Black box model

Interpretable AI method

Feature ranking $\hat{\mathbf{w}}$

Human interpretation

Feature name	Feature score
Flipper length	0.60
Bill length	0.25
Bill depth	0.10
Sex	0.05