

Ashutosh Joshi

📍 Baltimore, MD United States ✉ ashutxsh.jxshi@gmail.com ☎ 5513446092 📧 ashutosh--joshi 🌐 github.com/ashutoshjoshi1

SUMMARY

Results-driven Data Scientist and Software Engineer with over 4 years of experience in data analysis, machine learning, and cloud-based solutions. Adept at collaborating with cross-functional teams, including renowned organizations such as NASA and ESA, to develop and deploy scalable data pipelines and predictive models. Skilled in leveraging Big Data tools, including AWS S3, Athena, EMR, and Snowflake, to optimize data storage and real-time analytics. Proven expertise in building interactive dashboards and visualizations using Python, Power BI, and Tableau to drive data-driven decision-making and achieve significant cost reductions. Proficient in machine learning frameworks (Scikit-learn, PySpark, TensorFlow) and ensemble methods (LightGBM, bagging, boosting) for developing robust predictive models. Demonstrated success in designing ETL workflows, implementing CI/CD pipelines, and conducting exploratory data analysis (EDA) to uncover key business insights. Recognized for technical leadership, problem-solving aptitude, and commitment to continuous innovation in data science and engineering.

SKILLS

Programming: Python, Java, R, SQL, PySpark, Scala

Data Analysis: EDA, trends, patterns (Pandas, NumPy, Excel)

Statistics: Hypothesis testing, p-values, confidence intervals

Databases: SQL queries, joins, window functions, optimization

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

ETL: Pipelines, database schemas, normalization

Dashboards: Tableau, Power BI interactive reports

Big Data & Cloud: Apache Spark, Hadoop, AWS (Redshift, S3, Glue), Azure, Google Cloud Platform

Development: Python, Flask, ReactJS, REST APIs

Version Control: Git, GitHub, Jenkins

Cloud Platforms: AWS, Snowflake

Management: Agile, Scrum, JIRA, Kanban

EXPERIENCE

Software Engineer | SciGlob Instruments and Services | Columbia, MD | Nov 2024 - Present

- Collaborated closely with associates from [NASA](#) and [ESA](#) on the [Pandonia-global-network](#) research initiative.
- Develop and deploy automated ETL pipelines for structured and unstructured data using tools like PySpark, improving data ingestion and preprocessing for large-scale datasets.
- Work collaboratively with data engineers, scientists, and external research organizations (e.g., NASA, ESA) to ensure seamless integration of data pipelines and model deployments.
- Design, implement, and fine-tune machine learning models for predictive analytics, utilizing ensemble methods (e.g., LightGBM, bagging, boosting) to improve accuracy and performance.
- Provide actionable insights and recommendations through advanced data visualization techniques (e.g., Python libraries, Tableau) to help stakeholders and research partners make data-driven decisions.
- Leverage Big Data tools such as AWS S3, Athena, and EMR to process large datasets and optimize storage and querying performance for real-time analytics.
- Stay up-to-date with the latest trends in machine learning and data science (e.g., CNNs, LSTMs, and transformers) and propose innovative solutions to enhance ongoing research initiatives and optimize data analysis workflows.
- Conducted comprehensive analysis and exploratory data analysis (EDA) on log files from over 260 spectrometers deployed globally.
- Developed advanced visualizations to present complex data in a clear and insightful manner, enabling scientists and researchers to derive meaningful conclusions.

Data Analyst & Developer | 407 Associates LLC | Laurel, MD | April 2024 - Nov 2024

- Cleaned and transformed large datasets, ensuring accuracy and preparing data for analysis.
- Implemented data preprocessing workflows to handle missing data, outliers, and inconsistencies, ensuring high-quality input for machine learning models.
- Collaborated with product and engineering teams to design and deploy predictive models, contributing to improved operational efficiencies and customer insights.
- Developed machine learning pipelines in Python, using frameworks like Scikit-learn, PySpark, and TensorFlow, to create scalable solutions for real-time analytics.
- Designed and maintained data storage solutions in cloud environments (e.g., AWS S3, Snowflake), optimizing query performance and reducing costs.
- Created advanced visualizations and storyboards in Tableau to present technical findings in a clear, non-technical format for business stakeholders.
- Built and optimized ensemble models (e.g., bagging, boosting, LightGBM) to enhance model accuracy, ensuring robust performance across diverse datasets and use cases.
- Created interactive Power BI dashboards, enabling data-driven decisions and achieving a 25% cost reduction.
- Conducted exploratory data analysis, identifying key trends and patterns to support strategic planning.
- Developed SQL queries for complex data extraction and analysis, enhancing data accessibility for stakeholders.

