Paras Varshney

Data Scientist
Boston, MA | linkedin.com/in/blurred-machine | (857) 395 8732 | varshney.p@northeastern.edu
parasvarshney.com | github.com/blurred-machine

EDUCATION

Northeastern University, Boston, MA 2022 - 2024

Master of Science** in Data Analytics Engineering - **GPA: 3.95/4.0
Computer Science, Algorithms and Data Structures, Data Visualization, Advanced Machine Learning, Deep Learning, and NLP.
Responsibilities: **President** at the **Data Science Hub**, managing vibrant AI communities with **1500+** active members.

SKILLS

- **Languages**: Python, SQL, Java, R, Rust, Git, C, Matlab, MySQL, NoSQL, Spark, PySpark, Linux, and bash/shell scripting
- **Applied CS**: Data Science, ML, Deep Learning, Computer Vision, Natural Language Processing, Deep Learning,and Big Data
- **Tools and Frameworks**: Pandas, NumPy, scikit-learn, Keras, PyTorch, Tensorflow, OpenCV, Matplotlib, Seaborn, Docker, Kubernetes, git, SAS, Microsoft Excel, GPU, CUDA, Agile, Tableau, regression, clustering, Cassandra, AWS, GCP, and GitHub
- **Certifications**: Machine Learning Specialization by Stanford, Natural Language Processing Specialization by deeplearning.io, Data Scientist with Python Track at DataCamp #115,033

EXPERIENCE

Data Scientist (2.5+ years): LogicAI (Warsaw, PL) and IUDX, IISc.B Apr 2020 – Sep 2022

- Created a dynamic, stateless retraining strategy that makes models more flexible in **transportation datasets** and improved the performance of the **MLOps** architecture by building a strong **data warehouse catalog**. (Apache Airflow, cross-functional, data processing, API, statistics, communication skills, cloud computing, KPI analysis, Google Cloud, Lambda, ETL, MLOps)
- Developed workflow for an **open-source ITMS forecasting data ingestion** using version control, Kafka and Spark.
- Collaborate with backend engineers to build an ML platform for training and inference with CI/CD pipelines. Optimized **transport data governance** with Databricks **reduces processing time by 20%**. Improved MLOps efficiency by **15%** and **accelerated the ITMS forecasting pipeline by 1.5x**, handling **1.2 million data points** per minute. (Snowflake, JSON, elasticsearch, data-driven, data analysis, predictive models, decision trees, neural networks, Salesforce)

ML Research Assistant: Ocean Acoustic Sound Engineering Lab (Boston, MA) Feb 2023 – Present

- Implemented a state-of-the-art machine learning and data mining solution for **underwater mammal sound** analysis, achieving a **97.8% accuracy rate**. Processed predictive analytics for **200 million samples** across **4k–500k frequency bands**.
- Engineered a **geospatial** network **localization map**, revolutionizing real-time underwater tracking systems. Integrated **signal processing** and **spectrogram analysis** for acoustic signals, collaborating closely with **MIT researchers** for **data acquisition**.

Data Science Intern: Beyond Limits (Los Angeles, CA) Jun 2023 – Aug 2023

- Improved invoice data pipeline by researching **Form Recognizers** resulted in a **71% reduction in processing time** with real-time code suggestions and structured **batch mode ML** model implementation. (Hadoop, documentation)
- Reduced prediction errors by **0.6x** using **hierarchical classification** programming strategies and **drift analysis**, leveraging **AWS** EC2 and S3 **infrastructure**. Scheduling, executing, and monitoring data modeling pipelines to ensure robustness.

PROJECTS

Formula-1 Driver's Behavioral Modeling for Mind Wandering State Apr 2023 – Dec 2023

(TabTransformer, RNN, LSTM, A/B testing, embedding development, Artificial Intelligence, bioinformatics, Graph Database)

- Conducted **Master's thesis research**, improving driver safety by **35%** using advanced **physiological sensors** and **transformer models**, showcasing expertise in end-to-end networking complex interaction analysis. Employed **Hugging Face, TensorFlow** and **Azure** to analyze **EEG signals** for **drowsiness detection**, utilizing **BERT** and **RoBERTa** on unstructured time series data. Implemented a user-friendly data platform with **Docker** and **Kubernetes** for **containerization** and **orchestration**.

EXTRA-CURRICULAR ACTIVITIES

- **Global Data Science Ambassador, HP (2020 – 2023):** Recognized as one of the 16 Z by HP ambassadors worldwide, contributing to the data science BIPOC community through global grant program management and mentorship.
- **Interests**: Table Tennis (state finalist), Rubiks Cuber (under 30 seconds), and Hiking Himalayan Ranges.