

# PURVA JAVALE

## SOFTWARE ENGINEER | BIG DATA | MACHINE LEARNING

Kentucky, United States | +1(502) 650-1425 | purvajavale@gmail.com | [LinkedIn](#) | [Portfolio](#) | [GitHub](#) | [Medium](#)

**Work Authorization:** H4 EAD – No sponsorship required.

---

### EDUCATION

**Stanford University (Coursera)** - Machine Learning Specialization | 2024 - 2025

**TrendyTech** - Big Data Masters Program | 2023 - 2025

**Pune University** - Bachelor of Engineering (B.E.) - Information Technology | 2010 - 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:** PowerApps, 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, Delta Lake

**Legacy:** IBM Mainframe (z/OS), WebSphere

**Methodologies:** Agile, SDLC, Data Governance, Compliance

---

### PROFESSIONAL EXPERIENCE

**Software Engineer | Hexanika – Pune, India | Feb 2015 – Jan 2022**

Delivered scalable Big Data and Automation solutions for regulatory compliance and banking operations.

#### Key Achievements & Responsibilities:

- Architected Hadoop Spark-based data cleaning framework and ETL pipelines with schema validation, rule versioning, and audit controls, improving data quality and processing efficiency by 40%.
- Built and optimized pipelines using Java, Hadoop, Spark, and AWS, enhancing throughput by 30% and reducing manual effort by 40%.
- Developed and maintained data pipelines and dashboards (Power BI, Databricks) to track automation performance, operational metrics, and business impact.
- Modernized legacy workflows by deploying Spark on IBM z/OS mainframes for rule execution, accelerating compliance processing.
- Participated in process analysis and requirements gathering to ensure automation aligns with business goals and compliance standards.
- Developed SOAP APIs and dashboards to monitor pipeline performance, data controls, and end-to-end lineage.
- Design, develop, and maintain automation solutions using tools such as Power Automate, Power Apps, Power BI, Python, and PowerShell.
- Evaluated emerging tools and technologies in AI-driven automation, ML models, for decision support and intelligent task handling.
- Collaborated with business teams to identify automation opportunities and standardize solutions.
- Contributed to continuous improvement through innovation, documentation, knowledge sharing, and mentoring engineers.

#### Platforms Delivered:

- **Smart Join - Regulatory Data Preparation:** Built a standardized Hadoop Spark-based data cleaning and transformation framework, boosting data accuracy and reducing preprocessing workload by 40%.
- **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.