Pavani Jain

jain.pav@northeastern.edu | github.com/Pavanijain03 | linkedin.com/in/pavani-jain-ab6488204 | pavanijain03.github.io | (617)-516-9584

Experience

Data Analyst Intern **July 2024**

Verve Bridge *Delhi, India*

- Performed Exploratory Data Analysis (EDA) on 50K+ records to identify trends, correlations, and anomalies in structured datasets.
- Built interactive Tableau dashboards with KPIs and heatmaps, improving stakeholder reporting efficiency by ˜30%.
- Automated data preprocessing pipelines using Python (Pandas, NumPy), reducing manual analysis effort by ˜40%.

Member, Northeastern AI Club **Sep 2025 – Present**

Northeastern University *Boston, MA*

- Participated in 6+ AI-focused workshops and hands-on sessions covering machine learning and deep learning concepts.
- Applied learned techniques in small group exercises involving data preprocessing, model training, and result interpretation.

Skills

Programming Languages: Python, R, Java, C, HTML, CSS, Bash
Machine Learning & AI: Supervised & Unsupervised Learning, Deep Learning, Natural Language Processing (NLP), Neural Networks, Feature Engineering, Model Evaluation, Anomaly Detection, Gen AI
Data Science: Exploratory Data Analysis (EDA), Statistical Analysis, Data Visualization, Hypothesis Testing
Libraries & Frameworks: PyTorch, TensorFlow, Keras, scikit-learn, Pandas, NumPy, Seaborn, Matplotlib
Databases & Data Engineering: SQL, MySQL, Data Cleaning, ETL Pipelines
Tools & Platforms: Tableau, AWS (Cloud DevOps), Git, Linux

Projects

Genomic Text Curation & Topic Grouping **Jan 2026 – Feb 2026**

Python, NLP, Scikit-learn, TF-IDF, Topic Modeling

- Curated 50 genomics research texts using an NLP pipeline to extract structured biological information.
- Extracted 40 variants, 37 genes, and 19 diseases, generating 76 knowledge-graph–ready triples with 84% variant and 78% gene coverage (˜5s).
- Applied TF-IDF + K-Means, NMF, and LDA to cluster texts into 5 interpretable biomedical research topics.

Credit Card Fraud Detection **Apr 2025 – Aug 2025**

Python, Machine Learning, Deep Learning

- Developed anomaly detection models using autoencoders, achieving 92% recall in identifying fraudulent transactions.
- Analyzed Logistic Regression, Random Forest, XGBoost, and Neural Network models, with XGBoost achieving 94% ROC-AUC.
- Handled class imbalance and performed feature engineering to improve model robustness on highly skewed fraud datasets.

Amazon Product Recommendation System **Sep 2024 – Dec 2024**

Python, Machine Learning

- Built a collaborative filtering–based recommendation system, improving recommendation accuracy with an 18% reduction in RMSE.
- Evaluated recommendation quality using MAE and RMSE metrics across multiple model configurations.
- Analyzed user–item interaction patterns to identify latent preferences and enhance recommendation relevance.

Certifications

Deep Learning Specialization | Generative AI Certification | Java with Data Structures

Education

Northeastern University **Sep 2025 – Present**

Master of Science in Artificial Intelligence *Boston, MA*

- Relevant Coursework: Foundations of Artificial Intelligence, Machine Learning, Pattern Recognition, Algorithms