Harneet Kaur

Email: harneetkaur4464@gmail.com | **LinkedIn** | **GitHub** | **Website**

As a Data Science master's candidate with a distinguished academic background in Statistics, I'm driven by a natural curiosity for the stories hidden within data. I excel at building and refining machine learning models that turn complex information into clear, compelling insights. I am looking to join an innovative team where I can apply my skills to build products that solve real-world problems.

Education

- **M.Sc. Data Science** | VIT University, Vellore | 2024 - Present
  - **Current CGPA: 9.46** (Sem 1: 9.56, Sem 2: 9.35)
- **B.Sc. (Hons.) Statistics** | BJB Autonomous College, Bhubaneswar | 2020 - 2023
  - **Final CGPA: 9.3/10 (Ranked 2nd in department)**
- **Higher Secondary Education (CBSE)** | Dalmia Vidya Mandir | 2018 - 2020
  - **Class 12:** 93.4%
  - **Class 10:** 95.2%

Projects

NLP-Powered Mental Health Classification (Capstone Project) | Jul 2025 - Present

- Architected a multi-class NLP pipeline to classify 51k+ social media posts, resolving severe class imbalance using a hybrid sampling technique (strategic downsampling and SMOTE).
- Systematically evaluated 7 machine learning models, identifying a hyperparameter-tuned SVC (RBF) as the champion classifier with a peak F1-score of 0.721.
- Currently benchmarking the optimized SVC against advanced deep learning architectures, including a custom LSTM and a fine-tuned DistilBERT, to determine the state-of-the-art model for the task.

Hybrid Product Recommendation System | Sep 2024 - Sep 2025

- Developed a hybrid recommendation engine combining collaborative filtering with content-based filtering, using TF-IDF vectorization on product descriptions to enrich user profiles.
- Achieved a 480% increase in product discovery for users by delivering more relevant and personalized suggestions.

Skills

- **Languages & Programming:** Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly), R, SQL
- **Data Analysis & Statistics:** Exploratory Data Analysis (EDA), Hypothesis Testing, A/B Testing, Statistical Inference, Probability Distributions
- **Classical Machine Learning:**
  - **Regression:** Linear, Logistic, Regularization (Ridge, Lasso)
  - **Classification:** SVM, Decision Trees, Random Forest, KNN
  - **Clustering:** K-Means
  - **Techniques:** Feature Engineering (TF-IDF, Word2Vec), Hyperparameter Tuning (GridSearchCV, RandomizedCV), Handling Class Imbalance (SMOTE)
- **Deep Learning & NLP:**
  - **Frameworks:** TensorFlow, Keras
  - **Architectures:** LSTMs, Artificial Neural Networks (ANNs)
  - **Tasks:** Text Classification, Sentiment Analysis, Named Entity Recognition (NER), Topic Modeling
- **Tools, Systems & Big Data:** Git & GitHub, Jupyter Notebook, Google Colab, Relational Database Design (Normalization, ER Diagrams), Apache Spark, Hadoop Ecosystem (HDFS, MapReduce)

Certifications

- Getting Started with PowerBI - Linkedin Learning (Ongoing)
- 100 Days of Machine Learning by CampusX - YouTube (Ongoing)
- 100DaysOfCode in Python - Udemy (Ongoing)
- Education on Sustainable Development (100/100 Score) - NPTEL
- GenAI Cybersecurity: OWASP Top 10, MITRE ATLAS & API Attacks - Udemy
- Data Analytics using Excel - Udemy

Achievements

- Qualified IIT JAM (Statistics): All India Rank 607
- Qualified CUET PG (Statistics): All India Rank 348