Rishi Guptha Mankala

New York | [LinkedIn](#) | [Portfolio](#) | 934-263-3087 | rishiguptha.mankala@gmail.com

WORK EXPERIENCE

Stony Brook University   **Stony Brook, NY**
Data Science Researcher   *Jan 2025 - Present*
- Built Python-SQL pipelines for S&P 500 time-series data and trained LSTM, CNN, and SVM models using TensorFlow/PyTorch, improving model accuracy by 20% and risk scoring by 15% through Bayesian inference and backtesting.
- Delivered dashboards (Tableau, Python) integrating predictive outputs and risk metrics, enabling 30% faster portfolio decisions and surfacing volatility trends via SVM optimization and feature extraction.

Kimberly Clark   **Bengaluru, India**
Data Engineering Intern   *Jan 2023 - July 2024*
- Automated metadata tagging and validation in Collibra using Python/PostgreSQL workflows, earning 7 certifications and boosting governance team productivity by 25% while reducing manual effort by 30%.
- Migrated 10K+ assets into Collibra via Excel-based ETL with 100% data integrity, resolved 15% pipeline inconsistencies, and built Power BI/Tableau dashboards to visualize KPIs and reduce reporting errors by 18%.
- Designed and deployed 5+ advanced Collibra workflows to support real-time data quality checks and lineage validation, enabling faster stakeholder reporting and improved oversight across governance domains.

Matchday AI   **Hyderabad, India**
Data Engineering Intern   *Jan 2022 - Dec 2022*
- Built computer vision pipelines with OpenCV and deep learning to detect field lines, ball position, and player movement from broadcast footage, achieving 90%+ event tracking accuracy.
- Applied Homography Transformation to convert dynamic camera feeds into 2D field projections, powering spatial analytics for tactical coaching insights.
- Developed a full-stack annotation platform (Flask, JS, SQLite), reducing manual labeling time by 40% and enabling deployment of data-driven dashboards for 6+ ISL teams via Star Sports.

PROJECTS

NutriSmart: RAG-Driven Insights for Fitness and Nutrition
- Built a RAG pipeline using Sentence Transformers + FAISS to semantically index 1,200+ academic pages; containerized with Docker and deployed via FastAPI on AWS EC2, enabling low-latency fitness insights and boosting retrieval accuracy by 30%.

NYC Taxi Insights
- Engineered a GCP-based ETL pipeline for 10M+ ride records, modeled data in BigQuery using a star schema, and built Looker Studio dashboards; trained LightGBM models with BERT features for fare and duration prediction, boosting model efficiency by 30% and enabling real-time pricing strategies.

Sales Insights Dashboard
- Built a centralized reporting system in Power BI using SQL and DAX to integrate CRM/ERP data, applying time-series modeling and geospatial analytics to boost forecast accuracy by 30% and reduce sales cycle prediction time by 15%.

Spotify Playlist Analytics
- Deployed a serverless ETL pipeline on AWS (Lambda, Glue, S3) for 18K+ playlist logs, optimizing SQL performance in Athena to reduce processing time by 40% and enable daily insights into genre trends and user engagement.

EDUCATION

Stony Brook University   **Stony Brook, NY**
Masters of Science, Data Science

SKILLS

Programming Languages: Python, R, SQL, Java

Data Science & Machine Learning: TensorFlow, PyTorch, Scikit-Learn, Pandas, NumPy, Statistical Analysis, Statistical Modeling, Deep Learning, Large Language Models

Cloud Platforms: AWS (S3, EC2, Glue, Redshift, Lambda), Google Cloud (BigQuery, Cloud Storage, VertexAI)

Big Data Technologies/Frameworks: Snowflake, Airflow, Apache Spark, DBT, ,Data Modelling

Data Visualization: Power BI, Tableau, Looker Studio, AWS Quick Sight, Matplotlib

Tools: Linux, Docker, Astronomer, Git, GitHub, Microsoft Excel, Google Sheets, Collibra