

DET. NORMAN LAPID



CASE FILE # M L 2

2021

A Deeper Investigation into the Predictors for Police Sentiment Scores

Using TreeInterpreter + other
Methods



TEAM
NINE-NINE

CASEFILE # ML2

Recap of Case File #ML1

Used different ML regressors to identify predictors for trust scores in law enforcement.

- Random Forest performed well on accuracy and runtime.

New Objective:

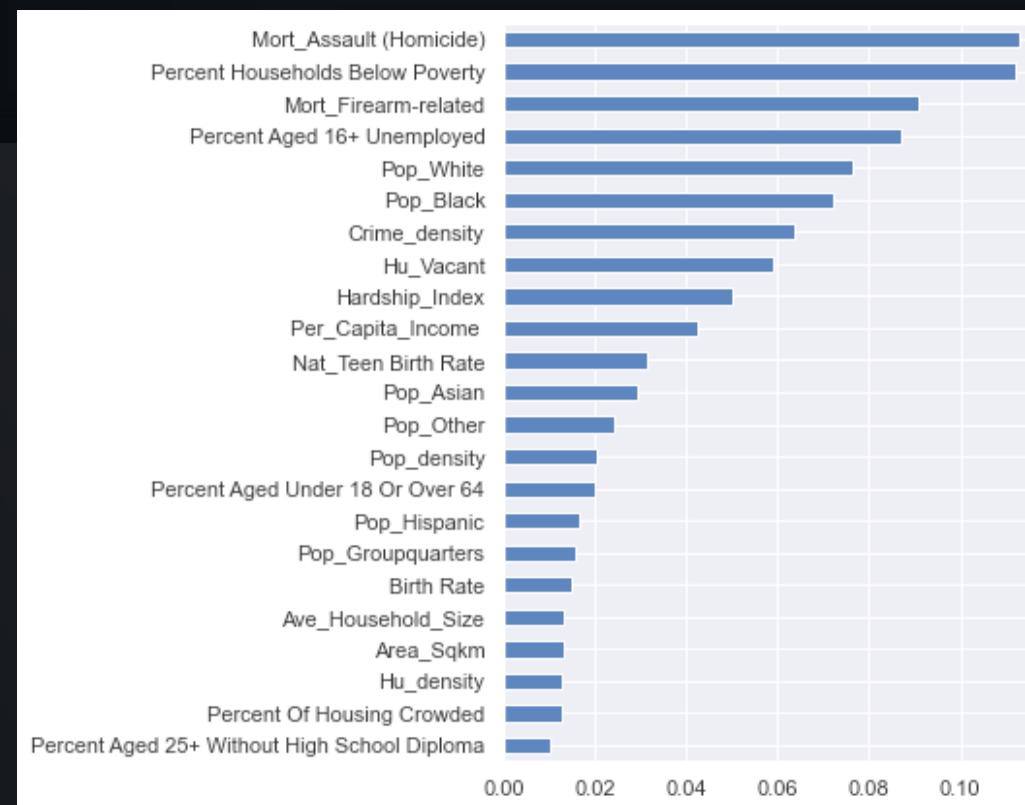
Apply interpretability methods to investigate a “black-box” Random Forest Model



