Project Proposal

Interpretability in Machine Learning

João Monteiro fc49821
Filipe Sousa fc52748

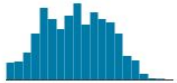
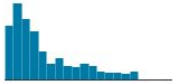
Goal

- Compare available Interpretability tools. (LIME & SHAP)
- How different algorithms learn from the same data.

Datasets

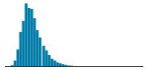
Customer Satisfaction in an airline company

- 130k Observations;
- 1 Target feature (Satisfaction);
- 4 Continuous Features;
- 18 Categorical / Boolean Features.

Gender	Age	Class	Flight Distance	satisfaction
Female 51% Male 49%	 7 85	Business 48% Eco 45% Other (1917) 7%	 31 4983	neutral or dissatisf... 56% satisfied 44%
Female	52	Eco	160	satisfied
Female	36	Business	2863	satisfied
Male	20	Eco	192	neutral or dissatisfied
Male	44	Business	3377	satisfied

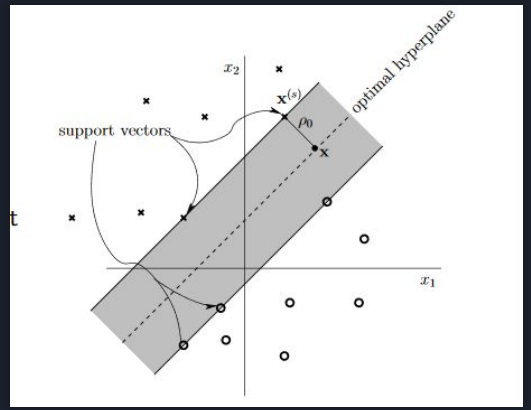
Hearth Disease Key Indicators

- 320K Observations;
- 1 Target feature (If Has Heart Disease)
- 4 Continous Features;
- 13 Categorical / Boolean Features.

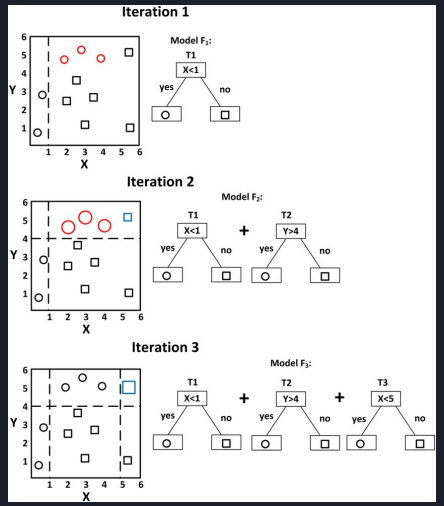
HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke
Respondents that have ever reported having coronary heart disease (CHD) or myocardial infarction (MI)	Body Mass Index (BMI)	Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes]	Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week	(Ever told) (you had) a stroke?
 true 27.4k 9% false 292k 91%	 12 94.8	 true 132k 41% false 188k 59%	 true 21.8k 7% false 298k 93%	 true 12.1k 4% false 308k 96%
No	16.6	Yes	No	No
No	20.34	No	No	Yes
No	26.58	Yes	No	No
No	24.21	No	No	No

Machine Learning Algorithms

Support Vector Machines

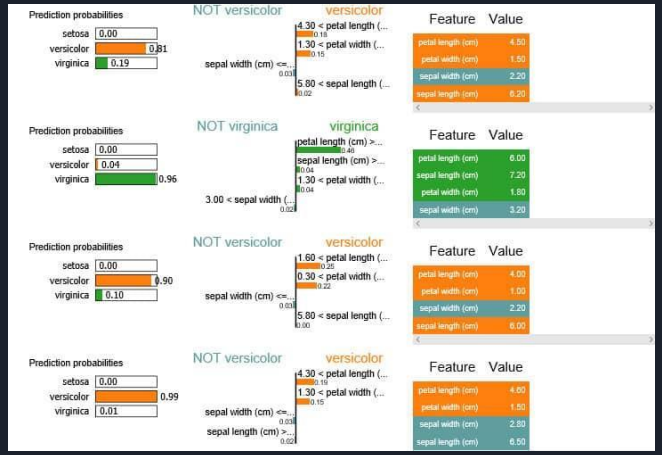


Gradient Boosts



SHAP & LIME

LIME
(Local Interpretable Model-Agnostic Explanations)



SHAP
(Shapley Addictive Explanation)