

Saumya Varshney

San Jose, CA 95134 | (+1) 669-225-5237 | saumya2603@gmail.com | [LinkedIn](#) | [GitHub](#)

EDUCATION

Master of Science, Data Analytics

May 2025

San Jose State University, California, USA (GPA - 3.6)

Coursework: Machine Learning, Big Data, Statistics, Deep Learning, Data Visualization & Database Systems.

Bachelor of Technology, Computer Science and Engineering

Jun 2017

Bundelkhand Institute of Engineering and Technology, India (GPA - 3.5)

Coursework: Data Structures, Data Mining & Warehousing, Distributed System & Design of Algorithms.

TECHNICAL SKILLS

Languages:	Python, SQL, NoSQL, Pyspark, Cypher QL, C#
Cloud Platforms:	AWS, Snowflake, GCP
Machine Learning:	AI/ML Models, NLP, Pytorch, TensorFlow
Tools/Platforms:	CI/CD, Docker, Google Apps Script, Tableau, Power BI, Microsoft Excel, ASP.NET
Data Warehouse/ETL:	Apache Airflow, Redshift, Google Big Query, AWS Glue, Apache Spark, SSIS, SSRS

WORK EXPERIENCE

Student Assistant, San Jose State University, CA

Feb 2024 - Present

- Orchestrated deployment and execution of **GPU-accelerated** deep learning models on HPC machines using **Docker**.
- Streamlined lab reservation process by developing **Google App Script** solution, resulting in a seamless and user-friendly booking experience.
- Automated user account management using **PowerShell remoting scripts**, boosting admin efficiency by **90%**.

Research Assistant, San Jose State University, CA

May 2024 – Aug 2024

- Integrated an **AI-enabled robot dog** with ML models, boosting autonomous data processing capabilities by **30%**.
- Augmented robot dog's functionality as a **voice assistant** by training it with LLMs revamping human-robot communication and interaction by **20%**.

Application Development Analyst, Accenture Services Pvt Ltd, Bangalore, India

Mar 2021 - Feb 2022

- Led an end-to-end **Azure migration**, navigating complexities to achieve system transition in 3 months using **agile method**.
- Analyzed production environment and collaborated with **cross-functional teams** to reduce code redundancy, improving throughput by **20%**.
- Converted Crystal Reports to **SSRS** and integrated with SaaS applications, reducing reporting time by **25%**.

Project Engineer, Wipro Limited, Hyderabad, India

Feb 2018 - Mar 2021

- Overhauled vital sales and manufacturing application utilizing **.Net, C#, SQL** scripting, delivering high-quality outcomes.
- Spearheaded enhancement of client's management application **backend** to boost page load time by **25%**.
- Directed **software release cycles** and implemented feature enhancements, achieving **99%** system **availability**.

PROJECT EXPERIENCE

Paraphrase Identification using Quora Question Pairs, San Jose State University, [Project Link](#)

Spring 2024

- Leveraged NLP with an **LSTM** model, utilizing word tokenization and embeddings, to achieve **81% accuracy** in predicting question similarity.
- Performed **PCA** on embeddings to reduce dimensions by **50%** and trained model, improving training efficiency by **30%**.

Bay Area VTA Ridership Forecasting, San Jose State University, [Project Link](#)

Spring 2024

- Forecasted ridership data with **80%** accuracy by integrating weather data and deploying advanced ML models, including Random Forest, Decision Trees, and **XGBoost**.
- Merged, cleaned, and transformed data, achieving model accuracy by **35%**, obtaining high accuracy with XGBoost.

AWS-Enabled Data Pipeline for NOAA Weather Data Analysis, San Jose State University, [Project Link](#)

Fall 2023

- Enhanced real-time analytics with **AWS Glue** by automating ETL processes, reducing processing latency by **25%**, and accelerating insights from weather data.
- Boosted efficiency by **20%** by integrating **Redshift**, optimizing query performance and storage in weather data pipeline.

Analysis of LinkedIn Job Market Data Using MySQL and MongoDB, San Jose State University, [Project Link](#)

Fall 2023

- Conducted SQL and NoSQL analyses on job data, deriving actionable job market insights; enhanced AWS **RDS** performance by **20%** with 3NF ER diagrams.
- Increased database efficiency by **15%** with optimized **NoSQL/MySQL** setups and complex **MQL** queries, integrated with MongoDB.