Naga Sai Sindhura Pandrangi
Fremont, CA | (669)-260-0662 | sindhura.naga98@gmail.com | linkedin.com/in/sindhurapandrangi

EDUCATION

San Jose State University **Jan. 2024 – Dec. 2025**
Master of Science in Data Analytics **GPA: 3.8/4.0**
Relevant Coursework: Machine Learning, Data Visualization, Statistical Analysis, Data Warehousing, Big Data Analytics

JNTUH, India **May. 2015 – May. 2019**
Bachelor of Technology, Electronics & Communication Engineering **GPA: 3.6/4.0**

EXPERIENCE

Teaching Assistant – Data Mining | San Jose State University **Aug. 2025 – Dec 2025**

- Mentored 60+ graduate students in machine learning, Python, and statistics, providing hands-on guidance in scikit-learn and data mining, and engineered automated grading pipelines that streamlined assessments and elevated feedback quality.

Data Analyst | S&P Global **Aug. 2022 – Aug. 2023**

- Extracted, transformed, and loaded large-scale structured and unstructured data into centralized data warehouses, improving data quality by 25% and enabling faster KPI-driven business intelligence.
- Performed data validation and cleansing using advanced SQL, increasing data integrity to 99.5%, reducing report errors in live dashboards, and optimizing complex SQL queries.
- Built 10+ interactive Power BI dashboards with DAX for KPI visualization, improving reporting efficiency by 30%, enabling real-time monitoring, and accelerating project delivery by 20%.

Data Engineer | Tata Consultancy Services **June. 2019 – Jul. 2022**

- Constructed and maintained scalable ETL pipelines with Apache Airflow and PySpark on GCP, managing ingestion from multiple data sources into BigQuery and GCP Buckets for optimized storage and high-volume processing.
- Refactored workflows for parallel processing, reducing pipeline execution time by 40%, improving data accuracy with advanced SQL transformations, and enabling near real-time healthcare data ingestion via Spark Streaming.
- Introduced monitoring and alerting within Airflow pipelines, reducing incident response time by 50%, while partnering with analysts to fine-tune data warehouses for seamless Power BI dashboard integration and KPI tracking.

PROJECTS GITHUB : github.com/NagaSaiSindhura

Smart Book Recommendation System |Apache Hadoop, Pyspark, Plotly **Aug. 2024 – Dec 2024**

- Developed a hybrid book recommendation engine using ALS matrix factorization and TF-IDF for collaborative and content-based filtering, improving recommendation relevance by 12%.
- Utilized PySpark for real-time data processing, created Plotly dashboards for user behavior insights, and deployed the system via a Flask web app for live recommendations.

Automated ETL & KPI Dashboard on GCP | Apache Airflow, Star Schema, BigQuery, Power BI **Aug. 2024 – Dec. 2024**

- Engineered an ETL pipeline with Apache Airflow and dbt tool to process 100K+ daily records from the New York Times dataset, streamlining ingestion, transformation, and loading into the data warehouse.
- Published Power BI dashboards connected to BigQuery deploying a star schema model, ensuring 95% on-time data delivery.

California Road Safety Dashboard | MYSQL, Microsoft PowerBI, Power Query, DAX **Jan. 2024 – May 2024**

- Performed end-to-end analysis of traffic accident data using MySQL and Power Query for data extraction and transformation, reducing preprocessing time by 25%.
- Crafted a Power BI dashboard with DAX-based KPIs, drill-throughs, and slicers by time, location, and severity. Also leveraged Power Pivot to model data and define relationships for accurate, responsive reporting.

Full-Stack Agentic Framework for NL to SQL Translation |Apache Hadoop, Pyspark, Plotly **Aug. 2024 – Dec 2024**

- Developed a Streamlit-based Natural Language to SQL platform using Hugging Face LLM fine-tuned on the Spider dataset, boosting BLEU score from 0.38 to 0.52 and enabling schema exploration, query generation, and interactive Plotly visual analytics.
- Optimized an 8B-parameter model for efficient CPU inference, cutting costs and latency, while building an intelligent query refinement system that improved SQL accuracy and delivered a smoother user experience.

Deepfake Video Detection | XceptionNet, MTCNN, PyTorch **Jan. 2025– May. 2025**

- Built a detection system leveraging XceptionNet, ResNet-Swish-BiLSTM, and Swiss Residual Blocks, attaining 97.01% accuracy and 0.988 ROC-AUC on Meta DFDC.

- Applied Focal Loss to manage class imbalance and used MTCNN for precise face localization and Accelerated training with PyTorch mixed precision, reducing compute time by 35% and GPU usage by 40%.

AI Mock Interviewer for Data Analyst/Data Scientist | LLM's , Fine Tuning, RAG , Docker, GCP **Jan. 2025 -Dec 2025**

- Developed Interview Prep, an AI-powered mock interview platform using four fine-tuned LLMs on 4,500+ Q&A pairs; integrated RAG for personalized question generation from resumes/job descriptions and real-time feedback, evaluated via BLEU, ROUGE, BERTScore, accuracy, and hallucination metrics**.**
- Engineered scalable full-stack system with React frontend (Vercel), FastAPI backend (Hugging Face/Docker), and GCP data pipelines; optimized with quantization for ~40% reduced memory and low-latency inference, delivering intuitive UI and performance dashboards for data analytics/science interview preparations.

TECHNICAL SKILLS

- **Languages & Tools:** Python (Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn), HTML, CSS, React, SQL (Microsoft SQL Server, MySQL), MongoDB, Spark, Git, Flask
- **Analytics & ML:** Regression, Classification, Clustering, Time Series, LSTM, A/B Testing, Hypothesis Testing, Statistical Inference
- **Cloud & Data:** AWS (S3, Redshift, Athena, Glue), Google Cloud Platform (BigQuery), Snowflake, Apache Airflow, DBT, Kafka, Hadoop, ETL Pipelines
- **Visualization & BI:** Tableau, Power BI, Looker Studio, Plotly (Dash), Excel