

 [prabhat.ds@outlook.in](mailto:prabhat.ds@outlook.in)  
 +91 99817-83437  
 Bhopal (M.P.)  
 [LinkedIn](#)

## Technical Mastery

### Programming & Fundamentals

- **Languages:** Python, R
- **Libraries:** Pandas, Scikit-learn
- **Databases:** SQL, Neo4j
- **Tools:** Git, GitHub

### Generative AI & Agents

- **Frameworks:** Agentic Workflows (LangChain, LlamaIndex, MCP, A2A)
- **Architectures:** Mixture of Experts (MoE), RAG, Multi-Agent Systems
- **Models:** LLMs (GPT-4o, Gemini), Hugging Face, Model Fine-tuning
- **Vector DBs:** OpenSearch, Qdrant, Pinecone

### Core AI/ML & Analytics

- NLP & Named Entity Recognition (NER)
- Time Series Analysis (Prophet, ARIMA, Holt-Winters)
- Deep Learning & Statistical Modeling
- Data Mining & Cluster Analysis

### Data Engineering & Platforms

- **Platforms:** Palantir Foundry & AIP, Databricks (Delta Lake), C3.ai
- **Processing:** PySpark, Spark SQL, ETL Pipeline Design

### Cloud, MLOps & Visualization

- **Cloud:** AWS (EC2, S3, Lambda, CaaS), GCP
- **DevOps:** Docker, Kubernetes, Terraform, CI/CD (GitHub Actions)
- **BI:** Power BI, Tableau, Streamlit, Plotly/Dash, Matplotlib/Seaborn

# Prabhat Shukla

(Lead Data Scientist | AI Architect)

## Executive Summary

Strategic AI leader with 12 years of professional experience, including 9+ years of specialized expertise in Machine Learning, NLP, and Generative AI. Proven track record of delivering high-scale industrial solutions, most notably an LLM-based warranty system that recovered \$55M+ for a Fortune 500 client. Expert in orchestrating Multi-Agent Systems (MAS) using Model Context Protocol (MCP) and Mixture of Experts (MoE). Deeply proficient in AWS, Databricks, and Palantir Foundry, with a focus on bridging the gap between experimental R&D and production-grade MLOps.

## Experience

**Yash Technologies Pvt Ltd - Lead Data Scientist | Pune**  
*Strategic Lead for John Deere & PALL Corporation*

**March 2020 - Current**

- **Agentic Translation Ecosystem (MCP & A2A):** Architected a multi-agent ecosystem on AWS (CaaS) utilizing Model Context Protocol (MCP) and MoE (Mixture of Experts) for optimized model routing.
  - **Agents:** Authoring Memory (RAG-based), Translation Memory (OpenSearch integration), and Quality Evaluation (Automated Audits).
  - **Impact:** Reduced localization turnaround by 35% and achieved a 90% consistency rate in technical terminology across multi-lingual deployments.
- **LLM-Based Translation Quality Evaluation (MTQE):** Engineered an automated audit system using LLMs and Galileo for real-time quality assessment. Enriched model prompts via Vector Database retrieval to inject Deere-specific style guides and approved terminology.
  - **Evaluation Framework:** Automated scoring across five dimensions—fluency, accuracy, terminology, style, and local conventions—with automated severity tagging (Major to Neutral).
  - **Impact:** Achieved a 20% reduction in manual linguistic effort in Phase 1, with a scalable architecture designed to reach 50% cost savings in translation QA.
- **Warranty Responsibility Code Assignment:** Designed and deployed a hybrid ML-driven solution leveraging Large Language Models (LLMs) to automatically assign responsibility codes to warranty claims (Supplier/Deere liability). The LLMs analyzed the complaint cause correction text to accurately predict the responsible party.
  - **Impact:** Recovered over \$55 million from previously rejected or pending claims, significantly increasing the claim acceptance rate and reducing rejections due to accurate and automated responsibility assignments.
- **Palantir Foundry Data Engineering:** Engineered high-scale data ingestion and ETL pipelines within Palantir Foundry using PySpark to unify disparate extended warranty datasets. Implemented complex data enrichment logic to transform raw telematics and claim data into actionable insights.
  - **Impact:** Drastically reduced time-to-insight for warranty analysts and improved data reliability, enabling more accurate long-term liability forecasting for key stakeholders.
- **Attachments Parts Forecasting (Time Series):** Developed an ML-driven forecasting framework for attachment parts using Prophet, ARIMA, and Holt-Winters models.

