

Akhlesh Rajput

Indore | akhleshrajput4@gmail.com | +91-7974949682 | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Aspiring Data professional with hands-on experience across data engineering, data analysis, and machine learning, including Python, SQL, Apache Spark (PySpark), Databricks, Azure Data Factory (ADF), ETL pipelines, and data warehousing. Skilled in building scalable data pipelines, distributed data processing systems, data modeling, workflow automation, and cloud-based analytics solutions, as well as developing machine learning models and business intelligence dashboards. Experienced in transforming large datasets into actionable insights using Spark, Power BI, and Python-based analytics. Strong foundation in DBMS, Operating Systems, Data Structures, and Algorithms. Actively seeking opportunities in Data Engineering, Data Analytics, Machine Learning, or Software Development roles.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, C
- **Data Engineering:** ETL Pipelines, ELT, Data Pipelines, Apache Spark (PySpark), Spark SQL, Databricks, Azure Data Factory (ADF), Data Warehousing, Data Modeling, Data Lakes, Batch Processing, Data Transformation
- **Cloud Technologies:** Microsoft Azure, Azure Storage, Azure Data Lake Storage (ADLS)
- **Big Data Technologies:** Apache Spark, Databricks
- **Databases:** MySQL, Relational Databases, Query Optimization, Schema Design, Data Integration
- **Machine Learning:** Machine Learning, Deep Learning, Medical Image Segmentation, Feature Engineering, Model Training
- **Frameworks & Libraries:** PyTorch, TensorFlow, Scikit-learn, OpenCV, NumPy, Pandas, Matplotlib
- **Visualization:** Power BI, DAX, KPI Reporting, Dashboard Development
- **Tools:** GitHub, Linux, Jupyter Notebook, Google Colab
- **Core Subjects:** DBMS, Operating Systems, Data Structures & Algorithms, Computer Networks

Internship Experience

AI/ML Intern | Maincrafts Technology

Feb 2026 – Jun 2026

- Developed a medical image segmentation system for brain tumor detection using Swin Transformer architecture.
- Performed MRI image preprocessing including resizing, normalization, augmentation, and mask generation using OpenCV.
- Implemented Swin Tiny Transformer using PyTorch for accurate semantic segmentation.
- Applied BCE Loss and Dice Loss to improve segmentation accuracy and model performance.
- Generated heatmaps and overlay visualizations for tumor localization and clinical interpretation.
- Evaluated model performance using Dice Score and IoU metrics in a GPU-enabled Google Colab environment.

PROJECTS

Brain Tumor Segmentation using Swin Transformer

- Developed a deep learning-based medical image segmentation system for accurate brain tumor detection from MRI images.
- Implemented Swin Transformer architecture using PyTorch to capture local and global contextual features.
- Performed image preprocessing including resizing, normalization, augmentation, and binary mask generation using OpenCV.
- Generated heatmaps and overlay visualizations for tumor localization.
- Evaluated model performance using Dice Score and IoU metrics.

Retail Sales Data Engineering Pipeline

- Designed and developed ETL pipelines using PySpark, Spark SQL, and Databricks for processing 50k+ retail transactions.
- Implemented data ingestion, cleansing, transformation, partitioning, and aggregation workflows.
- Applied data modeling and schema optimization techniques for analytical reporting.
- Improved distributed data processing performance by 30% using Spark optimization techniques.
- Generated business insights for customer segmentation and sales analytics.

Uber Trip Data Pipeline and Analytics Dashboard

- Built scalable data pipelines using Python, SQL, and PySpark for processing 100k+ Uber trip records.
- Performed data cleaning, transformation, aggregation, and analytics using distributed processing techniques.
- Identified demand patterns, peak hours, and revenue trends through large-scale data analysis.
- Developed interactive Power BI dashboards with 10+ KPIs for business intelligence reporting.
- Automated reporting workflows, reducing manual reporting effort by 40%.

Education

University of Hyderabad

2026

Master of Computer Applications (MCA)

Rajiv Gandhi Proudlyogiki Vishwavidyalaya (RGPV)

Bachelor of Technology (Civil Engineering)

CERTIFICATIONS

- **Introduction to Data Science — Cisco Networking Academy**
Learned fundamentals of data analytics lifecycle, visualization principles, and statistical concepts.
- **Programming in Python — Udemy**
Covered Python fundamentals, data structures, OOP concepts, and data analysis libraries.
- **MySQL Certification — Oracle**
Gained practical knowledge of SQL queries, joins, indexing, normalization, and relational database design.