# HIMANSHU PAITHANE

New York, NY | P: 201-241-9114 | hpaithan@stevens.edu | LinkedIn | github

## EDUCATION

**STEVENS INSTITUTE OF TECHNOLOGY**
Masters of Science
Major in Computer Science (GPA: 3.9/4.0)

## WORK EXPERIENCE

**PANASONIC NORTH AMERICA** May 2025 – Present
**Data Engineer**
- Developed and deployed a Databricks (PySpark) pipeline to transform and standardize **300,000+ customer names** using **NLP** + **fuzzy matching**, ranking candidate matches to improve downstream analytics reliability and reduce false matches.
- Built and deployed a scalable **PySpark** data validation framework to automate **ETL** reconciliation and migration QA (schema, completeness, metric parity), improving reporting trust and strengthening data-quality controls to support ongoing maintenance of downstream models and analytics.
- Automated **GA4 ETL** pipelines into **GCP BigQuery** for product engagement and lifecycle analytics, then integrated curated datasets into the Azure ecosystem (Microsoft Fabric/Power BI) to deliver dashboards that reduced reporting turnaround time and improved stakeholder decision-making.

**JAR** Mar 2024 – Aug 2024
**Data Scientist**
- Developed and deployed an internal ML-based affluence scoring model (**LightGBM**/**XGBoost**) to classify new users at sign-up into spending-affinity tiers; validated on holdout data and operationalized deployment with **MLflow**, replacing a third-party solution and saving **~$15k/month**.
- Developed a **Kafka**-to-**AWS S3** ingestion pipeline for transactional SMS events from **~20K DAUs**, orchestrated **Airflow** batch transforms and **Redshift** loads for analytics/modeling, and maintained operational datasets in **MongoDB** for downstream application use.
- Employed **NLP** techniques to process unstructured SMS data, converting it into structured formats for storage in **MongoDB**, enabling downstream analytics of user behavior and engagement.

**SYMPHONY AI** May 2023 – Aug 2023
**Data Science Intern**
- Built and optimized a **neural network** for motor health prediction using sensor-derived time-series features, improving accuracy by 23% through **feature engineering** and **hyperparameter tuning**, and delivering anomaly/risk scores to support maintenance decisions.
- Optimized production ML models (**SVM, XGBoost, Neural Networks**) through advanced **feature engineering, statistical analysis, and hyperparameter tuning**, boosting F1-score to 90% (confusion matrix/ROC-based evaluation) and enabling more reliable predictions at scale.
- Worked across **end-to-end ML workflows** including data preprocessing, feature engineering, model training, and validation across diverse model families (**tree-based**, **classical ML**, **deep learning**) to deliver reliable, production-oriented predictions.

**SHARP** Jun 2021 – Sept 2021
**Intern**
- Built end-to-end machine learning frameworks using **Python** libraries (**PyTorch, Scikit-Learn, TensorFlow, Keras**); performed preprocessing, feature engineering , and hyperparameter tuning to deliver high-accuracy, production-ready models
- Designed, executed, and analysed large scale **A/B tests** to evaluate business strategies and product features, applying statistical methods to generate actionable insights that directly guided product decisions and user engagement improvements.
- Optimized big data pipelines using **Apache Spark** & **Hadoop** to process large datasets, boosting efficiency by 4x and enabling scalable real-time analytics

## PROJECTS

**AlertVision - Computer Vision**
- Fine-tuned **YOLO** for high-accuracy detection of face, eyes, and head position to identify drowsiness
- Leveraged **CV2** (**OpenCV**) for real-time video input, image capture and preprocessing for the detection model
- Trained and deployed a deep learning model (**YOLO**, **PyTorch**) on captured images for real-time drowsiness detection and leveraged reinforcement learning techniques to improve decision-making accuracy in safety-critical use cases.

**FindMySquad - Full-Stack Web Development**
- Developed and deployed a full-stack web platform (**Node.js, Express.js , MongoDB, Handlebars, CSS, JavaScript**) for hosting/joining games, finding gym buddies, and organizing esports tournaments.
- Integrated real-time chat (**Websockets**), email reminders (**Nodemailer + Cron**), and interactive maps (**Leaflet.js**) to enhance collaboration and event coordination.
- Designed and implemented user engagement systems (profiles, ratings, karma leaderboard, achievements) to boost retention and fair play.

## SKILLS

**Data Science**: Python, Pandas, R, scikit-learn, Spark MLlib, Statistical Modeling, Model Validation, Classification, Time Series Analysis, A/B Testing, GBM (XGBoost/LightGBM), SVM, Neural Networks (PyTorch/TensorFlow/Keras), CNN, LSTM, spaCy, MATLAB, H2O
**Data Engineering**: SQL, PySpark, Databricks, ETL, MLflow, Kafka, Airflow, Azure, Data Modeling(Star/Snowflake), BigQuery, Redshift, GCP, AWS
**Software Engineering**: Git, Docker, CI/CD, JavaScript, Node.js, Express.js, React, REST APIs, MongoDB, WebSockets, Airflow, Kafka, System Design, OOP, Data Structures & Algorithms, HTML, CSS
**Independent Learning:** Generative AI / LLM tooling (Hugging Face, LangChain, vector databases) and model risk considerations for GenAI systems.