Sudarshan Dudhe

Boston, MA

857-398-5971 dudhe.s@northeastern.edu

Professional Summary

Data Scientist and Data Engineer with experience in developing robust ETL pipelines, predictive models, and cloud-based data solutions. Proficient in Python, SQL, AWS, and Snowflake. Strong foundation in statistics, machine learning, and data visualization, with a focus on optimizing data flow and model accuracy.

Education

Northeastern University **Sep. 2023 – May 2025**

Master of Science, Information Systems *Boston, MA*

Yeshwantrao Chavan College of Engineering **Sep. 2012 – May 2017**

Bachelor of Technology, Information Technology *Nagpur, Maharashtra, India*

Technical Skills

Programming Languages: Python, R, SQL, NoSQL, PySpark

Cloud Technologies: AWS (S3, Redshift, EC2, Lambda), Snowflake

Data Visualization: Tableau, Power BI, Matplotlib, Seaborn, Plotly

Data Science: Predictive Analytics, Machine Learning, Data Quality, Statistics, Feature Engineering

Data Engineering: ETL Pipelines, Data Warehousing, Airflow, Data Ingestion

Tools: Scikit-Learn, TensorFlow, PyTorch, Git, Docker, Kubernetes

Professional Experience

Arges Global Limited **January 2021 – August 2023**

Developer *Mumbai, Maharashtra, India*

- Developed and maintained a web application for textile data management, reducing order processing time by 30% through real-time data visualization.
- Engineered scalable ETL pipelines to extract and transform 10GB of quarterly data from multiple sources (CSV, APIs), enhancing data accuracy and analysis speed by 20%.
- Automated reporting processes using Python and integrated machine learning models, improving decision-making efficiency for the operations team.
- **Key Achievements**:
  - Led a team of 5 to design and implement an interactive dashboard that improved decision-making and operational insights.
  - Reduced data retrieval time by 40% by optimizing SQL queries and improving database schema design.

Academic Projects

Home Price Prediction Project — Data Science Project **September 2023**

- Built a predictive model for Bangalore home prices using a 13,320-record dataset, achieving 86.29% accuracy with **Linear Regression** and **Scikit-Learn**.
- Conducted feature engineering to transform raw data, improving model accuracy by 10% after dimensionality reduction and outlier detection.
- Deployed a web app using **Flask** for real-time price prediction, integrated with an **HTML/CSS/JavaScript** front-end.
- Key Technologies: Python, Pandas, NumPy, Scikit-Learn, Flask, HTML/CSS/JavaScript

CFA Institute Web Scraping & Analytics — Data Engineering Project **November 2023**

- Led a web scraping initiative using **Python** (Selenium, BeautifulSoup) to extract data from 220+ web pages, storing it in **AWS S3** and Snowflake.
- Designed an efficient ETL pipeline with **Airflow** for data ingestion, transformation, and storage, resulting in a 25% increase in processing speed.
- Applied Pydantic and Pytest for schema validation, ensuring data quality before storing it in Snowflake.
- Key Technologies: Python, Selenium, BeautifulSoup, Airflow, AWS S3, Snowflake, Git