Siddhi Kale

AI, machine learning and data science (fresher) | Pune, India | Open to internships and full-time roles
+91 85301 82828 | siddhi.kale2503@gmail.com | linkedin.com/in/siddhikale25 | github.com/siddhi-works | siddhi-kale.vercel.app

Summary

Final-year B.Tech student in AI and Data Science (CGPA 9.02). First author of a published machine learning paper (Springer Nature), First Prize at TechVerse 2026 and Rank 7 at the Shinchan Online Hackathon 2026. Builds data pipelines, models and agent systems in Python and R, from raw data to a deployed app.

Education

B.Tech in Artificial Intelligence and Data Science, Ajeenkya DY Patil University, Pune | Jul 2023 to Jul 2027 | CGPA 9.02/10 (up to 6th semester)

Experience

Data Science Intern, Aabhak Impex Pvt. Ltd. | Aug 2026 to Oct 2026

- Analysed sales and inventory data using Python and Pandas to identify business trends.
- Performed exploratory data analysis to understand monthly sales patterns and product demand.
- Developed a basic sales forecasting model using Linear Regression and visualised the results with Matplotlib.

Content Editor and Management Coordinator, Abhivyakti Editorial Team, ADYPU | Apr 2025 to Nov 2025

- Ran a content calendar across 5 platforms and a new workflow that raised production efficiency by 30%; errors down 15%.

Projects

Bot detection model for social media (R) | Published at ICAIML-AE 2026 | Code

- Built a Random Forest bot detector that reads what accounts post (TF-IDF) and how they behave (mentions, retweets, follower counts). Compared 5 models on 50,000 balanced Kaggle records with an 80:20 split.
- Random Forest was best: 90% accuracy, AUC 0.9685, F1 0.906. Supervised models beat K-Means by a wide margin, and behaviour features mattered more than text.

R, ranger, caret, text2vec, pROC, ggplot2

Chicago Crime Analytics: interactive dashboard | TechVerse 2026 First Prize | Live demo | Code

- Led Team Moleculens: cleaned 1,048,575 crime records into a live dashboard with 8 KPIs, 10 charts, a district map and an AI assistant.
- A Data Trust page lists every cleaning step and limitation (31% of coordinates missing, no rows removed). A fair-arrest analysis adjusts for crime type and shows 86% of the evening-to-night gap comes from what was reported, not the time of day.

Python, pandas, React, TypeScript, d3, Vercel

Plexis: Alexa+ add-on that plans trips and proves each plan | Code

- Built an MCP server with 10 tools for Alexa+. A CP-SAT solver finds the plan and an independent checker verifies it; impossible requests get the smallest fix, proven by solving again.
- On 80 seeded requests every plan passed every check and no impossible request got a wrong plan (greedy baselines passed 13.8% and 25.0%). 105 tests, Docker, CI.

Python, OR-Tools CP-SAT, Z3, MCP, LangGraph, Amazon Bedrock, Docker

ARIA: research agent | Code

- Built a LangGraph and LangChain agent that searches arXiv, reads and compares papers, finds gaps and writes a LaTeX paper exported as a PDF, with a Streamlit app and chat memory.

Python, LangGraph, LangChain, Gemini 2.5 Flash, arXiv API, LaTeX, Streamlit

Skills

Languages: Python, R, SQL, TypeScript

ML and data: scikit-learn, TensorFlow, Keras, OpenCV, pandas, NumPy, NLTK, Matplotlib, ggplot2, caret, text2vec

Agents and optimisation: LangGraph, LangChain, Gemini API, OR-Tools CP-SAT, Z3, MCP

Tools: Git, GitHub Actions, Docker, Streamlit, SQLite, Vercel

Publication and achievements

- First and corresponding author, "Hybrid Feature-Based Comparative Analysis of Supervised and Unsupervised ML-Based Intelligent Bot Detection Model for Social Media Data", ICAIML-AE 2026, Atlantis Press (Springer Nature), doi.org/10.2991/978-94-6239-791-0_10.
- First Prize, TechVerse 2026 (team leader). Rank 7 of 198 ranked projects, Shinchan Online Hackathon 2026. Smart India Hackathon 2026 internal-round winner. Winner, ADYPU internal hackathon.

Certifications

- AWS re/Start Graduate, Amazon Web Services. Artificial Intelligence and Machine Learning, Edunet Foundation.
- GenAI-Powered Data Analytics Job Simulation, Tata (Forage). Data Science, CodeWithHarry.