

ADITI MISHRA

aditi.smile28@gmail.com | +1 (774) 525-6880 | github.com/aditimishra28 | linkedin.com/in/aditimishra-im/

SUMMARY

Experienced Data Engineer with **2+ years** of expertise, specializing in ETL development, data warehousing and business intelligence. Demonstrated skills in Cloud computing, optimizing data pipelines and using SQL and Tableau to deliver scalable data solutions.

WORK EXPERIENCE

Data Engineer, IBM

June 2022 - August 2024

- Designed and optimized large-scale **ETL pipelines** using **Hadoop, Spark, and SQL**, improving **data availability by 30%** and reducing **latency by 25%** for advanced analytics.
- Developed and maintained scalable **data processing solutions**, integrating **structured and unstructured data** to support **real-time** and batch workloads across cloud platforms.
- Enhancing query performance **SQL queries and data models**, and **processing efficiency** for handling billions of records, enabling faster business insights.
- Built APIs for efficient data access, reducing **retrieval time by 25%**, and collaborated with cross-functional teams to ensure **data integrity**, quality, and compliance.

Lead of AI/ML/Data Science, Google

September 2021 - June 2022

- Led students on **building data pipelines and machine learning models** for real-world applications.
- Guided teams in designing and implementing **ETL workflows**, ensuring **efficient data processing and integration** for AI-driven projects.
- Data visualization using Tableau**, enabling teams to **present insights effectively** and drive data-driven decision-making.
- Collaborated on **end-to-end data projects**, applying **big data technologies** and best practices to enhance **model performance and scalability**.

TECHNICAL SKILLS

Data Engineering: Python, JavaScript, Docker, Kubernetes, Hadoop, Spark, Cassandra, Git, Cloud

Data Analysis & Visualization: Matplotlib, Tableau, R, Predictive Modeling, Hypothesis Testing, Regression

Database: MS SQL Server, MySQL, PostgreSQL, Oracle SQL

Frameworks: TensorFlow, Scikit-learn, Pandas, NumPy, Scipy, Matplotlib, Seaborn

CERTIFICATIONS

- Apache Spark: Databricks Certified Associate Developer Apache Spark3. September 2022
- Cloud: 30 Days Google Cloud April 2022
- Machine Learning: IBM Machine Learning with Python December 2022

PROJECTS

Exploring Fuel Efficiency (MPG) | Feature Engineering, Regression Analysis

September 2024 - October 2024

- Conducted feature engineering and regression analysis to predict MPG, achieving 90% accuracy and identifying key factors affecting fuel efficiency.
- Optimized model performance through data preprocessing, outlier detection, and feature selection, improving interpretability and prediction reliability.

KMeans Visualization | Clustering, Python, Matplotlib

July 2024 - August 2024

- Implemented the KMeans clustering algorithm from scratch, partitioning randomly generated data points into specified clusters.
- Created an animated visualization to demonstrate centroid convergence and data separation in clustering.

R Code Password Manager | Flask, Chrome Extension, Expo

May 2024 - June 2024

- Designed a secure, three-tier password management system integrating a Flask-based backend, Chrome extension, and Expo mobile app.
- Implemented encrypted credential storage and seamless authentication, enhancing security and accessibility for users.

EDUCATION

Clark University, Worcester, United States of America

August 2024 - December 2025

Masters of Science in Data Analytics

CGPA: 3.7/4

Relevant Courses: Machine Learning, Deep Learning, Natural Language Processing, Big Data Technologies, Cloud computing

Dr. Abdul Kalam Technical University, Uttar Pradesh, India

August 2018 - June 2022

Bachelor of Technology in Electronics Engineering

CGPA: 3.4/4

Relevant Courses: Digital Signal Processing, Communication Systems, Data Structures & Algorithms, Probability & Statistics