Aishwarya Gulab Thorat
+1 (408) 581-4992 · aishwaryathorat723@gmail.com · LinkedIn · GitHub · San Jose, CA

Education

Master of Science in Data Analytics, *San Jose State University, San Jose, CA* **Jan 2024 - Present**
- **Relevant Coursework:** Database Systems for Data Analytics, Big Data Technologies, Machine Learning, Business Intelligence, Distributed Systems GPA 3.66

Bachelor of Engineering in Information Technology, *University of Mumbai - Mumbai, India* **May 2019**
- **Relevant Coursework:** Business Intelligence, Design & Analysis of Algorithms, Database Management Systems, Software Engineering GPA 4

Work Experience

Magic9 Media & Analytics Pvt. Ltd **Mumbai, India**
Senior Data Scientist – AI/ML *Sep 2021 – Jan 2024*
- Built predictive models improving trend detection accuracy by 25% using **Python** (Pandas, NumPy, Scikit-learn) and **MATLAB**.
- Built interactive dashboards using **Tableau** and **Power BI** providing real-time insights to 500+ business users.
- Developed an ML-based fraud detection algorithms using **supervised** and **unsupervised ML** techniques.
- Automated ETL data pipelines using **Apache Airflow**, integrating **Amazon S3** and **SQL** Server for real-time analytics.
- Developed complex **SQL** queries and stored procedures to analyze viewership data and improve channel ratings.
- Improved trend forecasting and optimized media strategies using **A/B testing** and **statistical analysis**.

Mondelēz International **Mumbai, India**
Application Developer *Jan 2021 – Sep 2021*
- Designed a **Quickbase** and **Power Apps** solution to streamline reimbursement approvals reducing processing time by 40%.
- Developed an **ETL** pipeline for processing 100K+ records daily.
- Built interactive **Tableau** and **Power BI** dashboards to visualize sales performance enabling data-driven decision-making.

Sciative Solutions – Pricing Analytics **Mumbai, India**
Junior Data Scientist *Jun 2019 – Jan 2021*
- Built **regression** models for appliance price prediction optimizing pricing strategies and increasing sales.
- Developed **logistic regression** models to predict repeat purchases improving customer targeting by 15%.
- Implemented **PySpark** jobs for large-scale data processing in **Azure Databricks**.
- Analyzed large-scale product sales data using **Advanced Excel** and **SQL** to identify trends and anomalies.

Technical Skills

Programming Languages: Python, R, SQL, MATLAB, SAS
Data Engineering & Tools: DBT, Airflow, Spark, Lucidchart, DBT Lineage, Draw.io, familiar with ERWin and Visio for data modeling
ML & AI: TensorFlow, PyTorch, Scikit-learn, LLMs, GenAI, XGBoost, NLP, Supervised & Unsupervised Learning, Time Series Forecasting
Databases: Oracle, MongoDB, Google Big Query, Amazon RedShift, Snowflake, Relational Databases, NoSQL, MySQL, PostgreSQL
Big Data & Cloud Technologies: Apache Spark, Hadoop, AWS (S3, Redshift, Glue), GCP (BigQuery), Databricks, MapReduce
Business Intelligence & Visualization: Tableau, Power BI, Looker, Salesforce Reports, Metabase, Quicksight

Academic Projects

ETL Data Pipeline for Online Retail Transactions
- Designed and implemented an ETL pipeline storing data in **GCP** buckets and automating workflows with **Apache Airflow**.
- Ensured data quality with **SODA**, transformed data using **DBT**, and visualized insights in **Metabase**.

Deep Learning-based Medical Image Analysis
- Built a **CNN** model for rib fracture classification using 3D medical imaging data.
- Integrated residual blocks, dropout, and batch normalization for enhanced performance.

Real-time Sentiment Analysis on Yelp Reviews
- Designed a scalable pipeline integrating **Kafka**, **Spark** Streaming, **AWS** S3, Glue, Athena, SageMaker and Quicksight.
- Implemented **ML** models (Logistic Regression, TF-IDF, BERT) achieving 79% classification accuracy.

Zero-shot vs Fine-tuned Models for Sentiment Analysis
- Compared **DistilBERT** and **DistilRoBERTa** against fine-tuned models on the IMDB dataset.
- Integrated **OpenAI API** models (GPT-4o-mini, GPT-3.5) for classification and benchmarking.

Ride Management and Price Prediction System
- Developed a full-stack ride-hailing platform using **Django (backend)** and **ReactJS (frontend)**, integrating an **ML-based ride price prediction model** to provide real-time price estimations via **Streamlit**.
- Containerized the application using **Docker** and deployed the backend on **Kubernetes**, ensuring efficient scaling and management.