GLOBAL INTERPRETABILITY |

Feature Importance

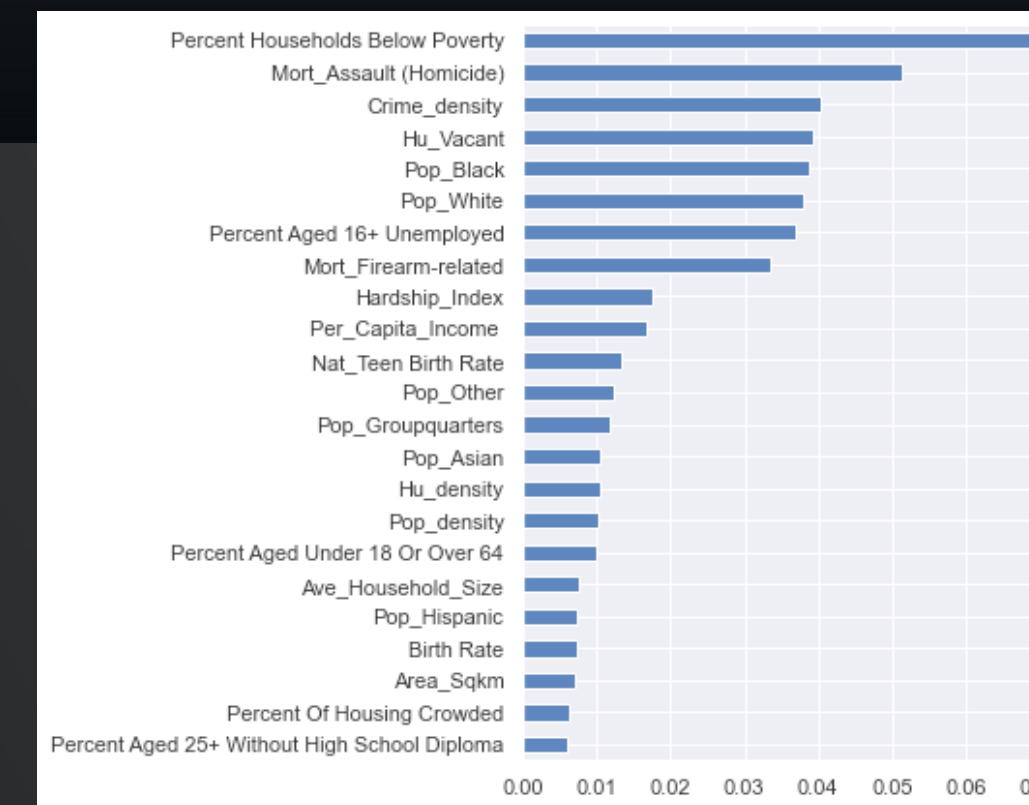
sklearn



Computed from
accumulation of the
impurity decrease within
each tree

Permutation Importance

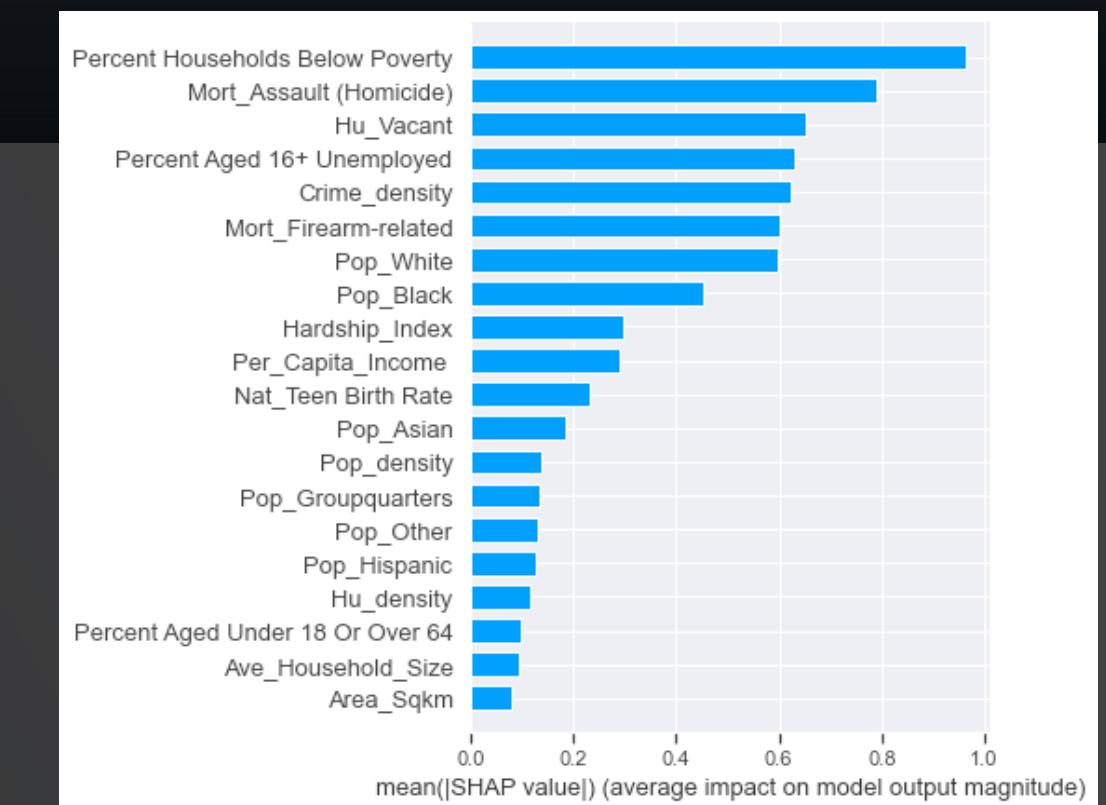
