

Tejaswi Neelapu

Seattle, WA | tejaswirupa2001@gmail.com | (206)856-4851 | [LinkedIn](#) | [Github](#)

EDUCATION

Master of Science in Data Science, GPA: 3.8/4.0

September 2023 - June 2025

Seattle University, Seattle, WA

- Course work: Machine Learning, Statistics, Probability, Big Data Analytics, Numerical Methods for ML.

Bachelor of Technology in Mechanical Engineering, GPA: 3.0/4.0

July 2018 - June 2022

Amrita Vishwa Vidyapeetham, Kerala, India

- Course work: Computer Architecture, Probability and Statistics, Numerical Methods, Operations Research.

TECHNICAL SKILLS

Languages: Python, SQL, R

Machine Learning: Supervised/Unsupervised Learning, Deep Learning, Scikit-learn, PyTorch

Data Engineering: Hadoop, Hive, Spark, Snowflake, dbt (Data Build Tool), ETL Pipelines, Docker

Databases & Cloud: AWS (EC2, Lambda, API Gateway, S3, QuickSight), MongoDB, MySQL, PostgreSQL

Data Analysis: Statistical Modeling, A/B Testing, Hypothesis Testing, Time Series Analysis, Tableau, Power BI

WORK EXPERIENCE

Data Engineer - WW Sustainability Science and Innovation, Amazon, Seattle, WA

October 2025 - Present

- Building **end-to-end data pipelines** to automate workflows for the **Material Flow Analysis (MFA)** program, enabling data-driven sustainability decisions at global scale.
- Designing **AWS QuickSight dashboards** to visualize environmental KPIs and model results, improving accessibility of sustainability metrics for scientists and leadership teams.

Data Scientist (Capstone), Costco Wholesale, Seattle, WA

January 2025 - June 2025

- Generated structured insights from 400K+ customer comments by gathering and analyzing feedback across 5 social platforms using **Natural Language Processing techniques** and AI-based sentiment tools.
- Boosted sentiment classification accuracy to 79.3% by designing a hybrid framework combining **VADER and BERT** in a weighted ensemble model.
- Spearheaded the development of an AI-powered chatbot using **Chainlit, Gemini Pro 1.5**, to facilitate access to real-time insights, integrating an interactive Power BI dashboard.

Data Analyst Intern, Genmab Inc., Princeton, NJ

June 2024 - December 2024

- Reduced clinical trial recruitment timelines by 15% by designing and deploying an AI-driven heatmap interface using **Tableau, dbt**, and participant geolocation data.
- Optimized ETL transformation efficiency by 25% by building automated **DataOps pipelines using SQL in Snowflake and dbt** for real-time data modeling and integration.
- Upheld **HIPAA compliance** and data privacy by implementing secure access protocols and anonymization techniques, maintaining regulatory integrity throughout the clinical data lifecycle.

Data Engineer, TATA Consultancy Services, Bangalore, India

May 2022 - August 2023

- Reduced ETL processing time by building and streamlining scalable data pipelines for 500K+ records using **AWS S3 and Snowflake**, ensuring seamless integration into downstream systems.
- Refined data accuracy by 15% by implementing data validation rules and **anomaly detection logic using SQL** and Snowflake scripting, reducing error rates across multiple operational datasets.
- Accelerated data storage efficiency and query performance by 40% by engineering optimized schemas and partition strategies using **AWS Glue**, supporting faster analytics on large-scale data.
- Drove operational reporting adoption by designing interactive **AWS QuickSight dashboards**, translating complex KPIs into actionable insights for stakeholders in finance and retail domains.
- Supported **data governance and reporting optimization** for a leading financial services client by documenting data lineage and implementing quality checks to ensure compliance with enterprise standards and audit readiness.

PROJECTS

Bird Species Identification using Deep Learning [Link](#) → Built a **Keras-based CNN model** on spectrogram-transformed audio calls, achieving **95.45% binary and 67.24% multi-class accuracy**; using feature extraction and generalization.

Predicting Diabetes using Health and Socioeconomic Indicators [Link](#) → Led a team of 3 to build a **logistic regression classifier** on 250K CDC records, achieving a **90.6% F1-score**; enhanced model reliability through data cleaning, preprocessing, and statistical feature selection.

ACHIEVEMENTS & AWARDS

On the Spot Award: Recognized as the top performer on the team at TCS for *outstanding project delivery and impact*.

Publication: Influence of Wake Splitter Plate on Flow, Springer.