

# VARUN VADDI

[varunvaddi30600@gmail.com](mailto:varunvaddi30600@gmail.com) | (713) 539-6996 | [www.linkedin.com/in/varunvaddi/](https://www.linkedin.com/in/varunvaddi/) | <https://varunvaddi.github.io/>

## SUMMARY

Data Science graduate with 3+ years building end-to-end ML solutions from pipelines to production. Proven track record partnering with stakeholders to deliver predictive models driving measurable impact: \$25M revenue growth and 70% process automation.

## WORK EXPERIENCE

### *Data Science Intern, Wild Genomics, CA*

May 2025 - Aug 2025

- Partnered with research scientists to address low-quality sequencing data blocking critical insights, designed automated **Python** pipeline with filtering & QC that increased **analysis-ready data from 30% to 80%**, enabling accurate predictive modeling for conservation decisions.
- Reduced **data redundancy by ~90%** through similarity-based clustering (97-99% identity), improving feature reliability for downstream ML models and accelerating time-to-insight for cross-sample biodiversity analysis.
- Enhanced species **classification accuracy from 85% to 95%** using optimized BLAST similarity thresholds, enabling genus-level precision that supported grant funding decisions and improved research outcome predictions.

### *Data Engineer, University of Houston, TX*

Jul 2024 - Dec 2025

- Partnered with admissions leadership to develop predictive model **analyzing 40K+ applications per cycle**, identifying high school GPA and test scores as key graduation predictors, translated findings into automated admission thresholds that reduced **manual processing by 70%**, accelerated **decisions by 5 days**, and improved predicted **graduate success rates by 12%**.
- Built automated data pipelines integrating multiple admissions systems and designed Tableau dashboards enabling officers to **prioritize high-risk applications**, improving decision consistency & supporting evidence-based policy changes.

### *Software Engineer, Accenture, Hyderabad, India*

Sep 2021 - Dec 2023

- Partnered with sales leadership to optimize customer data systems **managing 80M+ records**, streamlined analytics workflows & improved **deployment speed by 30%**, enabling faster identification of high-value leads and sales strategies.
- Led CRM data migration projects and developed SQL-based analytics pipelines that identified **conversion bottlenecks and optimized sales funnels**, directly **contributing to \$25M in revenue growth** through improved lead prioritization.
- Defined KPIs with business stakeholders and automated Tableau dashboards monitoring feature adoption and deal conversions, delivering actionable insights that **increased conversion rates by 10%** and informed product roadmap decisions.

## PROJECTS

### **NYC Taxi Data Pipeline**

- Built end-to-end analytics pipeline (Snowflake, dbt, Airflow, Kafka) addressing slow feature preparation for **8.6M+ taxi trips**, reduced **prep time by 95%** with 37 automated validations, enabling real-time demand forecasting to **optimize driver allocation** and reduce customer wait times.
- Analyzed **\$180M+ quarterly revenue** patterns using SQL and Tableau, uncovering **30% peak-hour demand lift** and **18% credit-card revenue premium** insights assisted in informed business strategies for **dynamic pricing optimization**.

### **Google Ad Policy Compliance with RAG**

- Developed hybrid RAG system (BM25, embeddings, FAISS) to address 5-10 min manual ad review delays, **reduced review latency to <1 sec** while achieving **80% accuracy and 78% Recall@5**, enabling scalable **compliance for 1,000+ ads/sec** and reducing costs.
- Implemented structured LLM outputs with confidence scoring to eliminate inconsistent policy decisions, achieving **95% policy coverage** and enabling real-time risk detection that **reduced compliance violations by 40%**.

### **Named Entity Recognition on X posts**

- Designed BERT-based NER pipeline **processing 11,000 X posts** to address inconsistent entity extraction, improved **macro F1 to 89%** across 7 entity types, enabling marketing teams to identify emerging trends **3x faster to optimize campaigns**.
- Optimized sub-word label reconciliation **reducing errors by 12%**, supporting scalable social listening for brand monitoring and competitive intelligence across 100K+ daily posts.

## EDUCATION

### **University of Houston**

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

**TX, USA**

Dec 2025

## SKILLS

**Programming Languages:** Python, SQL, R, Linux.

**Tools/Frameworks:** Tableau, Power BI, Excel, PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face Transformers, NumPy, Pandas, Seaborn, Matplotlib, Plotly, Snowflake, dbt, Airflow, Docker, Kafka, Kubernetes, Spark, Git, Agile, CI/CD, AWS, Azure.

**Competencies:** Machine Learning, Deep Learning, NLP (Gen AI, LLM, RAG), Predictive Modeling, Statistical Analysis, Feature Engineering, Forecasting, Anomaly Detection, A/B Testing.