

KAJAL MAHAJAN

Brampton, Ontario, Canada • kajalmahjan456@gmail.com • (647) 446-4316

LinkedIn: [linkedin.com/in/kajalmahajan5](https://www.linkedin.com/in/kajalmahajan5) • GitHub: [kajalmahajan23](https://github.com/kajalmahajan23) • Website: kajalmahajan23.github.io

PROFESSIONAL EXPERIENCE

Bank of America

Data Engineer

Mumbai, Maharashtra, India

March 2021 – June 2024

- Developed and deployed a fraud detection machine learning(ML) model using Python (Random Forest, Gradient Boosting, Linear regression, SVD) to analyze transaction patterns, improving the bank's fraud detection accuracy by 22%, reducing high-risk transactions.
- Built scalable ETL pipelines using Azure Data Factory, processing large datasets from SQL Server, PostgreSQL, and AWS S3, enhancing data processing efficiency by 40%.
- Automated data ingestion and transformation using AWS Glue and Athena, improving data transfer speed between AWS S3 and Snowflake by 35%, boosting reporting capabilities.
- Deployed Python-based APIs for data integration with Snowflake, leveraging AWS Lambda and PySpark for large-scale data migrations, significantly improving data processing speed in Azure Data Lake.
- Visualized complex financial data trends using Power BI, Salesforce, Matplotlib, and Tableau, creating real-time dashboards to drive data-driven decisions for business stakeholders.
- Conducted sentiment analysis on customer feedback using Python's NLP libraries (NLTK), providing insights that increased customer satisfaction by 10%.
- Responsible for MLOps, managing the DevOps toolchain, and ensuring Continuous Development, Continuous Integration (CI), and Automated Testing using Jenkins, which streamlined the machine learning model deployment pipeline.
- Developed and deployed custom machine learning models using Azure Databricks, integrated with Azure Data Lake to handle advanced analytics, improving operational efficiency by 15%.
- Optimized ETL pipelines using Python, SQL, and Snowflake, reducing data latency by 20% through seamless integration of data from AWS S3 into Snowflake.
- Utilized containerization technologies such as Docker and Kubernetes to deploy machine learning models and ETL pipelines in isolated, scalable environments, ensuring consistency across development and production.
- Followed Agile practices to manage projects, conducting Scrum meetings, sprint planning, and continuous delivery, ensuring timely delivery of machine learning and data engineering solutions.

Tej Infotech

Associate Data Engineer

Ahmedabad, Gujrat, India

June 2019 – Jan 2021

- Automated data extraction and processing workflows using Python (Pandas, NumPy), reducing manual effort by 40%, improving efficiency in transforming financial datasets for SQL Server.
- Leveraged PySpark to migrate customer data from Teradata DWH to Hadoop, reducing operational costs by 30% through optimized data storage and robust data modeling processes.
- Developed web scraping tools using BeautifulSoup and Selenium, accelerating historical data collection speeds by 40%, increasing dataset availability for machine learning and Salesforce integration tasks.
- Created interactive dashboards in Power BI and Tableau to monitor key metrics for the Cloudera platform, providing real-time insights, improving platform efficiency in PROD, DEV, and UAT environments.
- Demonstrated strong problem-solving capabilities by implementing CI/CD pipelines using Git and Jenkins, automating deployment processes for data workflows.
- Utilized Jupyter Notebook for exploratory data analysis (EDA), prototyping machine learning models, and testing hypotheses to enhance data-driven solutions.

EDUCATION

Centre for Development of Advanced Computing

Post Graduate Diploma, Big Data and Computer Science

Pune, India

2020, Grade: A

Savitribai Phule Pune University

Bachelor of Engineering, Electronics & Telecommunications

(Credentials evaluated by World Education Services, ON, Canada)

Pune, India

2016 – 2019

GPA: 3.3/4.0

Cusrow Wadia Institute of Technology

Diploma of Engineering, Electronics & Telecommunications

(Credentials evaluated by World Education Services, ON, Canada)

Pune, India

2013 – 2016

GPA: 3.28/4.0

SKILLS

- **Data Engineering:** Apache Spark, Airflow, Kafka, Hadoop, Alteryx, ETL Pipelines, Informatica
- **Data Warehousing:** Redshift, Big Query, Snowflake
- **Databases:** Hbase, MongoDB, Cassandra, MySQL, Postgres, Hive
- **Cloud Platforms:** AWS (EMR, S3, Lambda), Azure (Data Lake, Data Factory, Databricks, Synapse Analytics), GCP
- **Languages:** Python, SQL, Linux Shell Scripting, Java, R, Scala
- **Other Tools:** Git, Jenkins, Docker, Machine Learning/AI

CERTIFICATIONS & ACHIEVEMENTS

- **Microsoft** – Azure Fundamentals AZ-900
- **IBM** – Data Analysis with Python
- **Databricks** – Databricks Fundamentals
- **Duke University, Coursera** – Data Engineering with Spark, Hadoop, and Snowflake
- **Bank of America** – Silver Award 2023, Bronze Award 2022
- **LinkedIn** – Tableau Essential Training

ADDITIONAL LINKS

- Tableau Dashboard – [Link](#)
- Credly Badges – [Link](#)
- Coursera Accomplishments – [Link](#)