

KRISH BAKSHI

✉ work.krishb@gmail.com | [LinkedIn](#) | [GitHub](#) | [X](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Scientist / AI Engineer building **production AI systems** across **computer vision**, **Speech**, **forecasting**, **AI automation**, and **agentic workflows**. Experience **training**, **optimizing**, and **fine-tuning** ML/DL models, curating **datasets**, and deploying **end-to-end ML pipelines** from **data engineering** to **evaluation** and **production**. Built applied **LLM systems** and **AI agents** for **retrieval**, **workflow automation**, and user-facing **AI applications**, using **Azure Databricks**, **BigQuery**, **FastAPI**, **Docker**, **AWS**, and **GCP** in **production** settings. **Japanese-proficient (JLPT N3)** with experience working in **bilingual technical environments**.

WORK EXPERIENCE

DATA SCIENTIST *Nasiwak Services India Private Limited* 07/2025 – Present

- Built and deployed a **production CV pipeline** for **floor-plan analysis** within AI automation workflows, reducing **manual effort by 80%**, achieving **87.89% mAP**, and supporting **10,000+ orders/month**.
- Trained and **fine-tuned object detection** models for **building materials** and **85+ small electrical symbols**, while coordinating **320+ hours of annotation** and **augmentation** across a **5-member bilingual** team to improve **small-object detection quality**.
- Built and shipped a full-stack blueprint-editing MVP across **React**, **Next.js**, **FastAPI**, **Docker**, and **AWS EC2** for real-time prediction and editing workflows.
- Developed **AI-powered karaoke and audio-processing pipelines** using **yt-dlp**, **Demucs**, **TTS**, **Omnilingual ASR**, **FFmpeg**, and **C++**, generating karaoke videos in under **20 seconds** with real-time **autotune** and significant **noise reduction**.

DATA SCIENTIST INTERN: *Nasiwak Services India Private Limited* 04/2025 – 07/2025

- [AutoMailAI](#) my **AI Agent** helped me secure this internship opportunity through **LinkedIn outreach**.
- Developed **object detection pipelines** for floor plans data extraction **saving 60% manual effort**, using **YOLO**, **OpenCV** & **SAHI** achieving **mAP-0.5 82.06%** in production.
- Backed by a Japanese organization, contributed to delivering technical documentation to technical and non-technical stakeholders in **English** and **Japanese**.
- Integrated custom AI/ML pipelines into **Selenium**-based automation solutions for automated Invoice generation, making the process **faster** and **reliable**.

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

- Incubated by **Harvard Innovation Labs** and mentored by ex-IIT and ex-Harvard founder & senior staff.
- Engineered and deployed scalable data pipelines** for time-series forecasting on **Azure Databricks**, leveraging **PySpark** & **ETL workflows** to streamline ingestion into cloud-based data lakes.
- Built and optimized** an **XGBoost** model achieving **92.81% accuracy on regular sales** and **86.79% on promotional sales**, by engineering advanced features around **inventory**, **logistics**, **past sales** and **regional weather forecasts**.
- Experimented with **auto-regression models** **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.

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

PROJECTS

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

[\[Link\]](#)

- 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.*

[\[Link\]](#)

- 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.

SEC Filings QA Agent: *Python, Langchain, ChromaDB, Gemini, Streamlit, Financial RAG*

[\[Link\]](#)

- Built an end-to-end RAG pipeline using *Langchain* and *ChromaDB* to automate the extraction of insights from complex SEC filings (10-K/10-Q), streamlining financial document analysis.
- Developed a responsive Streamlit interface for real-time natural language querying of financial data, leveraging advanced indexing to handle both unstructured text and tabular data.

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

[\[Link\]](#)

- 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.

SKILLS

- **Programming Languages:** Python, SQL, C/C++, Javascript, Typescript, HTML, CSS.
- **Data & Visualization:** PySpark, Pandas, Matplotlib, Seaborn, Plotly.
- **ML & DL:** PyTorch, TensorFlow, Scikit-learn, OpenCV, HuggingFace, MLflow, langchain, LlamaIndex.
- **Web Frameworks:** React.js, Next.js, Node.js, Vue.js, Bun, Tailwind.
- **Databases:** PostgreSQL, MySQL, Pinecone, FAISS, Qdrant, ChromaDb.
- **API/Deployment:** Flask, FastAPI, Django, Docker.
- **Cloud Platforms:** Azure Databricks, AWS EC2, GCP, BigQuery, Weights & Biases, Langsmith.
- **Version Control:** Git, GitHub.
- **Languages:** English (Full professional), Japanese (JLPT N3), Hindi, Punjabi, Marathi (Bilingual).

EDUCATION

SAVITRIBAI PHULE PUNE UNIVERSITY || Pune

Bachelors of Computer Engineering [BE Computer]

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