Priyanka Nidadavolu

+1 (213) 760-5184 | pnidadavolu@ucsd.edu | LinkedIn | GitHub | Portfolio

Summary

Data Science MS student at UC San Diego, with 2.5 years at IBM where most of the meaningful work happened at the edge between data and the people using it. Good at figuring out why something isn't working for someone, then rebuilding it in a way that holds. Interested in the Design Lab because the research here starts with the human question, which is usually where the right technical answer eventually shows up.

Education

University of California, San Diego Sept 2025 – Present
Master of Science in Data Science GPA: 3.77/4.0 *San Diego, CA*

- Coursework: Machine Learning, Recommender Systems, Statistical Modeling, Big Data Analytics, Scalable Data Systems, Data Systems for ML

K L University 2019 – 2023
B.Tech in Computer Science GPA: 3.67/4.0 *Vijayawada, India*

- Coursework: Data Structures & Algorithms, Database Management, Data Warehousing and Mining, Data Visualization

Experience

IBM Jan 2023 – Jul 2025
Software Developer (promoted from Intern) *Bangalore, India*

- Worked on **WatsonX** AI evaluation: wrote test prompts, reviewed output quality for correctness and consistency, and documented what the model kept getting wrong so the product team had something concrete to act on.
- Rebuilt a data validation pipeline that was silently failing and taking down batch jobs with it. Spent time first understanding how engineers actually used it, then restructured around those workflows. Processing time dropped **90%**.
- Built an **LLM**-based router that analyzed code diffs and decided which cloud environments actually needed testing. Wrote custom metrics to evaluate its decisions. Cut redundant CI/CD runs and saved **40%** in compute costs.
- Ran clustering and time-series analysis on multi-cloud usage logs to find waste patterns. Wrote the findings up in plain language so they'd land with stakeholders who weren't looking at the raw data.
- Contributed to multi-cloud provisioning tooling: built backend services using Python and IBM Cloud APIs to automate infrastructure setup across AWS, Azure, and IBM Cloud.

Skills: Python, LLMs, scikit-learn, data validation, CI/CD, SQL, stakeholder communication

Samsung PRISM Jun 2021 – Mar 2022
Research Intern, NLP & Speech *Remote*

- Built a pipeline to generate phoneme data from **10,000**+ contact names using the CMU Pronouncing Dictionary. Ran accuracy checks against ground truth; the **10%** gain in recognition fed directly into the next model iteration.
- Worked closely with research leads to define what "good enough" looked like for the model, surface edge cases that automated checks missed, and communicate findings in a way that shaped follow-up decisions.

Skills: Python, pandas, NLP, Wav2Vec2, research collaboration

UC San Diego, Information Technology Services Oct 2025 – Present
Student Build & Release Engineer (Part-Time) *San Diego, CA*

- Keeps CI/CD running for campus engineering services: catches broken builds early, automates release checks, and writes deployment docs so the next person doesn't have to reverse-engineer what happened.

Skills: CI/CD, Git, Jenkins, technical documentation

Projects

GridGreen 2nd Place, Cloud Track – DataHacks 2026 Apr 2026

- ML engineers had no visibility into the carbon cost of a training run before starting it. Built a tool that reads their script, estimates the footprint, and recommends a leaner model with concrete trade-offs in plain language.
- Designed the carbon estimation engine around published scaling laws, with methodology and limitations surfaced to the user so they could trust the numbers (or know when not to).
- Shipped the full stack: Databricks DLT pipeline for EIA grid data, 48-hour Prophet forecasts, Monaco editor frontend, and MCP integration so it works inside Claude Desktop or Cursor without context-switching.

Skills: Python, FastAPI, RAG, Databricks DLT, AWS SageMaker, Snowflake, Prophet, Next.js

MediDB Clinical Drug Safety & Recommendation Apr 2026

- Built a system that surfaces medication options and flags unsafe drug combinations before they reach a patient. The architecture was designed around how clinicians actually reason, not just what made the ML easier.
- Wired together four databases so each handled the query it was built for: graph queries for interactions (Neo4j), vector search for adverse events (Qdrant), relational for EHR (PostgreSQL), documents for audit trails (MongoDB).
- Wrote ETL from synthetic Synthea EHR records with documented assumptions at every stage, so anyone picking this up later would know where the data came from and what had been cleaned.

Skills: Python, SQL, PostgreSQL, Neo4j, Qdrant, MongoDB, FastAPI

Bank Marketing Campaign Analysis 2024

- Dug into 40,000+ customer records from a Portuguese bank to find what actually predicted term deposit sign-ups. Turned those findings into Tableau dashboards that non-technical stakeholders could use directly.
- Caught and flagged a data leakage issue (call duration) early on, before it could make the model look better than it was. Wrote up the implications so the team understood the constraint.

Skills: Python, Tableau, pandas, Matplotlib, Seaborn, data storytelling

Socially-Aware Recommendation System live demo Dec 2025

- Recommendation on Epinions is hard because 99.99% of the rating matrix is empty. Standard collaborative filtering just gives up. Used social trust networks as a signal instead, and beat a published IJCAI 2017 baseline by 6%.
- Deployed as a live interactive app with A/B model comparison, so anyone could test how the approaches differed on real queries. Containerized with Docker, hosted on Hugging Face Spaces.

Skills: Python, NetworkX, XGBoost, FastAPI, Docker

Achievements & Certifications

- Won **2nd Place in the Cloud Track** and the **Best Use of Snowflake Cortex Award** at DataHacks 2026, a 36-hour hackathon at UC San Diego, for GridGreen, a carbon analytics platform built on Snowflake Cortex, Databricks, AWS SageMaker, and MCP.
- Received the **Star of the Month** award at IBM in November 2023 for consistent contributions across ETL pipeline rebuilds, SQL warehouse optimization, and delivering analytical reports adopted by Finance and Engineering teams.
- Received the **People's Choice Award** at IBM for the multi-cloud provisioner project, which standardized usage data across AWS, Azure, and GCP and led to a **40%** reduction in cloud costs.
- Placed in the top **30 nationally** in the **Codehers Coding Challenge** in 2022.
- **Certifications:** AWS Academy Cloud Foundations, Big Data Analytics (UCSD/Coursera), IBM Data Science Foundations Level 1 & 2.