Esha Teware

+1 (717) 903-4428 | eshateware44@gmail.com | Washington, DC, USA | LinkedIn | GitHub | Portfolio

EDUCATION

American University **January 2025 - May 2026**
Master's, Data Science *GPA: 3.92*
- TA, Advanced Machine Learning; TA, Data Management
- Dean's Advisory Board
- Coursework: Machine Learning, Applied NLP, Regression, Bayesian Statistics, Statistical Programming, Database Systems

Pennsylvania State University **August 2018 - May 2023**
Bachelor's, Data Science *GPA: 3.5*
- Coursework: Big Data Analytics, Cloud Computing, Intro to Artificial Intelligence, Data Structures & Algorithms, Data Mining

SKILLS

- **Programming:** Python (pandas, NumPy, Matplotlib, scikit-learn, statsmodels), PyTorch, TensorFlow; SQL; R
- **Tools:** Docker, Git, Snowflake, Tableau, Power BI, R Shiny
- **NLP Specializations:** BoW/TF-IDF, Word2Vec embeddings, text classification (ML/CNN/RNN), NER, topic modeling (LDA)

PROFESSIONAL EXPERIENCE

Perry International Inc. **New York, NY, USA**
Data Analyst *February 2024 - January 2025*
- Supported a financial reallocation project by analyzing cross-market KPIs (Churn rate, Seasonality Adjusted Capital Efficiency ratios) using regression and data manipulation tools on Python to deliver insight that drove 15% reduction in digital footprint costs.
- Delivered a supply chain transformation project by building and validating demand-forecast models (MAPE < 17%) informing inventory and media-buy schedules and reducing stock-outs by 10%.
- Engaged with cross functional teams to define ELT pipelines to support the development of optimized Tableau dashboards by using custom queries and automated data extracts.

Koch Industries **Auburn Hills, MI, USA**
Data Scientist (FP&A) *May 2023 - February 2024*
- Designed modular Python ETL pipelines using pandas and NumPy for Guardian Glass production data, standardizing analytics across five North America plants and reducing data-prep time 40%.
- Built Dockerized analytics workflows and Snowflake-connected R Shiny dashboards delivering real-time KPIs to engineering and finance teams, eliminating manual refresh cycles.
- Implemented statistical anomaly-detection on energy-usage streams; insights drove process adjustments that successfully reduced utility expenses by 4%.

PROJECTS

Multi-Modal ML for Molecular Property Prediction (BACE-1) **Washington, DC, USA**
- Built an end-to-end multimodal pipeline to predict pIC50 for BACE-1 inhibitor molecules, using SMILES as the core molecular representation and feature source.
- Generated features across three modalities: NLP text embeddings from molecule-specific descriptions, Sum-over-Bonds (SoB) structured descriptors, and 2D molecular images rendered from SMILES for vision-based embeddings.
- Implemented a tri-modal comparison and fusion workflow (chemical + visual + textual embeddings), training and benchmarking NN models in PyTorch/TensorFlow using consistent splits with MAE/RMSE and error analysis.

Bank Marketing Success - Predictive Modeling & Campaign Strategy **Washington, DC, USA**
- Improved targeting for a Portuguese bank outreach campaign using 64K call data to predict term-deposit subscription; addressed strong class imbalance (~11.3% “yes”) and removed duration-based leakage.
- Built model evaluation in R using train-test split and 10-fold cross-validation (logistic, Ridge/LASSO, decision tree, random forest, SVM, neural network) evaluated with AUC/accuracy.
- Delivered best performance with Random Forest (AUC 0.802), translating key drivers into recommendations and a monitoring plan (threshold tuning, AUC/error tracking, transparent drivers).

Drowsiness.TV - Real-Time Drowsiness Detection POC **State College, PA, USA**
- Led a 6-person cross-functional team, managing timelines and stakeholder communication; delivered drowsiness-detection POC to incubation investors using facial-landmark analysis to auto-pause video platforms.
- Trained and validated a CNN classifier on eye-state images; improved accuracy from 85% to 94% and tested robustness across glasses wearers and multi-user scenarios.
- Built a full-stack pipeline (React + Node/Express + Python) achieving sub-100 ms response, enabling real-time drowsiness detection for evaluation and rapid iteration.

TRAININGS

NVIDIA Deep Learning Institute
- Fundamentals of Accelerated Computing with CUDA Python; Accelerate Data Science Workflows with Zero Code Changes; Accelerating End-to-End Data Science Workflows; Generative & Agentic AI Explained; Build Deep Research Agent; Intro to Multi-Modal Data Curation