## **Education & Certifications**

- **PGBDA:** Siddaganga Institute of Technology (2017)
- **B.E.(CS):** RKDF College of Engineering (2011)

## **Certifications**

- C3.AI V8 Data Science
- Data Scientist with Python Track
- Deep Learning Specialization
- Graph Analytics for Big Data.

## **Accomplishments**

- Star Achiever Award
- Prime Player Award
- Super Star Award

Incorporated variables such as historical sales and base-coded/non-base-coded attachments, evaluating performance via RMSE and MAPE.

- **Impact:** Reduced inventory discrepancies by integrating high-accuracy forecasts.
- **Enterprise Data Automation & Business Intelligence:** Engineered end-to-end data pipelines in Databricks and developed a suite of automated Power BI dashboards to drive data-driven decision-making. Automated complex reporting for Order Response, Site Logistics, and Factory Performance.
  - **Impact:** Eliminated 1,100+ manual hours per year and accelerated time-to-insight for critical logistics and factory performance metrics.
- **Predictive Analytics (PALL Corp):** Created data pipelines to feature-engineer XML log files from Flowstar machines. Developed an ML model to predict integrity test outcomes at an early stage.
  - **Impact:** Enabled proactive failure identification in medicinal drug manufacturing, drastically reducing production downtime.

### **Accenture - Sr. Data Scientist | Gurgaon**

**July 2018 – Feb 2020**

**Client:** Google, Inc. (Google Retail)

- **GCP-Based Prototype Development:** Architected and deployed end-to-end ML prototypes using Google Cloud Platform (GCP) and Python to solve high-priority business challenges for Google Retail.
- **Market Basket Analysis:** Applied the Apriori algorithm to identify high-confidence purchasing patterns, optimizing cross-selling and product bundling strategies.
- **Email Send-Time Optimization:** Developed regression models to predict optimal email dispatch times, maximizing CTR and campaign engagement.
- **User Segmentation:** Leveraged PySpark K-means to segment millions of site visitors, enabling highly targeted marketing and personalized UX.

### **Northout Solutions - Data Scientist | Indore**

**Dec 2017 – July 2018**

**Client:** John Hancock Financial.

**Financial Spending Analysis:** Analyzed user transactional behavior and predicted future transactions using ARIMA and regression models, providing deep insights for personalized financial planning.

### **Bonsmat Group - Machine Learning Engineer | Ludhiana**

**Sept 2017 – Nov 2017**

- **Conversational AI Assistant:** Developed a REST API-based chatbot with weather forecasting and news summarization capabilities using NER and text classification.

### **Constalytics - Data Engineer | Mohali**

**April 2017 – Aug 2017**

- **Knowledge Graph Platform:** Architected an NLP platform using Neo4j to visualize entity relationships and perform automated sentiment analysis.

### **DSRI – Machine Learning Researcher | Bengaluru**

**Aug 2016 – March 2017**

- **Big Data Research:** Conducted large-scale cluster analysis on California weather data using PySpark as part of PG Diploma research.

### **Early Career: Software & Web Development**

**Jan 2012 – July 2016**

- **Predictive Research (Software Developer):** Focused on full-lifecycle development and testing of data-centric applications.
- **Freelance (Web Developer):** Architected and maintained custom web solutions and database schemas for various clients.