

Mayur Kumar

Data Scientist

www.linkedin.com/in/mayur-kumar-sourav | candidates.intervues.club/u/mayur-kumar | github.com/mayurks23 | candidates.intervues.club/u/mayur-kumar

SUMMARY

Data Scientist with a background in data modeling, statistical analysis, and machine learning pipeline development. Experienced in executing end-to-end ETL processes, exploratory data analysis, and deploying cloud services using AWS EC2 and Flask. Holds an M.Sc. and B.Sc. in Data Science and Artificial Intelligence.

EXPERIENCE

Data Science Intern

Innomatics Research Labs

Jan 2024 – Mar 2024

- Collected and cleaned operational data across multiple sources, building end-to-end ETL processes to stage analysis-ready datasets.
- Executed statistical hypothesis testing and exploratory data analysis on 10,000+ structured records using Python (Pandas, NumPy, SciPy) and SQL to validate operational assumptions.
- Built and deployed a Python and Flask-based regex matching web application with REST API integration on AWS EC2, implementing runtime monitoring to track service performance.
- Refactored backend application workflows to enhance code structure, technical documentation, and overall reliability.
- Delivered data-driven analytical insights to cross-functional teams to support delivery-performance and revenue-management strategy.

Data Science Intern

Innomatics Research Labs

Jan 2024 – Mar 2024

- Collected and cleaned operational data across multiple sources, building end-to-end ETL processes to stage analysis-ready datasets.
- Executed statistical hypothesis testing and exploratory data analysis on 10,000+ structured records using Python (Pandas, NumPy, SciPy) and SQL to validate operational assumptions.
- Built and deployed a Python and Flask-based regex matching web application with REST API integration on AWS EC2, implementing runtime monitoring to track service performance.
- Refactored backend application workflows to enhance code structure, technical documentation, and overall reliability.
- Delivered data-driven analytical insights to cross-functional teams to support delivery-performance and revenue-management strategy.

EDUCATION

Vellore Institute of Technology · M.Sc., Master of Science - MS · 2024 – 2026

Christ (Deemed to be) University · B.Sc., Bachelor of Data Science and Artificial Intelligence honours · 2021 – 2024

SKILLS

Languages and Frameworks: Bash, HTML, JavaScript, Python, R, SQL, Flask, Keras, NumPy, Pandas, PyTorch, scikit-learn, SciPy, TensorFlow

Databases: Database Management System (DBMS), MongoDB, MySQL, PostgreSQL

Tools: Amazon EC2, Amazon S3, Apache Airflow, ARIMA, Astro CLI, AWS EC2, CNN, Code Refactoring, Data Modeling, Data Pipelines, Debugging, Docker, FinBERT, Git, GNN / GraphSAGE, Google Colab, Jupyter, LLM, LSTM, Matplotlib, Microsoft Excel, Microsoft Power BI, Mistral-7B, MLflow, Pipeline Engineering, Prefect, REST API, Seaborn, Tableau, Transformers, VS Code

PROJECTS

GNN + LLM Integration for Stock Portfolio Optimization

Designed and implemented a GraphSAGE-based GNN using PyTorch to model relationships between Nifty-50 stocks using financial and textual features. Integrated FinBERT for sentiment. Modeled Nifty-50 stock relationships in PyTorch using a GraphSAGE architecture combining financial metrics and text features. Integrated FinBERT for financial sentiment scoring to inform stock portfolio optimization. GraphSAGE, GNN, PyTorch, FinBERT, Mistral-7B, LLM, NLP

Sentiment Classification with Deep Learning

Processed 8,518 unstructured text records and built LSTM and CNN-based classifiers using PyTorch and TensorFlow. Applied NLP preprocessing including tokenization, vectorization, f... Processed 8,518 unstructured text records using NLP preprocessing routines including tokenization and vectorization. Constructed and trained LSTM and CNN classification models using PyTorch and TensorFlow. LSTM, CNN, PyTorch, TensorFlow, NLP

Weather ETL Pipeline

Built an automated ETL pipeline for data extraction, transformation, and loading into PostgreSQL using Apache Airflow DAGs. Containerized services using Docker for reproducible de... Automated extraction, transformation, and loading of weather data into PostgreSQL using Apache Airflow DAGs. Containerized core application services with Docker to ensure reproducible local and cloud deployments. Apache Airflow, Docker, PostgreSQL

Additional projects: Civil Aviation Time Series Analysis, SLA & Revenue Impact Analysis – Pizza Delivery Operations, SENTIMENT ANALYSIS FOR PRODUCT REVIEWS, Life Expectancy Prediction Using Machine Learning, Weather ETL Pipeline using Apache Airflow, SLA & Revenue Impact Analysis – Pizza Delivery Operations, SENTIMENT ANALYSIS FOR PRODUCT REVIEWS, Life Expectancy Prediction Using Machine Learning, Weather ETL Pipeline using Apache Airflow, GNN + LLM Portfolio Optimization, Sentiment Analysis with Deep Learning and MLOps, Time Series Forecasting for the Aviation Industry, Retail Sales Analytics Dashboard, GNN + LLM Integration for Stock Portfolio Optimization, Sentiment Classification with Deep Learning, Weather ETL Pipeline, Civil Aviation Time Series Analysis, SLA & Revenue Impact Analysis – Pizza Delivery Operations, SENTIMENT ANALYSIS FOR PRODUCT REVIEWS, Life Expectancy Prediction Using Machine Learning, Weather ETL Pipeline using Apache Airflow, Backend Refactoring and Bug Fixing for Note Taking Application Development and AWS Deployment of Regex Matching Web Application with Flask

AWARDS

2nd Rank - Fanalytics 2022, Python For Everybody Certification Coursera Certification, Python For Everybody Certification, Python For Everybody, Second Rank, Fanalytics 2022