HRUSHIKESH THANIKONDA
Jersey City, NJ | +1 (551)-331-9712 | thanikondahrushikesh@gmail.com | LinkedIn | GitHub

Data Engineer with experience in financial time-series analysis, ETL pipeline development, and cloud-based data platforms. Skilled in building reliable, auditable data systems using Python, SQL, and Azure to support analytics and forecasting workflows.

EDUCATION

Stevens Institute of Technology, Hoboken, NJ
Master of Science in Data Science (Sept 2024 – Present)
Blekinge Institute of Technology, Sweden
Bachelor of Science in Computer Science and Engineering (May 2021 – Aug 2023)

PROJECT EXPEREINCE

Forecasting & Market Trend Analysis (Sept 2025 – Dec 2025)

- Analyzed BlackRock equity price time-series using ARMA-based models to support short-term market trend evaluation and liquidity related forecasting , selecting parameters via ACF/PACF diagnostics.
- Validated financial forecasts using holdout testing and confidence intervals improving reliability of time series predictors for decision-support analytics.
- Modeled long-term macroeconomic time-series (California unemployment, 1976–2025) using SARIMA to capture seasonality and cyclical signals relevant to market and liquidity conditions.
- Applied residual diagnostics and statistical validation to ensure robustness, data integrity, and controlled use in forecasting workflows.

Cloud-Based Data Ingestion & Storage Platform **(**Sept 2025 – Dec2025)

- Designed and implemented a cloud-based data ingestion platform supporting user authentication, role-based access control (user, approver, admin), and governed data access.
- Built backend services using .NET, persisting structured authentication and metadata in Azure SQL with defined schemas and constraints.
- Ingested and stored unstructured image data in Azure Cosmos DB, separating storage layers for binaries and relational metadata.
- Implemented human-in-the-loop approval workflows, ensuring data validity prior to downstream access and consumption.
- Tracked application-generated metadata (image views, access events) in Azure Storage for auditing and analytics.
- Secured secrets and credentials using Azure Key Vault, eliminating hard-coded keys in pipeline logic.
- Implemented data validation and failure-detection logic to ensure reliable data ingestion and storage.

Experimental Data Pipeline – Action Generation (EPIC-Kitchens) **(**Sept 2025 – Dec2025)

- Engineered a video-to-image data pipeline, converting raw video into timestamp-aligned image frames using label metadata.
- Designed structured datasets using stacked image pairs (t, t+Δt) as model inputs.
- Proposed and implemented data augmentation (image flipping) to mitigate bias and increase dataset size (40 → 80 samples).
- Implemented preprocessing and filtering logic using MediaPipe to reduce noise and ensure data consistency.
- Improved data governance, access controls, and auditability, supporting risk-aware analytics and reviewable reporting workflows.

Insurance Data Engineering & Reporting Pipeline (Jan 2025 – May 2025)

- Built Python- and SQL-based ETL pipelines to process,validate and transform structured datasets improving data reliability and enabling downstream analytics.
- Defined schema design, validation rules, and data quality controls to ensure consistency and accuracy across pipelines.
- Implemented data quality checks (null handling, outlier detection, distribution shifts) to detect pipeline issues early.
- Developed downstream datasets consumed by reporting and modeling layers.
- Produced Power BI dashboards to surface trends, anomalies, and performance metrics, supporting stakeholder decision-making.

Performance of different ML models - Cyberbullying Dataset (Sept 2024 – Dec 2024)

- Annotated and reviewed 40,000+ text samples, applying consistent labeling rules for cyberbullying sentiment classification.
- Resolved ambiguous and mislabeled samples through iterative human-in-the-loop review cycles, improving label consistency and reducing bias.
- Applied structured preprocessing and validation techniques, contributing to 84% classification accuracy.
- Documented annotation decisions and edge cases to ensure guideline adherence and reproducibility.

Research Thesis – Sentiment Analysis of Cyberbullying Tweets **(**May 2021 – Sept 2023)

- Developed an SVM-based sentiment classifier achieving 84% accuracy across 40,000+ labeled samples.
- Published findings and presented analytical insights in academic research settings.

TECHNICAL SKILLS

Data Engineering & Pipelines: Python, SQL, PostgreSQL, data ingestion, data transformation, ETL/ELT workflows, schema design, data validation, data quality checks
Cloud & Storage: Azure SQL, Cosmos DB, Azure Storage, Key Vault, secure access control
Analytics & Processing: Time-series data processing, regression pipelines, forecasting validation
Engineering Practices: Git, GitHub, Command Line (CLI), pipeline reliability, documentation

CERTIFICATIONS

- Microsoft Power BI Data Analyst Professional Certificate
- Meta Database Engineer Professional Certificate