sklearn | eli5
manual implementation



Permute features and
compare accuracy against a
baseline

SHAP Summary Plot

shap



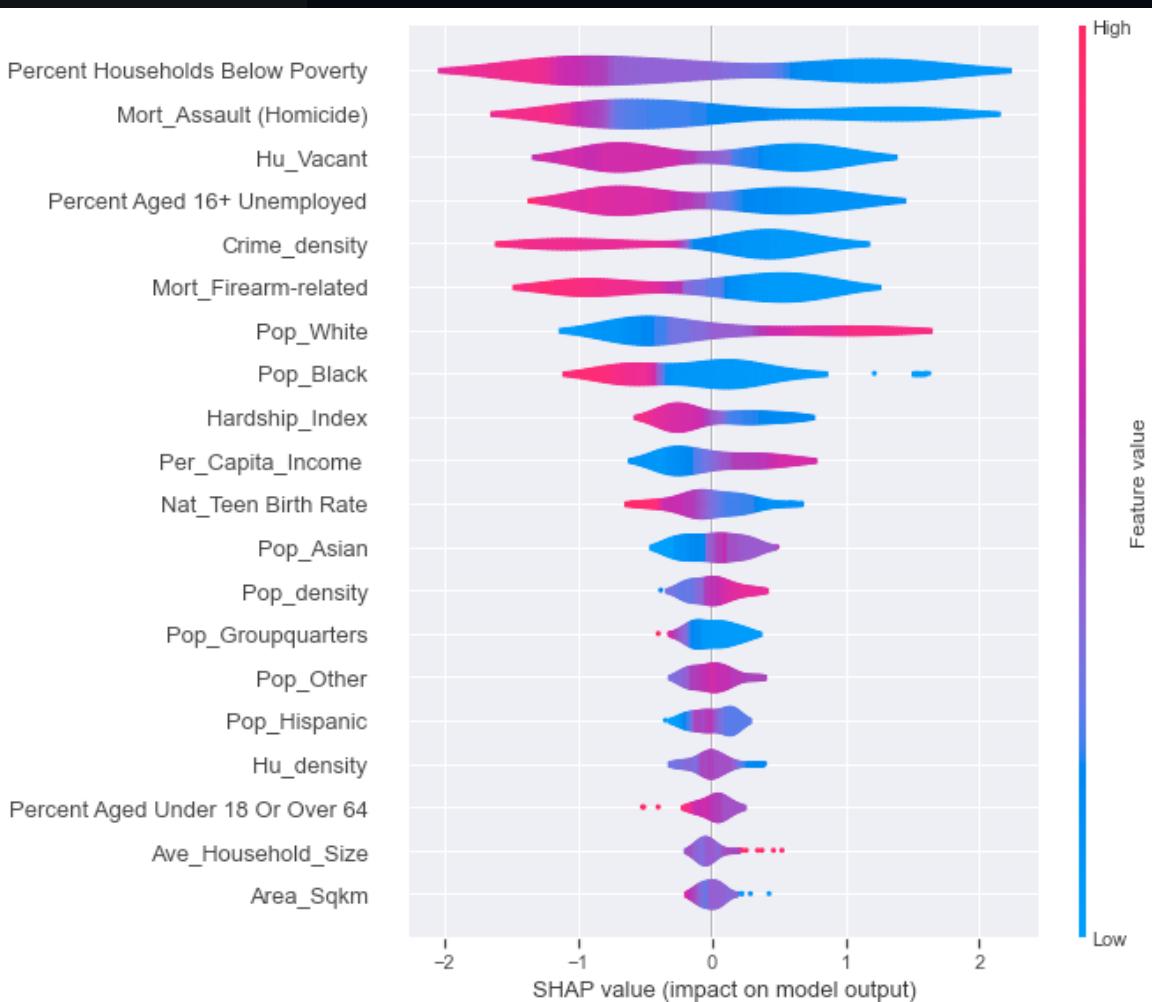
Mean absolute SHAP values
per feature across all
samples

