Prashanth Javaji

346-599-8688 | Prashanth.jp1029@gmail.com | linkedin.com/in/prashanth-javaji/ | github.com/PRASHANTHJAVAJI

Education

Rice University — Houston, TX
Master of Data Science (GPA: 3.85/4.00) — *Aug 2024 – Dec 2025*

SRM University — Chennai, India
Bachelor of Computer Science and Engineering (GPA: 9.42/10.00) — *Sept 2020 – Jun 2024*

Technical Skills

Programming Languages: Python, C/C++, SQL, R

Machine Learning & AI: LLMs (BERT), NLP, PyTorch, Pandas, XGBoost, LightGBM, CatBoost, Scikit-learn, TensorFlow

Data Engineering & Cloud: Snowflake, AWS, Databricks, Git, PySpark, Django, MongoDB, ELT Pipelines

Data Visualization: Tableau, Power BI, Excel, Matplotlib, Seaborn

Certifications: Machine Learning, Meta Backend Developer (Coursera)

Experience

Rice University Facilities Engineering & Planning — Houston, TX
Data Analyst — *Dec 2024 – Present*

- Consolidated 25 Snowflake and NoSQL tables into 10 optimized analytical views, improving data consistency and enabling reliable facilities and capital planning insights.
- Built interactive Tableau dashboards with standardized keys and drill-through capabilities, enabling stakeholders to perform granular analysis across complex datasets.
- Automated recurring reporting workflows using reusable ETL pipelines, reducing reporting workload by ~80% and eliminating manual reconciliation errors.

EDP Renewables North America LLC — Houston, TX
Summer Associate, Energy Analytics — Link — *Jun 2025 – Aug 2025*

- Developed a Day-Ahead spread forecasting System, supporting bid strategy across 26 renewable assets.
- Engineered ETL pipelines with QA checks and a small feature store; trained weighted ensemble models (XGBoost, LightGBM, CatBoost, MLP) with calibrated probabilities.
- Implemented autoencoder-based feature selection, cutting retrain time from 31 to 5 min (~89%) while retaining ≥86% ROC-AUC and stabilizing model drift.

Rice University — Houston, TX
Teaching Assistant: Computer Networks, Statistics, Data Privacy — *Aug 2024 – Dec 2025*

- Taught and mentored 50+ students across Computer Networks, Data Privacy, and Statistics courses; led labs and assignments to support faculty and peers through collaborative problem-solving.

Machine Learning Projects

Streamflow Forecasting with Transformer Models — Link
LSTM, Temporal Fusion Transformer(TFT) and Patch Time Series Transformer(PatchTST) — Aug 2025 – Dec 2025

- Engineered a data pipeline for 310 Texas gauges using the CAMELS dataset, implementing rainfall-triggered segmentation with a 120-hour lookback window.
- Achieved a Kling-Gupta Efficiency (KGE) score of 0.8532 with the PatchTST model by optimizing patch-based temporal attention mechanisms for 24-hour forecasting.

Semantic Movie Search Engine & Recommendation System — Link
BERT, Autoencoder, Cosine Similarity, GraphSage — Feb 2025 – Apr 2025

- Developed a content-based recommendation system on ~7,570 IMDB titles using BERT embeddings to encode metadata into a semantic vector space for personalized suggestions.
- Benchmarked similarity models (cosine, transformer, GNN), achieving performance gains through optimized embedding strategies and hybrid model fusion techniques.

Detection of AI-Generated Text with BERT Model — Publication
BERT, Augmented LLMs, Tokenization — Oct 2023 – May 2024

- Designed and fine-tuned a BERT model, achieving a 92% accuracy rate in distinguishing AI-generated vs. human-written text from 28,000 essays using hyperparameter tuning.
- Applied advanced data preprocessing techniques (tokenization, text transformation) to improve model performance, leading to high F1 scores and balanced precision/recall trade-offs.