

VARUN VADDI

varunvaddi30600@gmail.com | (713) 539-6996 | Houston, TX | www.linkedin.com/in/varunvaddi/

WORK EXPERIENCE

Data Science Intern, Wild Genomics, CA

May 2025 - Aug 2025

- Built an end-to-end **Python** preprocessing pipeline for high-throughput sequencing data (demultiplexing, trimming, filtering, clustering), increasing analysis-ready data from 30% to 80% and enabling consistent, scalable downstream analytics workflows.
- Designed similarity-based **clustering** pipelines for sequences at 97-99% identity, reducing redundant data by ~90% and minimizing storage and computation costs across multi-sample datasets.
- Developed threshold-based **classification** workflows (BLAST, 95% identity) with automated reporting, improving accuracy from 85% to 95% and generating structured, high-quality labeled datasets for downstream analyses.
- Automated **batch processing** of multiple sequencing datasets using Python with logging, monitoring, validation, and filtering, reducing runtime by 40% and supporting rapid retraining of analytical pipelines.

Software Engineer, Accenture, Hyderabad, India

Sep 2021 - Dec 2023

- Optimized **Salesforce** data systems managing 80M+ records by refactoring 30+ Apex classes and process builders, improving workflow efficiency and deployment speed by 30% for large-scale data operations.
- Led large-scale data migrations and built interactive **CRM Analytics** dashboards using SOQL reports and analytics tools, analyzing conversion metrics and sales funnels to improve strategy, driving \$25M in revenue growth.
- Defined core performance metrics and built KPI dashboards using Salesforce **reports** to track feature impact and closed-won deals, enabling data-driven decisions and driving a 10% lift in conversion.

Data Engineer, University of Houston, TX

Jul 2024 - Dec 2025

- Automated **ETL** pipelines for admissions data (test scores, applications, transcripts) from multiple sources into **PeopleSoft**, ensuring accurate, clean, and timely data for enrollment and operational analytics.

PROJECTS

Customer Review Analytics

- Engineered 45+ **NLP**-derived features from 42K+ customer reviews, processing and transforming raw text into structured datasets suitable for downstream analytics and trend analysis.
- Built interactive **Tableau** dashboards revealing location preference patterns (nearest park bias), sentiment trends, and relative footfall proxies, enabling comparison of performance and improvement across Disneyland parks.

NYC Taxi Data Pipeline

- Eliminated 4-6-hour manual processing bottleneck by architecting automated **Snowflake + dbt + Dockerized Airflow** ETL pipeline processing 2M+ NYC taxi records with 37+ validation checks, reducing processing time by 95% and data quality incidents by 90%.
- Enabled data-driven optimization strategies by building dbt **analytics** models that identified 5-6 PM peak demand (30% higher volume), 15% tip premium on credit card payments, and high-traffic route corridors for dynamic pricing and fleet positioning decisions.

Sentiment Analysis

- Built data preprocessing and feature engineering pipelines for LSTM/GRU models on 70K+ reviews, standardizing raw text into embeddings and structured inputs for efficient model training.
- Conducted cross-dataset evaluation and ablation studies to streamline pipeline efficiency, reducing data preparation and model selection time by 20%.

EDUCATION

University of Houston

Master of Science (MS), Engineering Data Science | GPA: 3.7/4

TX, USA

Dec 2025

SKILLS

Programming Languages: Python, SQL, Java, R, MATLAB, MySQL, Oracle

Tools/Frameworks: Advanced Excel, Tableau, Power BI, Snowflake, dbt, Airflow, Docker, Git/GitHub, Kubernetes, Agile, CI/CD, Pandas, NumPy.

Competencies: Machine Learning, Deep Learning, Natural Language Processing, Generative AI, Transformers, LLM, RAG.