

# Henil Gajjar

Boston, MA • +1 (857) 492 8741 • gajjar.he@northeastern.edu • [LinkedIn](#) • [Github](#)

## Education

**Master's in Data Science - Northeastern University** (Boston, USA)

August 2023 - Present

**Related Coursework:** Large Language Models, Supervised and Unsupervised Machine Learning, Data Mining

GPA: 4.0/4.0

Head Teaching Assistant for CS2810 (Data Modelling) and DS3000 (Foundations of Data Science)

**Bachelor's in Electronics and Communication Engineering with Minor's in CSE - Nirma University** (Ahmedabad, India)

June 2023

**Related Coursework:** Database Management, Applied Statistics, Machine Learning, Computer Vision

GPA: 3.9/4.0

## Skills

**Languages**

Python, C++, R, Dart, MATLAB

**ML Frameworks**

Hugging Face, PyTorch, NLTK, Scikit-Learn, TensorFlow, OpenCV, Keras, Pandas, Flask

**Database and Visualization**

MySQL, MongoDB, Google Firebase, Power Bi, Tableau

**Tools and Platforms**

Github, Amazon AWS, Docker, MLFlow, Data Version Control (DVC), Google Cloud Platform, Flutter

**Techniques**

Generative AI, Natural Language Processing, Statistical Modeling, Data Modeling, Predictive Modeling

## Experience

**Head of AI/ML - Hyperlab Sportech Pvt. Ltd.** (Gandhinagar, India)

January 2022 - August 2023

- Engineered a ML-driven mobile application for an athlete training device 'Helios' resulting in over 5k+ downloads and 4.7-star rating on App store and Play Store within first week of the launch.
- Employed LSTM model for timeout drills utilizing past athlete trainings, pivotal in securing \$25M valued investment on Shark Tank.
- Established streamlined CI/CD deployment on AWS through Github Actions, reducing deployment time by 100%.
- Optimized MongoDB data schema post ETL, achieving 30% reduction in query response time through refined data modeling and indexing.
- Developed a CNN-based novel Table Tennis ball prediction system, driving 15% increase in a player's ball placement accuracy post training.

**Student Researcher - Nirma University** (Ahmedabad, India)

May 2021 - December 2021

- Curated a dataset of a Li-ion battery pack encompassing over 150k rows, extracting actionable insights for data-driven strategies.
- Utilized Random Forest regression model to predict optimal temperature during charging/discharging, improving battery health by 12%.
- Employed clustering techniques to understand current rates and temperature of the battery pack over 1300 charge-discharge cycles.

## Projects

**End-to-End Kidney Tumor Classification Using Enhanced VGG16 Model** [\[Github\]](#)

June 2024

- Developed an end-to-end classification pipeline using modified VGG16 model, achieving 97% accuracy compared to original model's 92%.
- Implemented comprehensive data ingestion, model training, and evaluation workflows utilizing MLflow and DVC for effective experiment tracking and version control.
- Deployed the model as web application on AWS using Github Actions, Docker and EC2, providing a platform for real-time classification.

**Rent the Runway Fashion Recommender System** [\[Github\]](#)

Feb 2024 - April 2024

- Scraped the product website to collect product attributes and user reviews, generating comprehensive datasets for products and reviews.
- Employed matrix factorization techniques for collaborative filtering, cosine similarity for content-based filtering, and user attribute matching to recommend products, addressing cold start problems for new users and products.
- Implemented incremental SVD for updating recommendations to handle cold start issues, reducing system update time by 40%.

**Fantasy Team Recommendation for IPL 2024** [\[Github\]](#)

Feb 2024 - April 2024

- Fine-Tuned Gemma & Llama-2 using cricket dataset resulting in 10% improvement in ROUGE score for cricket-specific text generation tasks.
- Established algorithm to extract structured data from unstructured IPL historical data including player stats and match scorecards.
- Leveraged Prompt engineering (2-shot) to enhance prediction accuracy resulting in 85% accurate team prediction for the IPL matches.

**Advanced Performance Metrics for Ultimate Frisbee Athletes**

November 2023 - January 2024

- Integrated machine learning and mixed effects linear modeling techniques to develop non-box score player rating systems.
- Implemented on-off plus-minus models resulting in 20% increase in predictive accuracy, revealing performance on offense and defense.
- Identified top-performing players with a 15% boost in composite ratings, offering a comprehensive evaluation of player contributions beyond traditional box score statistics.

## Patent and Selected Publications

**Patent:** Steering System for Autonomous Solar Electric Vehicle. IP India, 363614-001, Issued July 12, 2022

[Patent](#)

**Publications:** A Comprehensive Study on Lane Detecting Autonomous Car using Computer Vision (Elsevier ESA), June 2023,

[Article](#)

A Comparative Analysis of Various Deep-Learning Models for Noise Suppression (EAI endorsed Publication), July 2023

[Article](#)