

Harneet Kaur

Data Science | Machine Learning | Deep Learning | NLP & Statistics

Phone No.: +91 8018724774 Mail: harneetkaur4464@gmail.com | [LinkedIn](#) | [GitHub](#) | [Website](#)

PROFILE

MSc Data Science candidate with a strong foundation in Statistics and hands-on experience working with large, real-world datasets. Experienced in building NLP and recommendation systems, handling imbalanced data, and designing evaluation-driven ML pipelines. Seeking data science or machine learning internship roles focused on applied problem solving and measurable impact.

EDUCATION

- VIT University, Vellore, India | M.Sc. Data Science | 2024 – 2026**
 - CGPA: 9.46 / 10.0 (Sem 1: 9.56, Sem 2: 9.35)
- BJB Autonomous College, Bhubaneswar, India | B.Sc. (Hons.) Statistics | 2020 – 2023**
 - CGPA: 9.29 / 10.0 — Rank 2 in Department (Distinction)
 - Achievements:** AIR 607 (IIT JAM – Statistics), AIR 348 (CUET PG)
- Class 12th: 93.4% | Class 10th: 95.2

ACADEMIC PROJECTS

NLP-Powered Mental Health Status Detection | *Python, Transformers, Scikit-learn* | Jul 2025 – Nov 2025

- Built an NLP pipeline classifying 51,000+ social media posts into 7 mental health categories; addressed severe class imbalance using stratified sampling and metric-driven evaluation.
- Benchmarked MentalBERT, BioBERT, and PubMedBERT; selected MentalBERT based on superior minority-class performance ($F1 = 0.84$), especially for Personality Disorders.
- Designed an emotion-aware preprocessing pipeline preserving linguistic cues (capitalization, punctuation), improving Logistic Regression baseline performance by 15%.

Hybrid E-commerce Recommendation System | *Sentence-BERT, TensorFlow* | Sep 2024 – Sep 2025

- Engineered a hybrid recommendation engine on 75,000+ user-item interactions to mitigate the cold-start problem in sparse environments.
- Generated 384-dimensional semantic embeddings using Sentence-BERT from product descriptions to enable context-aware recommendations beyond ratings.
- Achieved $\text{Precision@10} = 13.26\%$ and $\text{NDCG@10} = 29.37\%$, delivering a 4× improvement over sentiment-based baselines.

TECHNICAL SKILLS

Languages: Python, R, SQL.

Frameworks & Libraries: TensorFlow, Keras, Hugging Face Transformers, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn.

Core Competencies: Machine Learning, Deep Learning, NLP, Statistical Inference, Feature Engineering, Model Evaluation.

Tools: Git, GitHub, Apache Spark, Hadoop, Jupyter.

CURRENT FOCUS

- Agentic AI and Retrieval-Augmented Generation (RAG)
- Model evaluation and trade-offs across classical ML and transformers
- End-to-end ML pipelines with error analysis and reproducibility
- Baseline modeling and ablation studies to justify model complexity
- Data quality checks and leakage prevention in ML workflows