Graduate Student Assistant | University of Maryland Baltimore County | Baltimore, MD | January 2023 - December 2023

- Performed data cleaning and statistical analysis on research datasets, increasing efficiency by 20%.
- Developed and implemented end-to-end data pipelines using PySpark and SQL for preprocessing large research datasets, improving scalability and reducing data processing time.
- Applied machine learning algorithms, such as regression models and clustering, to uncover hidden patterns in research data and provide actionable insights for stakeholders.
- Integrated cloud-based solutions to automate data ingestion and processing, enabling seamless access to large datasets for real-time analysis.
- Built Excel and Python-based tools for automating data processing, enabling faster and more accurate analysis.

System Engineer | Tata Consultancy Services LTD | Bangalore, IN | April 2021 - August 2022

- Conducted data cleaning and prepared data for analysis in large-scale data engineering projects.

- Collaborated with cross-functional teams to optimize data integration workflows by implementing automated validation checks and improving data transformation logic, enhancing data accuracy and consistency.
- Created comprehensive reports and visualizations using tools like Python and Power BI to effectively communicate insights and support design decision-making in large-scale data engineering projects.
- Analyzed key metrics and presented findings to project teams, improving decision-making on data integration.
- Assisted with design reviews by recommending ways to incorporate requirements into design and information/data flows.

Assistant System Engineer | Tata Consultancy Services LTD | Bangalore, IN | June 2020 - April 2021

- Conducted performance testing and optimized cloud applications, enhancing system reliability.
- Developed and maintained scripts for automated deployment and monitoring of cloud-based applications, improving the efficiency of DevOps workflows.
- Assisted in debugging and troubleshooting production issues, contributing to faster incident resolution and minimizing system downtime.
- Collaborated with cross-functional teams to refine application performance metrics, ensuring alignment with service-level agreements (SLAs) and business requirements.
- Contributed to the implementation of CI/CD pipelines, ensuring seamless integration, testing, and delivery of software updates across multiple environments.
- Wrote detailed technical documentation and user guides to support knowledge transfer and streamline onboarding for new team members.
- Participated in agile development cycles, collaborating on continuous improvement and feature delivery.
- Took an active role in code reviews, ensuring solutions were aligned to predefined architectural specifications.

PROJECTS

Development Performance Dashboard | 407 Associates, LLC

- Created a Power BI dashboard for construction data, providing insights into cost trends, seasonality, and performance metrics.
- Visualized data with Tableau, allowing stakeholders to monitor KPIs and make informed business decisions.

Customer Behavior Analysis

- Conducted A/B testing and statistical analysis on customer engagement data, providing actionable insights for marketing strategies.
- Applied hypothesis testing to determine statistically significant trends, enhancing campaign effectiveness.

Fraud Detection System

- Spearheaded the fraud detection system, leveraging Python(Pandas and NumPy) and SQL forthorough data analysis.
- Authored anomaly detection algorithms using ML (scikit-learn, TensorFlow), reducing fraudulent transactions by 18%.

EDUCATION

Master of Science in Data Science | University of Maryland Baltimore County | **Baltimore, MD** | **2022 - 2024** | **3.89/4.00 GPA**
Bachelor of Technology in Computer Science | Medi-Caps University | **Indore, IN** | **2016 - 2020** | **3.7/4.00 GPA**

CERTIFICATIONS & ACHIEVEMENTS

- Google Analytics Certification
- Certification in Data Structures with Python – University of Michigan
- Winner, Google Devfest 2018 Hackathon – Developed a decentralized application for asset donations
- Certification in Machine Learning with Andrew Ng | Stanford University