LIME vs SHAP for Explainable AI (XAI)

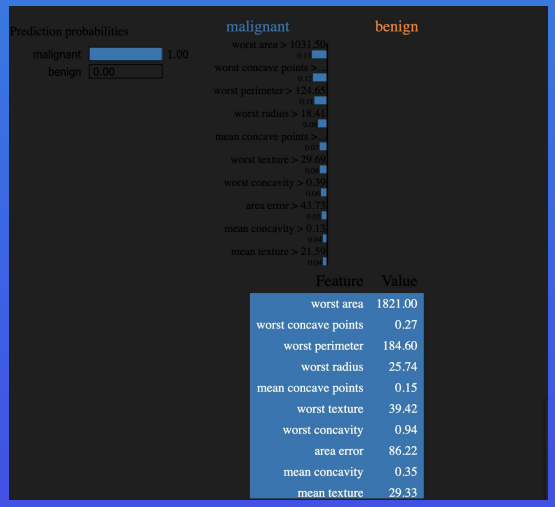
- Breast Cancer dataset (scikit-learn)
- Random Forest model (200 trees)
- Focus: interpreting a single prediction + global behavior

TL;DR

- LIME → simple, human-readable rules around one prediction
- SHAP → Shapley-value attributions with global + local views
- For our model, worst area / concavity / radius dominate decisions
- SHAP is more stable; LIME is more story-like for clinicians

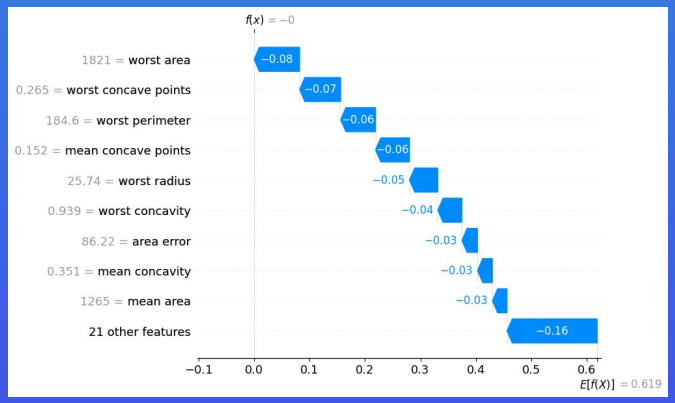
LIME Local Explanation – How to Read This

- Left: model says malignant with probability 1.00 (blue bar)
- Middle: blue bars = features pushing toward malignant; orange would be toward benign
- Each rule (e.g., "worst area > 1031.50") is a local decision boundary
- The larger the bar, the larger that feature's impact on this patient's malignancy score
- Table at bottom lists the actual values (e.g., worst area = 1821, worst concave points = 0.27)



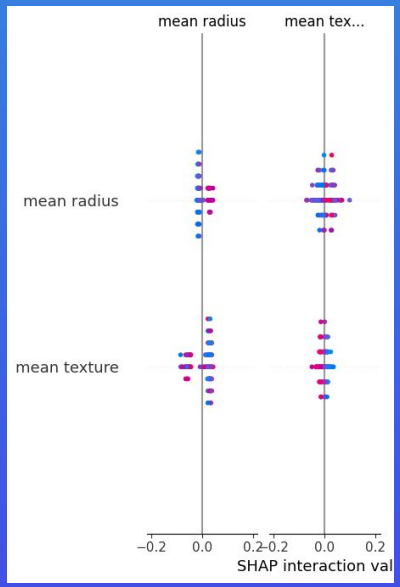
SHAP Waterfall Plot – Local Explanation

- Baseline on the right: E[f(X)] = average model output over all patients
- Blue blocks show how each feature moves us from baseline to this patient's prediction
- Rightward contributions increase malignancy probability; leftward decrease it
- In this example, high worst area and concavity-related features strongly push toward malignant



SHAP Interaction / Additional View

- SHAP can visualize interactions between features (e.g., mean radius × mean texture)
- Each point is a patient; color and position reflect joint influence on prediction
- Helps reveal nonlinear patterns that simple feature importance misses
- Complements LIME by showing how groups of patients behave, not only one case



LIME vs SHAP – Comparison

LIME:

- Pros: very interpretable (if/then style rules), easy to show to clinicians
- Cons: explanations can change with different sampling; no direct global picture

SHAP:

- Pros: theoretically grounded, consistent global rankings, rich plots (summary, waterfall, interaction)
- Cons: slower to compute; visualizations slightly harder to explain to non-technical audiences

Project Takeaways

- For this Breast Cancer model, both methods highlight similar key features (worst area, concavity, radius)
- LIME is great for case-by-case storytelling; SHAP is better for auditing the model as a whole
- Using both together provides a more robust explanation than either alone
- Next steps: measure explanation stability and extend to fairness / subgroup analysis