facial recognition and authentication, succeeding with a **97.5%** recognition rate on a dataset of **5000+** students.
- Implemented **gaze point estimation** for attendance tracking using PyTorch with OpenCV and Flask, adhering to data science best practices and facial data **privacy measures**. Applied **FaaS** principles for deployment.

Bayesian Statistical Approaches to develop Dynamic Pricing Models

Oct. 2019 – Dec. 2019

- Deploying Bayesian statistics, improved revenue generation by **17%**. Conducted churn analysis and market segmentation using **RFM scoring** and **Uplift Modeling** based on Bayesian Networks.
- Automated pricing model through feature engineering, statistical inferences, and **probabilistic modeling**, including **A/B testing**. Utilized Python3 and PyCharm for development of the project.

Video frame Analysis with edge estimation for Road Congestion Control

Jul. 2018 – Sep. 2018

- Developed a Smart AI Traffic Light System using CCTV to monitor **road congestion** by analyzing **video frames** and **semantic segmentation** based road edging. Switched streetlight signals based on **road lane density statistics**, while logging emergency vehicles prioritization. Employed Python3, Raspberry Pi, and OpenCV.

SKILLS

Languages: Python3, R, C/C++, JAVA, SQL, Rust

Frameworks: Scipy, Scikit, PyTorch, Keras, NLTK, Machine Learning, Deep Learning, Tensorflow

Tools: Jupyter, Pycharm, VSC, Git, Tableau, Superset, H2O, MongoDB, Hadoop, Flask, Spark

Soft Skills: Public Speaking, Accountable and Responsible, Time Management, Collaborative, Goal driven

ACTIVITIES & LEADERSHIPS

- VP of Projects & Relations** at Data Science Hub, Northeastern University - 2022.
- Recognized as one among **16 Global Data Science Ambassadors** with Z by HP & NVIDIA, 2020 - Present.
- Speaker**, NVIDIA GTC Global AI Conference, April 2021, **Data Science Mentor**, Hack4She global Hackathon.
- President**, Cultural Affairs Society, **Coordinator**, Co'digo Club. Organized techno-cultural fest SOLASTA-19.
- Attained **Kaggle Master** tier with a global rank under 63 out of 194K active Data Scientists (LB - Top 1%).
- Interests: Lead Guitarist (IIIT Band), Speed Cuber (30s), Table Tennis (CBSE clusters, inter-IIIT).