

# KRISH BAKSHI

☎ +91-8483035625 | ✉ [krishbakshi23@gmail.com](mailto:krishbakshi23@gmail.com) | [🌐 LinkedIn](#) | [🐙 GitHub](#)

## SKILLS

---

- **Programming Languages:** Python, SQL, C/C++.
- **Data & Visualization:** PySpark, Pandas, Matplotlib, Seaborn, Plotly.
- **ML & DL:** PyTorch, TensorFlow, Scikit-learn, OpenCV, statsmodels, HuggingFace, MLflow, Databricks.
- **Big Data:** Apache Spark, BigQuery.
- **Deployment:** Flask, FastAPI, Docker, GCP.
- **Version Control:** Git, GitHub.
- **Languages:** English (Full professional), Japanese (JLPT N3), Hindi, Punjabi, Marathi (Bilingual).

## WORK EXPERIENCE

---

**DATA SCIENTIST INTERN:** *Nasiwak Services India Private Limited* 04/2025 – Present

- [AutoMailAI](#) helped me secure this internship opportunity through LinkedIn outreach.
- Backed by a Japanese organization, contributed to delivering technical documentation to technical and non-technical stakeholders in English and Japanese.
- Researched and developed object detection pipelines for niche-specific tasks, such as floor plans, using YOLO and OpenCV.
- Integrated custom ML & AI pipelines into applications such as Selenium-based automation solutions, etc.

**DATA SCIENCE INTERN:** *Metafied (Formerly PROFCESS)* 07/2024 – 01/2025

- Incubated by Harvard Innovation Labs and mentored by ex-IIT and ex-Harvard founder and senior staff.
- Contributed to designing and deploying scalable data pipelines for time series forecasting, utilizing the Azure Databricks platform and PySpark (Python API for Apache Spark) for data extraction, transformation, and loading (ETL) into cloud-based data lakes.
- Implemented and optimized auto-regression models such as XGBoost, ARIMA, and SARIMA, enhancing the accuracy of predictive insights for future sales.
- Developed data queries to surface insights and trends from large datasets from BigQuery, worked on various projects involving Time series, Computer vision, and LLM-based applications.

**RPA INTERN:** *PROAZURE SOLUTIONS PVT LTD* 12/2023 – 01/2024

- Led a team to design and implement Robotic Process Automation (RPA) solutions, driving automation initiatives for repetitive tasks.
- Automated workflows, including web scraping, online data collection, and Excel automation, leading to a 30% improvement in task efficiency.

**WEB DEVELOPMENT INTERN:** *RB TECH SERVICES* 07/2020 – 09/2020

- Led a team of developers in building a dynamic website, overseeing both frontend and backend development processes.
- Managed database design and real-time connectivity using phpMyAdmin and Wamp Server, ensuring seamless data flow between the server and client-side application.

## PROJECTS

---

**AutoMailAI:** *Python, Gemini 2.0 Flash, Gradio, Gmail API, AI Automation.*

- Built an AI-based tool for personalized cold email generation using prompt engineering with dynamic templates, enabling tailored outreach for Jobs & Internships from user inputs and resumes.
- Designed a Gradio interface with single & batch processing (Excel), automated DOCX & ZIP generation, and integrated Gmail API for auto-drafting emails at scale.

**ImaginAIry:** *Python, Stable Diffusion XL, Gemini 2.0 Flash, Gradio, Text-to-Image.*

- Built a text-to-image generation pipeline using Stable Diffusion XL with prompt augmentation via Gemini 2.0 Flash.
- Fine-tuned Gemini API on a custom prompt dataset for improved image description generation.

**PulseMate:** *Python, LLM, RAG, GPT 3.5, Decision Tree, Medical Sciences.*

- Developed an AI-driven web application using BRFSS data and Decision Tree models to assess cardiovascular risk based on user inputs (health, demographics, lifestyle).
- RAG pipeline with MedQuAD and PubMedQA to generate personalized, evidence-based medical advice.

**Time Series Forecasting For Future Sales:** *Python, XGBoost, PySpark, Gradio*

- Developed a sales forecasting model using XGBoost, accelerated with CUDA for efficient training, achieving a training time of 7 seconds. Data was transformed using PySpark for enhanced scalability and performance.
- Initially, the model had a 60% Mean Absolute Percentage Error (MAPE). After feature extraction and training, the model achieved 10% MAPE and 90% accuracy.

**AI Integrated Analytics Dashboard:** *Python, PySpark, BigQuery, Gemini, Streamlit, Plotly.*

- Designed and deployed a live analytics dashboard using Streamlit, integrated with Gemini to summarize key performance indicators (KPIs).
- Utilized PySpark to surface trends and retrieved data from a BigQuery data lake for efficient analysis.

## RESEARCH & PUBLICATIONS

---

*Krish Bakshi, Sunil Bade, Mayur Bhand, Siddhesh Dhindale, Dr.H.B.Jadhav: Crop Classification using Convolutional Neural Network | IJSREM Accepted in February 2025 DOI:10.55041/IJSREM41238*

## EDUCATION

---

**SAVITRIBAI PHULE PUNE UNIVERSITY** || Pune

08/2022 – 06/2025

*Bachelors of Computer Engineering*

**Relevant Coursework:** Data Structures and Algorithm, Probability and Statistics, Machine Learning.