

---

**WORK EXPERIENCE****Data Engineer**

Accenture Private Solutions, Mumbai, IN

Sep 2021 - Jul 2023

- Developed and maintained scalable data pipelines using Python and Shell scripting to automate data ingestion, transformation, and processing for large datasets.
- Optimized ETL processes using Hive and Unix commands, reducing data processing time by 30% and improving data accessibility.
- Processed and integrated structured and unstructured data into AWS cloud storage, ensuring efficient data management and scalability.
- Collaborated with cross-functional teams to deliver high-quality data solutions, aligning data pipelines with business and analytics requirements.
- Monitored and maintained cloud-based data environments (AWS), ensuring data security, integrity, and compliance with industry standards.

---

**INTERNSHIP****Machine Learning Intern**

ZS Analytics LLC, New York, USA

May 2024 - Aug 2024

- Developed and optimized supervised and unsupervised machine learning models to address business challenges, utilizing Python, Scikit-learn, and TensorFlow.
- Collaborated with cross-functional teams, including data scientists and business analysts, to gather requirements and deliver insights through Exploratory Data Analysis (EDA) and model performance evaluation.
- Deployed machine learning models into production environments, integrating them into analytics pipelines with AWS and Docker for seamless scalability and performance monitoring.

---

**EDUCATION****Master of Science in Data Analytics & Visualization**

Yeshiva University, Katz School of Science and Health | New York, USA

Aug 2023 - Aug 2024

**Bachelor of Engineering in Information Technology**

University of Mumbai, Bharati Vidyapeeth College of Engineering | Mumbai, IN

Aug 2018 - July 2021

---

**ACADEMIC PROJECTS****Analyzing Chicago Crime Dataset Using Machine Learning Models**

Feb 2024 - May 2024

- Led a project to analyze the Chicago crime dataset from Google Big Query, uncovering trends and patterns.
- Built and evaluated machine learning models (Logistic Regression, Random Forest, XG Boost), and developed an ensemble model (AdaBoost with Decision Trees) to predict crime types.
- Conducted geospatial analysis using Matplotlib, Seaborn, Folium to visualize crime distribution, identify hotspots.
- Leveraged data-driven methods to enhance public safety and foster safer communities.

**Movie Analysis System Using AWS**

Mar 2024 - May 2024

- Developed a system on AWS to provide directors with insights on market trends and viewer preferences.
- Integrated and visualized data from Movie Lens and TMDb, offering actionable insights for decision making.
- Utilized AWS services (S3, Lambda, RDS, Glue), executed ETL processes, created visualizations using Tableau.

**Healthcare Management System**

Sep 2023 - Dec 2023

- Tracked disease prevalence, understanding characteristics and making informed decisions to improve healthcare management.
- Leveraged PostgreSQL, ETL processes, and SQL Workbench for data handling and modeling.
- Developed visualizations and dashboards using Power BI.

---

**TECHNICAL SKILLS**

- **Programming Languages:** Python, Shell Scripting, R, Java.
- **Cloud Services:** AWS (S3, Lambda, RDS, Glue, Sage maker, Redshift), Azure
- **Big Data Tools:** Apache Spark, Apache Hadoop, Apache Kafka
- **Databases:** SQL, NoSQL, PostgreSQL, Apache Hive, Snowflake
- **Data Analysis:** Machine Learning, Predictive Modeling, Data Mining, Statistical Analysis
- **Data Modelling:** Relational Data Modelling, Schema Design, ETL
- **Data Visualization:** Tableau, Power BI
- **Development Tools:** Git, GitHub, Jira, Jupyter Notebook, Visual Studio Code, Microsoft Excel.

---

**CERTIFICATIONS**

- Microsoft Certified: [Azure Data Engineer Associate](#) | July 2024
- Amazon Web Services Certified: [Cloud Practitioner](#) | April 2024