MORE SHAP SUMMARY PLOTS

Beeswarm plot



Violin plot



Decision plot



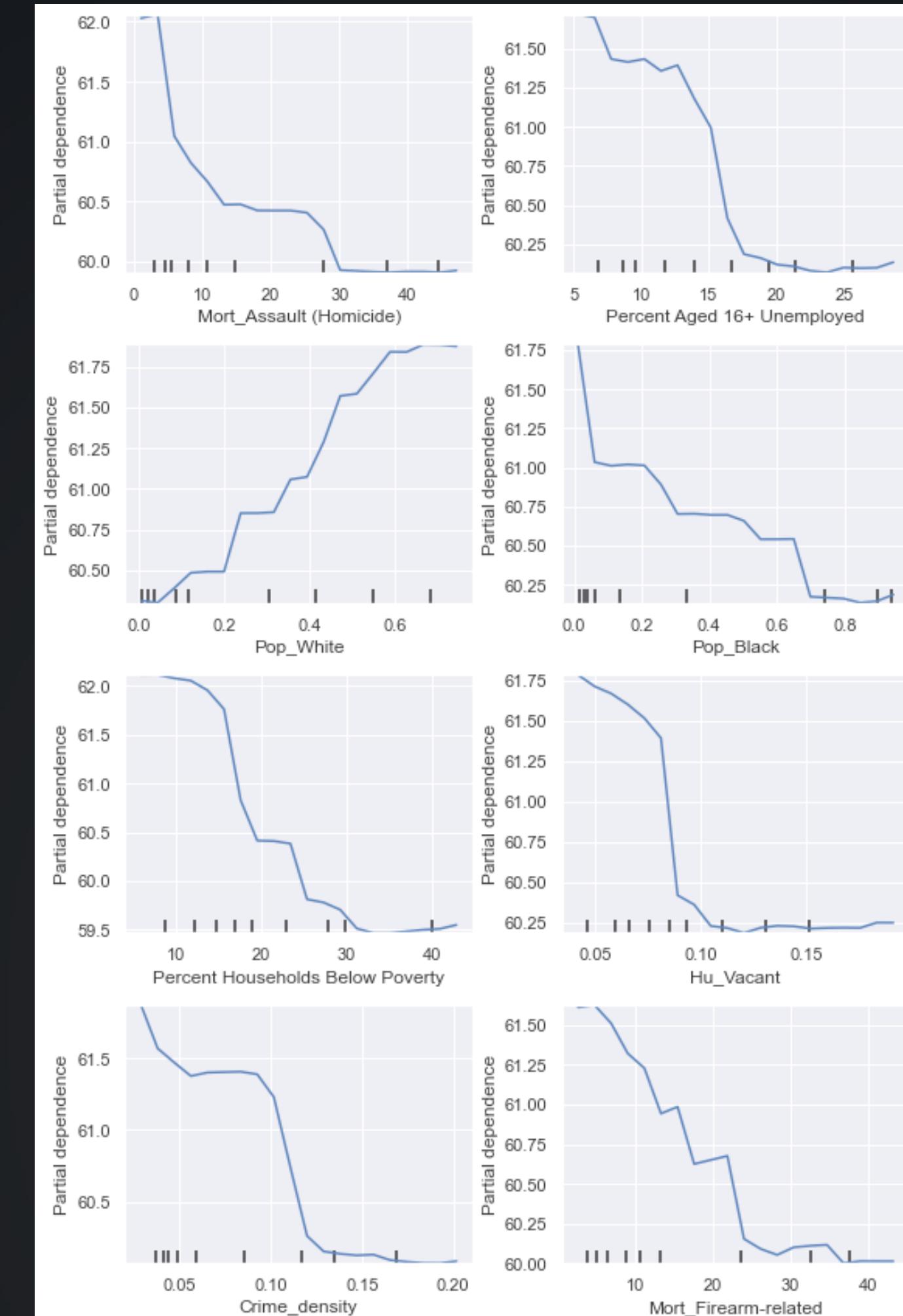
So far the evidence agrees...

Top predictors

- mortality from homicides/firearms
 - % unemployed
 - % households below poverty
 - vacant housing
 - crime density
 - % black population
-
- % white population

