

# Vibin Ravikumar

## Data Scientist

✉ vibin.ravikumar01@gmail.com    ☎ +91 8369707681    📍 Chennai, India

🌐 [linkedin.com/in/vibin-ravikumar-963314a4](https://www.linkedin.com/in/vibin-ravikumar-963314a4)

Data scientist with 3+ years of experience building scalable data pipelines and deploying advanced ML models in computer vision and multimodal LLMs. Combines technical leadership with hands-on skills to integrate infrastructure and analytics for solving complex business challenges.

### Professional Experience

01/2024 – present  
Chennai, India

- Lead Data Scientist, AiSPRY**
- Designed and optimized scalable PySpark and Hadoop pipelines, integrating ML models into production and accelerating data processing for analytics and feature engineering.
  - Developed robust ETL workflows that transformed raw data into structured formats, improving data quality for business intelligence and predictive modeling.
  - Delivered end-to-end machine learning solutions, combining data engineering and data science to boost model accuracy and streamline deployment.

04/2022 – 01/2024  
Chennai, India

- Associate Data Scientist, 360DigiTMG**
- Engineered scalable data preprocessing and validation pipelines for terabytes of multimodal data using distributed processing to handle high-volume workloads efficiently.
  - Designed and optimized ML models for large-scale multimodal applications, applying advanced feature engineering and hyperparameter tuning to extract actionable insights.
  - Collaborated with engineering and business teams to translate requirements into technical solutions, while establishing documentation and data governance for enterprise-grade ML operations.

### Projects

#### HOSPITAL ANALYTICS DASHBOARD PROJECT

**Business Problem:**

Healthcare organization requires a unified analytics platform to integrate data sources and deliver real-time insights for optimizing clinical, operational, and financial performance.

**Key Responsibilities:**

Architecting PySpark pipelines to process terabytes of healthcare data from EHR, billing, and scheduling systems into a scalable data lake. Developing predictive models for admissions, resource use, and financial outcomes with HIPAA-compliant ETL. Creating cross-functional KPIs and interactive dashboards delivering role-based insights for clinical and operational teams.

**Progress & Outcomes:**

Established a solid data foundation with dashboards and predictive models achieving 85%+ accuracy, enabling a 25% faster decision-making process and 15% improved resource allocation for proactive hospital management.

#### LEGAL CHATBOT PROJECT

**Business Problem:**

Legal professionals lacked efficient access to insights from India's new legal codes and thousands of court judgments, requiring an AI solution to convert vast documents into accurate, context-aware legal guidance.

**Key Responsibilities:**

Built end-to-end PySpark pipelines to process thousands of legal PDFs into optimized datasets stored in S3. Developed a scalable FAISS vector search system to index terabytes of embeddings for fast retrieval and implemented a retrieval-augmented generation model using Google AI Embeddings and Groq’s LLaMA 3, achieving 95% accuracy with strong context filtering.

**Outcomes:**

Delivered a production-ready legal AI assistant that reduced legal research time by 60%, enabling precise extraction from gigabytes of complex legal text.

**MEDICAL INVENTORY OPTIMIZATION**

**Business Problem:**

Healthcare provider struggled with inefficient pharmaceutical inventory management, resulting in both stockouts affecting patient care and excess inventory tying up capital.

**Key Objectives:**

Engineered data processing pipeline to handle close to a gigabytes of historical utilization data for analysis, with batch processing of gigabytes of transaction data for forecasting. Developed an ensemble forecasting system integrating LSTM, GRU, and RNN architectures to predict 3-4 month demand patterns for pharmaceutical products with varying consumption patterns.

**Outcomes:**

Delivered forecasting system with <5% MAPE accuracy that processed gigabytes of monthly data, optimizing inventory across facilities while reducing holding costs and maintaining high drug availability rates.

Skills	
<b>Machine Learning &amp; Deep Learning</b> <ul style="list-style-type: none"><li>Neural Networks, Ensemble Methods, Deep Learning Architectures.</li><li>Computer Vision</li><li>Facial Recognition and Biometric Authentication</li><li>Time Series Analysis and Forecasting</li><li>Multimodal Language Models (LLMs)</li><li>Natural Language Processing</li><li>Regression and Classification Models</li><li>Cluster Analysis &amp; Dimension Reduction</li></ul>	<b>Data Engineering</b> <ul style="list-style-type: none"><li>Data Pipeline Architecture &amp; Implementation</li><li>ETL/ELT Process Design and Optimization</li><li>Data Lake &amp; Data Warehouse Solutions</li><li>Real-time &amp; Batch Processing Systems</li><li>Database Management</li><li>Big Data Technologies</li><li>Data Modeling &amp; Schema Design</li><li>Data Quality &amp; Validation Frameworks</li><li>API Development &amp; Integration</li></ul>
<b>Exploratory Data Analysis</b> <ul style="list-style-type: none"><li>Statistical Analysis and Hypothesis Testing</li><li>Data Visualization</li><li>Pattern Recognition and Trend Identification</li><li>Outlier Detection and Anomaly Analysis</li><li>Feature Importance Analysis and Selection</li><li>Dimensionality Reduction</li><li>Distribution Analysis and Transformation</li><li>Interactive Dashboard Creation</li></ul>	<b>Programming &amp; Libraries</b> <ul style="list-style-type: none"><li>Python, SQL, PySpark</li><li>TensorFlow, PyTorch, Keras</li><li>Flask, FastAPI, Streamlit</li><li>Apache Spark, Hadoop, Airflow</li><li>OpenCV, scikit-learn, pandas, NumPy</li><li>FAISS, Hugging Face Transformers, LangChain</li><li>MySQL, MongoDB</li><li>Docker, Kubernetes</li><li>Seaborn, Plotly</li><li>InsightFace, DeepSORT, YOLO</li><li>Google Generative AI, Groq, Mistral, OpenAI</li></ul>
<b>Soft Skills</b> <ul style="list-style-type: none"><li>Team Leadership and Mentoring</li><li>Research Publication &amp; Technical Documentation</li><li>Client Communication &amp; Agile Methodologies</li><li>Public Speaking</li></ul>	

Publications	
31/03/2025	AI-Powered Legal Chatbot for Litigative Situations: Leveraging BNS, BNSS, and BSA with Contextual Document Integration - IJISRT <a href="#">🔗</a>
16/09/2023	Forecasting Drug Demand for Optimal Medical Inventory Management: A Data-Driven Approach with Advanced Machine Learning Techniques - IJISRT <a href="#">🔗</a>

---

## Education

---

Chennai, India

**B.E Mechanical Engineering**, *Anna University Chennai*

---

## Languages

---

English

Hindi

Malayalam

Tamil

Marathi

---

## References

---

**Bharani Kumar Depuru**, *CEO & Founder*, AiSPRY, 360DigiTMG  
Bharani@360digitmg.com

---

## Declaration

---

I hereby declare that the information provided in this resume is true and accurate to the best of my knowledge. All achievements and responsibilities mentioned reflect my genuine professional experience. I understand that any misrepresentation may lead to disqualification from employment consideration or termination if discovered after employment begins.

---

**Vibin Ravikumar**  
Chennai, 16th May 2025