



JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING

685.621 Algorithms for Data Science

Supervised Learning: Model Interpretability & Explainability

The Importance of Model Interpretability

- **Trust & Transparency:** Users need to understand why a model makes certain predictions.
- **Regulatory Compliance:** Industries like healthcare and finance require explainable AI.
- **Debugging & Improvement:** Interpretability helps identify model biases and errors.
- **Human-in-the-Loop Decision Making:** Decision-makers rely on interpretable outputs.

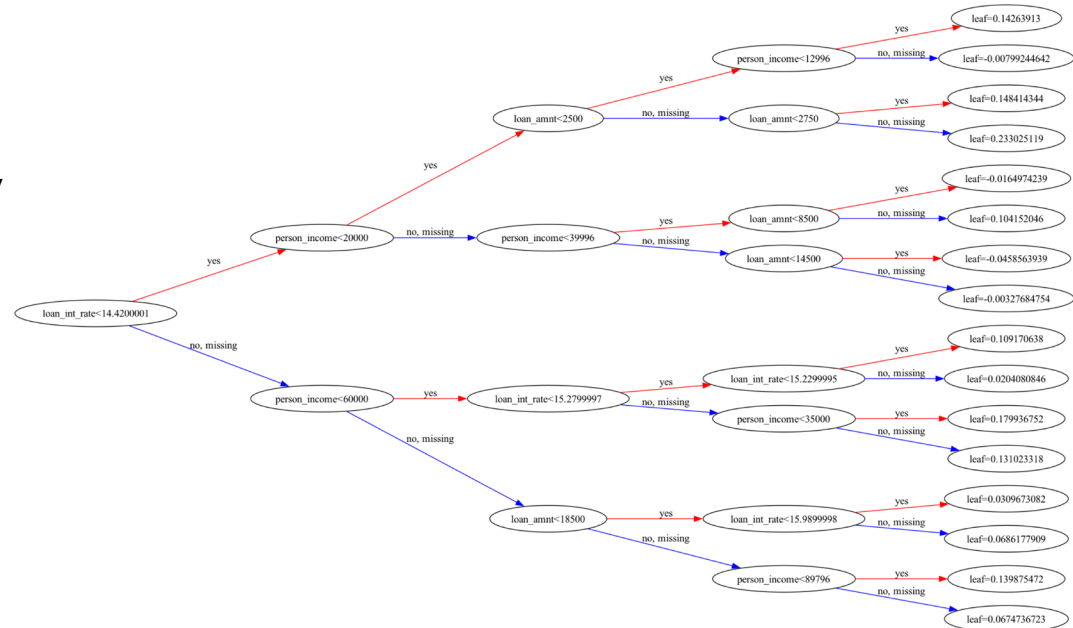
How Do We Interpret Models?

- **Intrinsic Interpretability:**

- Simple models like Decision Trees & Logistic Regression.
- Built-in transparency—directly explainable.

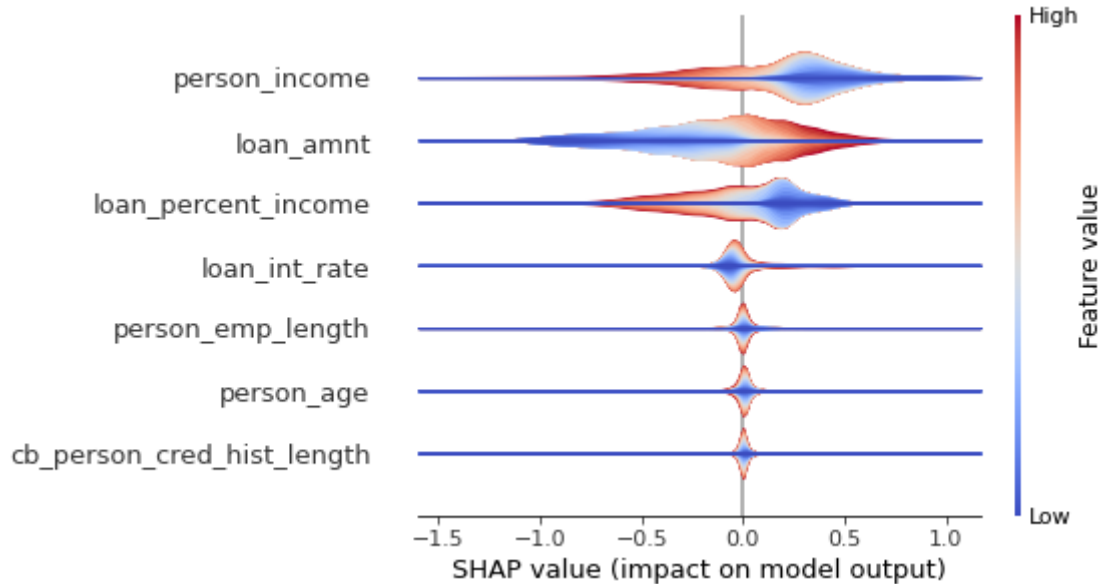
- **Post-Hoc Interpretability:**

- Needed for complex models like Neural Networks & Random Forests.
- Uses techniques like Feature Importance, SHAP, and LIME.

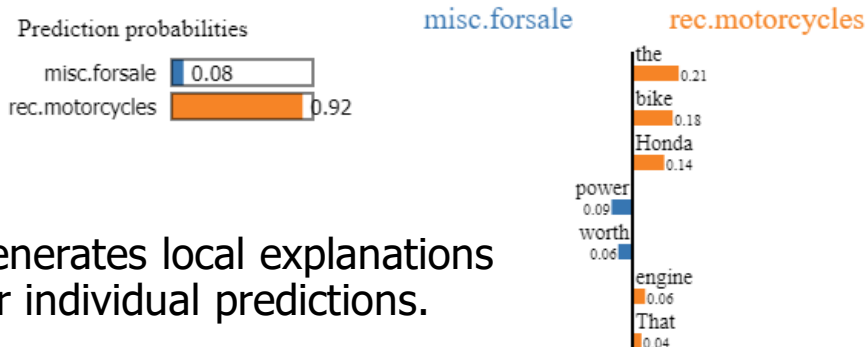


Explaining Predictions Using SHAP Values

- Provides global and local feature importance scores.
- Helps interpret black-box models like Random Forest and Neural Networks.
- The summary plot shows which features impact predictions the most.



LIME: Explaining Individual Predictions



- Generates local explanations for individual predictions.
- Creates a simplified linear approximation of the complex model.
- The LIME explanation shows how each feature contributes to a specific prediction.

Text with highlighted words

From: britt@leland.Stanford.EDU (Britt Park)
Subject: Power Loss under load of a Honda CB360T
Keywords: Honda CB360T power loss carburetion jets timing
Organization: DSG, Stanford University, CA 94305, USA
Lines: 24

I have a 1974 Honda CB360T which for most of my purposes runs well. However I experience a severe power drop at cruising speeds under load. That is, on a mild upgrade @ 50 mph in 4th or 5th I'm lucky if I can hold speed. If I try to add throttle much past 5000 rpm, power drops drastically. Put simply, under load, the engine won't rev past 5000 rpm. The top third/half of the throttle range is dead. Standing still the engine runs fine up to red line (9-10K).



JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING

© The Johns Hopkins University 2024, All Rights Reserved.