Trust

↑
Trust



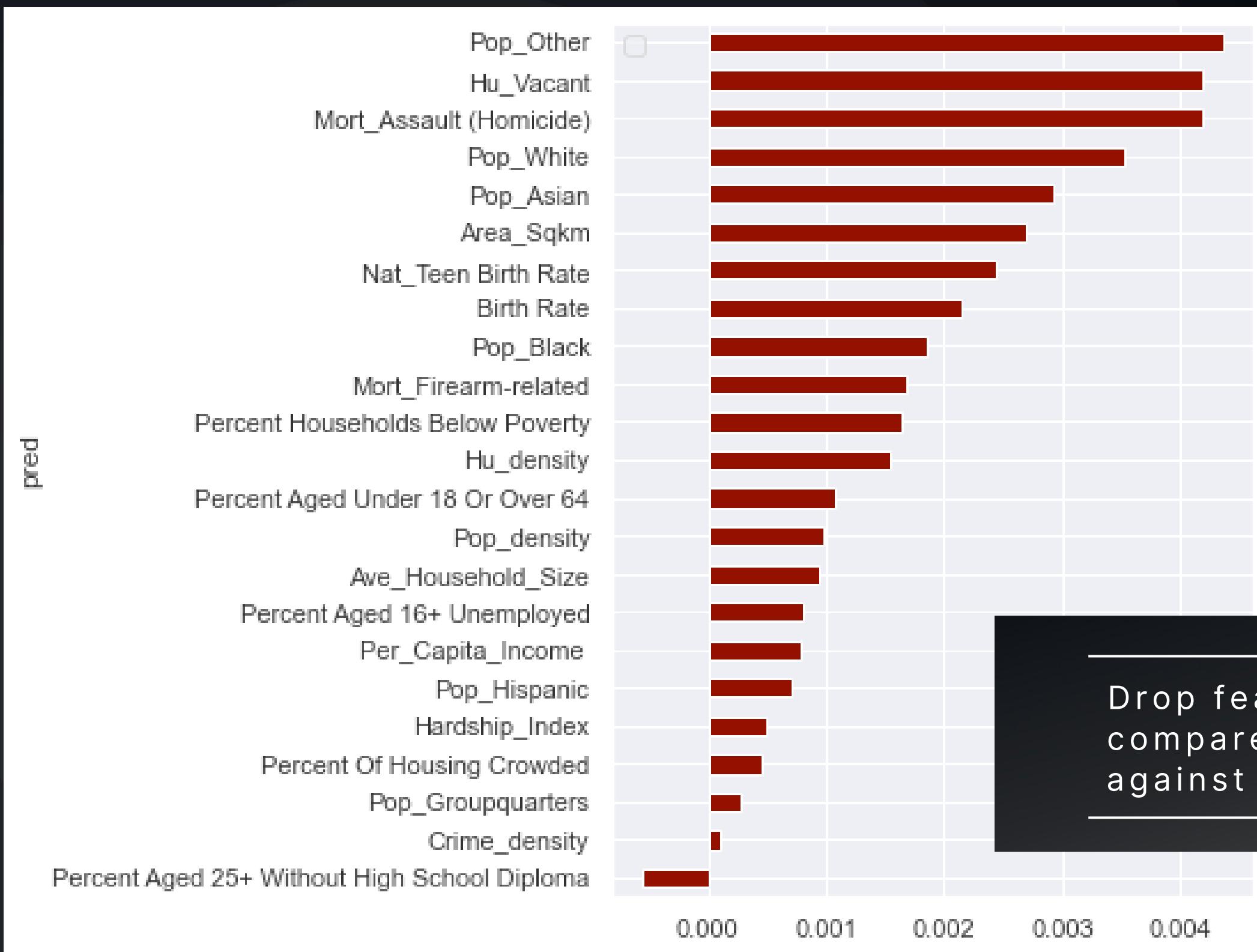
Partial
Dependence
Plots

Attempt "Drop Column Importance"

Results were not comparable since retraining on reduced features results in a different model

