

MAYURESH PRAMOD PANDEY

San Jose, CA | mayurp.pandey@gmail.com | +1 (669) 340-6006 | [linkedin.com/in/mayureshpp](https://www.linkedin.com/in/mayureshpp) | github.com/mayu99

EDUCATION

Master of Science, Data Analytics

May 2026

San Jose State University, CA, USA

Courses – Mathematics, Big Data, Data Visualization, Machine Learning, Deep Learning, Data Warehouse and Pipeline, Distributed Systems for Data Engineering, Generative AI Applications.

Bachelor of Engineering, Information Technology

June 2021

University of Mumbai, Maharashtra, India

Courses – Data Structure and Algorithm, DBMS, Data Mining, Big Data, Artificial Intelligence, Cloud Computing.

RELEVANT EXPERIENCE

Data Science & Engineering Intern | Schneider National, WI

May 2025 - Aug 2025

(Tech Stack: - Python, PySpark, Spark SQL, Prophet, Kalman Filtering, Azure, Oracle, Snowflake, Hive.)

- Rebuilt a **Bid Optimization Framework** from scratch by engineering **ETL pipelines** in **Azure Databricks** (Oracle, Snowflake, Hive) to process **2B+ records**, automating cleaning, deduplication, mapping, reducing data prep time by **40%**.
- Designed a **Cost Forecast Model** using **Prophet** with hierarchical forecasting and Kalman filtering, generating 365-day carrier cost predictions across lanes, submarkets, and improving cost accuracy by **15%** over legacy methods.
- Built a **Bid Acceptance Model** with **LASSO logistic regression** and 10+ engineered features, including log-distance, historical win behavior, and rate-per-mile sensitivity modeled through sigmoid win-rate curves.
- Delivered a **30% boost in prediction accuracy** and powered a Contract Price Optimization engine that raised margin recommendations by **8–12%** without reducing competitive win rates.

Data Scientist | Mu Sigma, Bangalore, India

Aug 2021 - June 2024

(Tech Stack: - PySpark, Hive, Azure Data Lake, Azure Data Bricks, Azure Blob, Azure Key Vaults, Azure DevOps.)

- Developed **automated forecasting tool** using ML/statistical models (ARIMA, ARIMAX, SMA, Prophet, Holt-Winters), improving **forecast accuracy by 15–20%** and streamlining supply-demand planning across Spain, Portugal, and Poland.
- Designed and deployed scalable **ETL pipelines** and data assets using **PySpark, Hive, and Azure** (Data Lake, Data Factory, Databricks, Blob, Key Vaults, DevOps).
- Automated data workflows and enabled app integration via React, improving efficiency in order and shipment forecasting, enabling **end-to-end automation** and reducing **manual processing time by 30%**.

PROJECTS

[Precision Banking: Customer Response Prediction](#)

- Developed and tuned ML models (KNN, Random Forest, SVM, Naïve Bayes) on the Bank Marketing dataset, achieving **94% accuracy and ROC-AUC 0.94** through SMOTE balancing, EDA, and hyperparameter tuning.
- Deployed the best-performing **Random Forest model via Streamlit web app**, enabling real-time customer response prediction that improved campaign targeting, reduced unnecessary outreach, and increased marketing ROI.

[MyNewsMate: AI-Powered Personalized News Recommendation System](#)

- Developed a **scalable distributed system** (Django REST, ReactJS, PostgreSQL, Docker, AWS) for real-time personalized news feeds.
- Engineered a **hybrid NLP-driven recommender** (TF-IDF, cosine similarity, Hugging Face BART, VADER) with Celery + Redis async processing, boosting engagement **30%**, CTR **40%**, and reducing bounce rate **25%**.

[End-to-End E-Commerce Data Pipeline on Azure](#)

- Designed and built an **end-to-end e-commerce data pipeline** on **Azure** using Data Factory, Data Lake Gen2, and Databricks (PySpark) to ingest data from MySQL and MongoDB.
- Cleaned and transformed data using the **Medallion Architecture (Bronze–Silver–Gold)** to ensure data quality.
- Loaded curated data into Azure Synapse Analytics and created **Tableau dashboards** for key business KPIs, improving performance and scalability using **partitioning, caching, and Parquet compression**.

SKILLS

Programming Languages: Python, R, JavaScript, C, C++, SQL.

Databases: MongoDB, MySQL, PostgreSQL, Hadoop, Spark SQL, Hive, Neo4j.

Cloud & DevOps: Amazon AWS, Microsoft Azure, Snowflake, Jenkins, Docker, Kubernetes.

Frameworks: TensorFlow, PyTorch, Keras, Scikit-Learn, PySpark, FastAPI, Airflow.

AI & Generative AI Tools: Azure OpenAI, OpenAI API (GPT-4), Hugging Face Transformers, LangChain, FAISS

Other: Git, Tableau, PowerBI, Excel, Sphinx, Machine Learning, Deep Learning.

PUBLICATIONS AND CERTIFICATIONS

- Mental Health Prediction for Juveniles using ML – Elsevier Conference, 2021 (SSRN: 3867291)
- Smart Emergency Vehicle Detection using IoT & ML – IEEE INAC-4, 2019
- Identity Resolution in Social Networks using Recommender Systems – Journal of AI & Systems, 2019 (ISSN: 2642-2859).
- **AWS Certified Solutions Architect – Associate** ([Credential](#))