Trade-off: model performance vs. explanation quality
(Bubble size ≈ explanation robustness, 1 − MaxSensitivity)