Prashanth Javaji

Prashanth.Javaji@rice.edu | (346)599-8688 | linkedin.com/in/prashanth-javaji/ | prashanthjavaji.github.io/prashanth | github.com/PRASHANTHJAVAJI

EDUCATION

RICE University, Houston, United States - ***Master, Data Science*** – ***GPA – 3.83/4*** Aug 2024 – Dec 2025
SRM University, Chennai, India - *Bachelor,* ***Computer Science and Engineering - GPA - 9.42/10*** Sept 2020 – Jun 2024

TECHNICAL SKILLS

Programming Languages	:	Python; SQL; R; C/C++
Machine Learning / Frameworks	:	clustering & segmentation; recommendation systems; NLP; LLMs; feature engineering
Libraries / Models	:	scikit-learn; PyTorch; TensorFlow; BERT; XGBoost; LightGBM; CatBoost; autoencoders
Data Engineering & Platforms	:	Snowflake; AWS; Databricks; Apache Spark/PySpark; Hadoop; MongoDB; ETL/ELT; Git
Analytics and Visualisation	:	Tableau; Microsoft Power BI; Matplotlib
Web technologies	:	HTML, CSS, Bootstrap, Web Development (Django)
Certifications	:	Machine Learning, Meta Backend Developer (Coursera)

WORK EXPERIENCE

Data Analyst

Rice University Facilities Engineering & Planning, Houston, Texas Dec 2024 – Present

- Consolidated 25 tables into 10 Snowflake views (star schema); standardized Snowflake→Tableau keys/filters, TRIRIGA-validated; analysis-ready tables; reduced manual reconciliations and cross-filter/drill-through errors.
- Cleansed and integrated Snowflake/NoSQL into analysis-ready tables; delivered reusable ETL for capital projects with cross- workbook filters and drill-through navigation, reducing reporting workload ~80%.

Summer Associate, Energy Analytics

EDP Renewables North America LLC, Houston, Texas | **Link** Jun 2025 – Aug 2025

- Built a day-ahead price/DA-RT spread forecasting system to guide bid strategy and risk management across 26 renewable assets.
- Engineered a Snowflake → pandas/PySpark ETL with QA checks and a small feature store; trained Supervised weighted ensemble model (XGBoost, LightGBM, CatBoost, MLP) with calibrated probabilities and real-time alerts.
- Proposed and implemented **unsupervised autoencoder**-based feature selection, cutting retrain time 31→3 min (~85%) while retaining ≥86% ROC-AUC and stabilizing model drift.

Teaching Assistant, Rice University, Houston, Texas

Introduction to Computer Networks, Intro to Statistics for Engineers & Scientists, Data Privacy & Security

- Taught and mentored 50+ students across computer networks, data privacy, and statistics courses, leading labs and assignments, and supporting both faculty and peers through collaborative problem-solving.

PROJECTS

Critics and User-Based Movie Recommendation System Using BERT Embeddings Feb 2025 – Apr 2025

github.com/PRASHANTHJAVAJI/Movie_Recommendation

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

Detection of AI-Generated Text with BERT Model Oct 2023 – May 2024

ieeexplore.ieee.org/abstract/document/10692072

- Designed and fine-tuned a BERT model, achieving a 92% accuracy rate in distinguishing AI-generated vs. human-written text from 28,000 essays by applying 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.

Prediction of Obesity Risk with Logistic Regression May 2023 – Nov 2023

github.com/PRASHANTHJAVAJI/Prediction-of-obesity-risk-with-Logistic-Regression

- Implemented logistic regression, random forest, and gradient boosting to predict obesity risk, achieving **85% accuracy** by optimizing hyperparameters on a dataset of **1,100 individuals**
- Identified **key obesity risk predictors**, providing actionable insights that informed public health strategies and supported targeted intervention efforts.