PURVA JAVALE
SOFTWARE ENGINEER | BIG DATA | MACHINE LEARNING

Kentucky, United States | +1(502) 650-1425 | purvajavale@gmail.com | <u>LinkedIn</u> | <u>Portfolio</u> | <u>GitHub</u> | <u>Medium</u>
Work Authorization: H4 EAD – No sponsorship required.

EDUCATION

Stanford University (Coursera)
Machine Learning Specialization | 2025
TrendyTech
Big Data Masters Program | 2025
Pune University
Bachelor of Engineering (B.E.) - Information Technology | 2014

TECHNICAL SKILLS

Big Data: Hadoop, Spark, Hive, HBase, Kafka, Airflow
Database: MySQL, SQL Server
Formats: TXT, CSV, XML, JSON, AVRO, ORC, PARQUET
Machine Learning: Data Preprocessing, Model-Ready Pipelines
Visualization: Power BI, Dashboard Design
Tools: Team Foundation Server, GitHub, Eclipse (Spring Boot), Visual Studio, PyCharm, Microsoft Project, Jira, Microsoft Office

Programming: Java, Python, SQL, .NET Core
Cloud & DevOps: AWS, Databricks, Maven
Database: MySQL, SQL Server
Legacy: IBM Mainframe (z/OS), WebSphere
Methodologies: Agile, SDLC, Data Governance, Compliance

PROFESSIONAL EXPERIENCE

Software Engineer | Hexanika – Pune, India | Feb 2015 – Jan 2022
- Scaled Hexanika's regulatory data platform from inception to enterprise adoption, automating compliance reporting for U.S. banks.
- Architected Spark-based data cleaning framework and designed ETL pipelines with schema checks, rule versioning, and audit controls, improving data quality and processing efficiency by 40%.
- Built and optimized pipelines using Java, Hadoop, Spark, and AWS, enhancing performance by 30% and reducing manual effort by 40%.
- Modernized legacy workflows by deploying Spark on IBM z/OS mainframe for parallel rule execution and faster compliance processing.
- Integrated processed datasets with Power BI dashboards for real-time regulatory insights and operational analytics.
- Developed SOAP APIs and dashboards to monitor pipeline performance, data controls, and end-to-end lineage.
- Mentored engineers on Spark optimization, product design, and data quality best practices, recognized as a thought partner in problem-solving and algorithm design.

Platforms Delivered
Smart Join - Regulatory Data Preparation
- Designed and implemented a data cleaning framework with standardized mapping, boosting data accuracy and reducing preprocessing workload by 40%.
- Implemented Spark-based transformation and validation engine with audit checks, schema validation, and governed rule execution.

Smart Reg - Regulatory Reporting
- Developed a Spark SQL rules engine covering FRY-9C, Call Reports, HMDA-LAR, and AML-SAR, accelerating project creation by 70%.
- Integrated predictive ML models for rule applicability scoring, reducing false positives, and improving compliance accuracy.

CERTIFICATIONS

Stanford University (Coursera) - *Machine Learning Specialization | 2025*
TrendyTech - *Big Data Masters Program | 2025*
Google (Coursera) - *Agile Project Management | 2024*
IBM (Coursera) - *IBM Mainframe | 2024*
Databricks - *Databricks Fundamentals | 2025, Databricks Platform Architect (AWS) | 2025*

PERSONAL PROJECTS

Data Quality & Cleaning Pipeline: Python pipeline on AWS S3 to clean, deduplicate, and validate datasets for ETL readiness.
Data Professionals Insight Dashboard: Interactive Power BI dashboards delivering insights on trends, tools, salaries, and skill gaps.
End-to-End Data Engineering Pipeline: Databricks pipeline powering business-ready dashboards from raw datasets.
Clustering & Fuzzy Matching Engine: Python engine combining K-Means and fuzzy matching to detect patterns in messy data.
Movie Review Sentiment Classification: Real-time sentiment analysis using Python and Streamlit, achieving 85%+ accuracy.

AWARDS & ACHIEVEMENTS

- Authored technical articles on Medium, covering data engineering and sharing practical insights for professionals.
- Took on diverse responsibilities across engineering, client delivery, and hiring to support organizational growth.
- Delivered guest lectures on Big Data at MIT Pune, mentoring and guiding aspiring data engineers.
- Recognized as Star Performer of the Year and Well-Done Achiever for driving innovation and delivering high-impact solutions.
- Achieved 2nd place in Hexanika's FinTech Hackathon for designing creative and effective data engineering solutions.