

# Eshita Gupta

gupta.eshita01@gmail.com | (669) 577-6251 | Sunnyvale, CA | [Portfolio](#)

## WORK EXPERIENCE

### San Jose State University

Graduate Research & Data Science Assistant  
San Jose, CA

Jan 2025 – Present

- Investigating generative modeling approaches (GNNs, diffusion models) for inverse RNA design, enabling data-driven generation of stable 3D RNA wireframe polyhedra not observed in nature.
- Supported instruction for **Computational Programming for Data Analytics**, mentoring 100+ students in Python, object-oriented design, and data structures.
- Implemented an **LLM-based grading assistant to auto-match submissions** to baseline solutions and generate personalized feedback, reducing grading time by 45%.

### Canara Bank

Data Scientist

Dec 2020 – Jun 2023

Delhi, India

A leading bank with 10,000+ branches and 80000+ employees, managing a multi-billion-dollar retail and MSME credit portfolio.

- Built supervised **credit risk models (XGBoost, LightGBM)** with IV-based feature selection, improving loan approval precision 27% and reducing delinquency 15%.
- Implemented **fraud detection pipelines using ensemble models** and graph analysis with velocity checks and device fingerprinting at 100K+ daily transaction scale
- Developed **LTV segmentation models using supervised (gradient boosting) and unsupervised learning (k-means)**, optimizing underwriting and cross-sell strategies for 12% conversion lift.
- Created **real-time Tableau dashboards monitoring fraud, charge-offs, and delinquency** across \$50B portfolio, enabling proactive risk management
- Led **A/B testing framework for credit strategies** using propensity models and champion-challenger methodology, achieving 12% incremental cross-sell lift.
- Performed **NLP sentiment analysis (BERT, topic modeling)** on customer feedback and social media posts, driving 25% improvement in satisfaction scores by data driven service optimization.

### Canara Bank

Data Analyst

Aug 2019 – Dec 2020

Delhi, India

- Created **interactive Tableau dashboards** tracking delinquency roll rates and recovery curves with vintage and segment drilldowns, enabling early intervention strategies that improved recovery rates 25%.
- Automated data pipelines (SQL, Python)** feeding Tableau dashboards with daily refresh schedules, streamlining credit portfolio reporting and reducing manual processing 50% (~3.5 hours per day)
- Established data quality frameworks with automated validation checks in Python, reducing data discrepancies by 40%.

## SKILLS

- Data Science & Analytics:** Predictive Modeling, Classification, Clustering, Forecasting, Causal Inference, A/B Testing, Metric Design, ETL/ELT
- Programming & Engineering:** Python, SQL, PyTorch, TensorFlow, PySpark, dbt, Linux, Docker, GitHub, REST API, CI/CD
- Cloud & Visualization:** AWS (S3, Lambda, RDS, Athena, EC2, SageMaker), Kafka, Airflow, Hadoop, Tableau, Looker, Power BI, MS Excel
- Certification:** Google Data Analytics Professional Certificate
- Interests:** Swimming, Painting, Plant mom, Yoga

## PROJECTS

### LLM Tuning for Chain-of-Thought Reasoning

- Built an AI math tutor application by fine-tuning LLMs (LLaMA, Gemma, Phi, DeepSeek) with COT reasoning, improving accuracy by 15% over the baseline LLM models. Integrated a continuous retraining workflow using Airflow hosted on Oracle Cloud Instance.

### Text Toxicity Classification - NLP

- Developed models for multi-class toxicity classification using LR, SVM, XGBoost and BiLSTM leveraging FastText and Glove embeddings. Achieved 95.8% accuracy and 18% false positive reduction on a 1.8 million record dataset.

### World Bank Open Data - Study into Food Insecurity

- Developed Food Insecurity Severity Index (FISI) by analyzing 44 global indicators (1963–2023) and applying PCA + k-means clustering, uncovering region-specific patterns and generating insights that informed policy recommendations for reducing food insecurity worldwide.

### Anime Recommender System (Wide & Deep Learning)

- Implemented a hybrid recommendation system using Google's Wide & Deep Learning approach on 23M+ records, achieving 0.0729 RMSE after optimized sampling and feature engineering.

## EDUCATION

### San Jose State University

Master of Science, Applied Data Science

Aug 2023 – May 2025

San Jose, CA

**Coursework:** Data Structures & Algorithms, Object Oriented programming, Distributed Computing Systems, Database Management System, Machine Learning, Deep Learning, Big Data, Data Mining, Artificial Intelligence (AI)

### National Institute of Technology, Allahabad

Master of Science, Mathematics and Computer Science

Allahabad, India