# Gnana Prasuna Nimeesha Vakacharla

San Jose, CA | nimeeshav02@gmail.com | +1 (203) 809-7684 | Portfolio | LinkedIn | GitHub

## Education

**San Jose State University**  <span style="float:right">San Jose, United States</span>

Master's in Data Analytics  <span style="float:right">January 2024–December 2025</span>

## Technical Skills

**Languages**: Python, C, R, JavaScript, HTML, CSS, ReactJS, NoSQL, SQL

**Data Engineering and Cloud**: Apache Spark, Hive, Hadoop, AWS (S3, EC2, Lambda), GCP (BigQuery), Azure

**Libraries and Visualization**: Numpy, Pandas, Matplotlib, SciPy, Seaborn, Tableau, PowerBI, Plotly, MS Excel

**Machine Learning**: Scikit-learn, XGBoost, LightGBM, OpenCV

**Deep Learning and Generative AI**: TensorFlow, PyTorch, CNN, LSTM, RAG, NLP, LLM, Computer Vision

## Experience

**Data Scientist**, iQuardra – Remote  <span style="float:right">May 2024 – July 2024</span>

- Collaborated with cross-functional teams within iQuadra's AI division to **develop enterprise book recommendation system, processed 5.6 GB dataset with 3M+ reviews** using Hadoop HDFS and PySpark
- Implemented 6 MapReduce jobs analyzing user behavior patterns and preferences, **reduced data processing time by 65%** while ensuring scalable architecture for enterprise deployment requirements
- Engineered collaborative filtering models using Apache Spark MLlib's ALS algorithm (**RMSE 1.28**) and content-based filtering using MinHashLSH for **212K books**, **achieving 78% user engagement improvement and 42% recommendation accuracy boost**
- Deployed large-scale data transformation pipeline converting 2.89 GB of raw data into actionable insights, integrated Flask web application with interactive Plotly/Dash visualizations, **decreased response time by 60%**

## Academic Projects

**Cloud Analytics on New York Times Data** | Python, Apache Airflow, GCP, BigQuery, Power BI

- Automated ELT pipeline extracting 500K+ JSON records from NYT APIs using Python & Airflow, **achieving 99.9% reliability and processing 1TB+ historical data through orchestrated workflows**
- Created GCP data warehouse with BigQuery and star schema design using DBT transformations, **delivering 40% query performance improvement while maintaining zero data corruption across 1,000+ daily records**
- Built ML forecasting model with polynomial regression achieving **85% prediction accuracy** and Power BI dashboards showing **11.7% growth across key performance indicators for stakeholder decision-making**

**DIS-EASIFY** | Python, Django, Azure, Scikit-Learn, TensorFlow, Docker, Kubernetes, MLOps

- Developed AI-powered healthcare platform using Python, Django, TensorFlow **achieving 97.6%-100% accuracy** across heart disease, diabetes, pneumonia, breast cancer detection for global populations
- Architected scalable Azure cloud infrastructure with Docker/Kubernetes, CI/CD pipelines **serving 1000+ concurrent users** globally with **<2s response time** and HTTPS security
- Engineered full-stack solution preprocessing 5+ Kaggle datasets, implementing Naïve Bayes, Random Forest, CNN, Decision Trees with **80:20 validation** and comprehensive testing
- **Delivered 60% healthcare accessibility** improvement through probability-based assessments, HIPAA-compliant frameworks **reducing false positives by 85%** via intuitive symptom-to-diagnosis workflow

**Interview Prep-AI** | Python, RAG, Vector DB, Airflow, GCP, Mistral-7B, Llama-4

- Streamlined self-executing data pipeline web-scraping 4,500+ interview datasets using Airflow Composer on GCP with RAG architecture and vector databases, **achieving 40% reduction in preprocessing time**
- Fine-tuned ML models (Mistral-7B, Llama-4) with agentic AI framework **achieving 90% question generation and 85% answer evaluation accuracy using multimodal speech-to-text** integration, **improving simulation effectiveness by 30%**
- Architected scalable system with vector database **optimization achieving 97% response uniqueness** and **<5 second query time**, supporting 100+ concurrent mock interviews with agent-based evaluation framework

## Achievements

- Secured **top 8% placement in NVIDIA AI** Hackathon by identifying emerging industry patterns and directing technical strategy to **surpass 200+ competing teams**
- **Coordinated departmental operations** as Department Representative for Civil Engineering Department, overseeing **academic and co-curricular activities** for **500+ students** while managing **cross-departmental requirements**