

SIYAL KAMBALE

Pune, Maharashtra, India | siyalkamble9106@gmail.com | +91-75585 01452

github.com/Siyalkamble | linkedin.com/in/siyalkamble

OBJECTIVE

Final-year BSc Computer Science student focused on Python, machine learning, and model evaluation. Building projects in classification and regression with scikit-learn, with interest in NLP, LLMs, and practical AI engineering.

EDUCATION

G H Raisoni College of Arts, Commerce and Science, Pune

Expected Graduation: 2027

- Pursuing Bachelor's degree, focused on AI/ML engineering through self-driven, project-based learning.
- GPA: 9.18/10.0
- Relevant Coursework: DSA, OOP, DBMS, Machine Learning, Deep Learning.

TECHNICAL SKILLS

Languages: Python, SQL, C(Basics)

ML/DL: scikit-learn, PyTorch, Fastapi, Pydantic, Streamlit

Data Tools: NumPy, Pandas, Matplotlib, Seaborn

Tools: Git/GitHub, Linux (Arch), VSCode, Google Colab, Kaggle API, Jupyter Lab

PROJECTS

Medical Insurance Cost Predictor (Regression) [GitHub Link](#)

- Problem: Predict medical insurance charges from 1,338 records for pricing/underwriting use cases.
- Key decision: Built a **ColumnTransformer** + **Pipeline** + **GridSearchCV** setup to fairly benchmark 5 regression models instead of hand-tuning one.
- Result: Random Forest won with $R^2=0.865$, MAE=\$2,674 — beat linear models by ~9 points.
- Insight: Linear models plateaued at $R^2 \approx 0.78$ because they can't model the smoker $\times$ BMI interaction (non-linear cost explosion for smokers with high BMI)
- **Tech:** Python, scikit-learn, Pandas · Pipeline · GridSearchCV · Random Forest · Ridge/Lasso · CV

Employee Turnover Prediction (Classification) [GitHub Link](#)

- Problem: Predict employee attrition from 1,350 HR records to flag at-risk employees before they leave.
- Key decision: Compared baseline, L1, and L2 regularized logistic regression instead of jumping to ensemble models — needed interpretable coefficients for HR stakeholders, not just raw accuracy.
- Result: L2 (Ridge) gave 88.5% accuracy, 0.958 ROC-AUC, 86.6% recall.
- Insight: Detected multicollinearity via VIF/ $R^2=0.97$ on **Annual_Bonus_Squared**, dropped it — improved coefficient stability without sacrificing recall.
- **Tech:** Python, scikit-learn, Pandas · Logistic Regression · L1/L2 · ROC-AUC

Neural Network from Scratch in Numpy [GitHub Link](#)

- Problem: Understand backprop and gradient flow at the tensor-math level before relying on framework autograd.
- Key decision: Implemented forward + manual backward pass (no autograd) starting from XOR (non-linearly separable, forces needed for hidden layers), then a general ANN, then a char-level LSTM with gate computations written explicitly.
- **Tech:** Python, NumPy · Manual backpropagation · Gradient descent · LSTM gate mechanics

CERTIFICATIONS

- Apna College — Prime AI/ML (In Progress)
- Machine Learning — Fundamentals of Machine Learning using Python ([Udemy](#))
- Master Python — with numpy and pandas for Machine Learning ([Udemy](#))
- Python for beginners— Learn all the basics of Python ([Udemy](#))