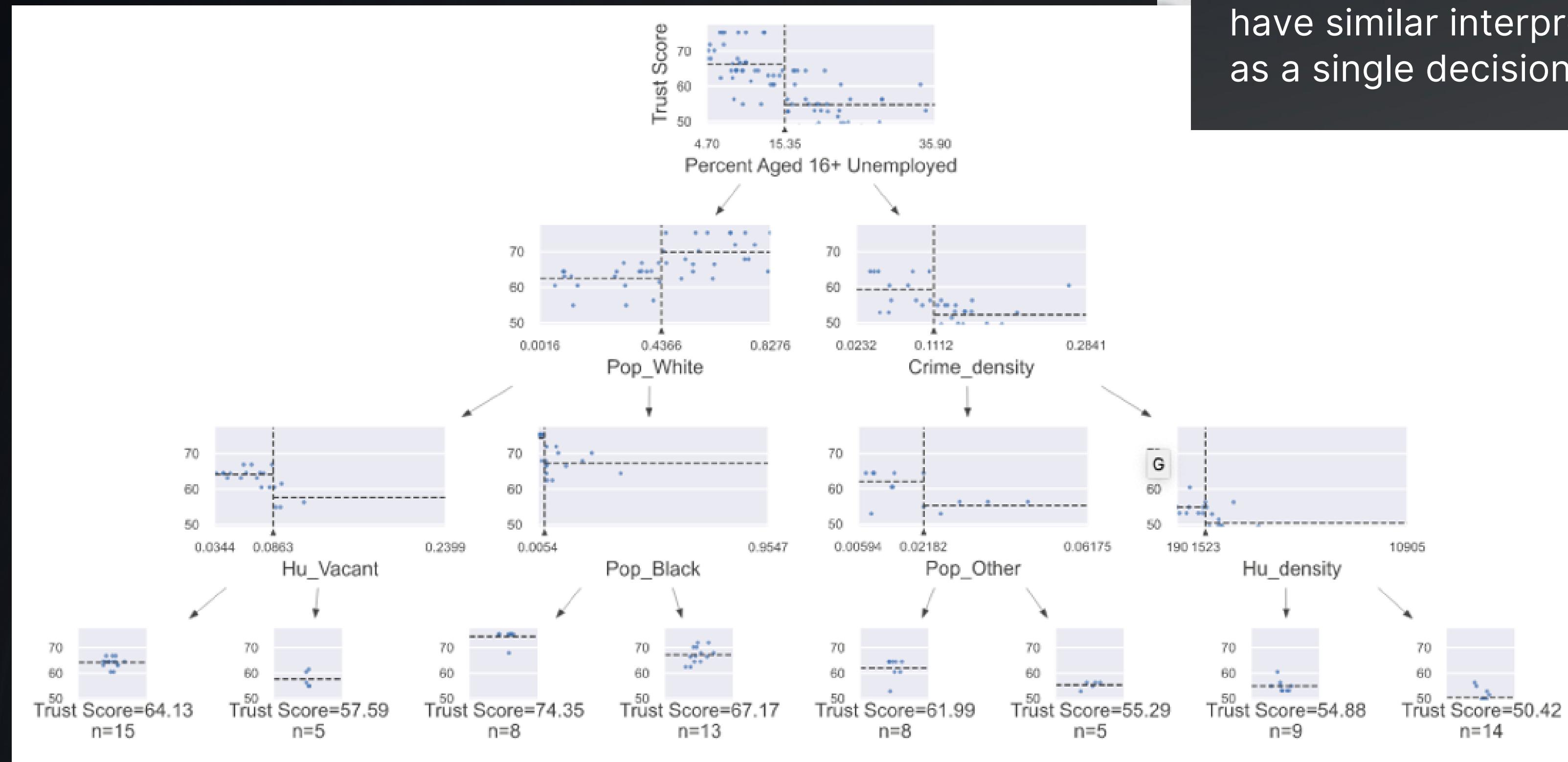
Negative importance

Removing these improves accuracy



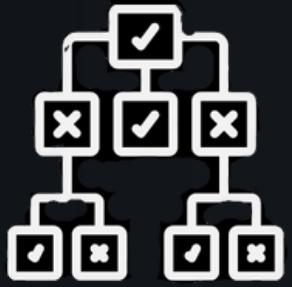
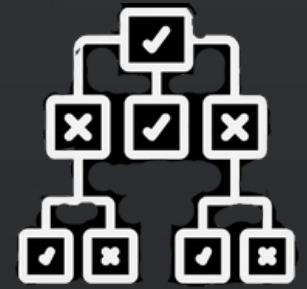
What about for Random Forests?

Ideally, we would want to have similar interpretability as a single decision tree

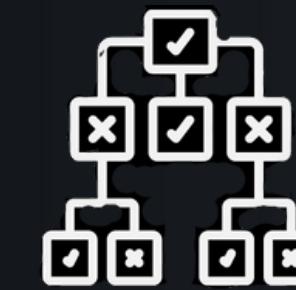
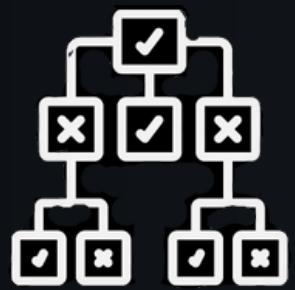


n_estimators of deep trees

Tree Interpreter



⋮



For a single tree, the prediction is the mean of the target variable (bias) plus the sum of the contributions at each node.

$$f(\mathbf{x}) = c_{full} + \sum_{k=1}^K contrib(\mathbf{x}, k)$$

Since the prediction of a forest is the average prediction for all trees:

$$F(\mathbf{x}) = \frac{1}{J} \sum_{j=1}^{JK} f_j(\mathbf{x})$$

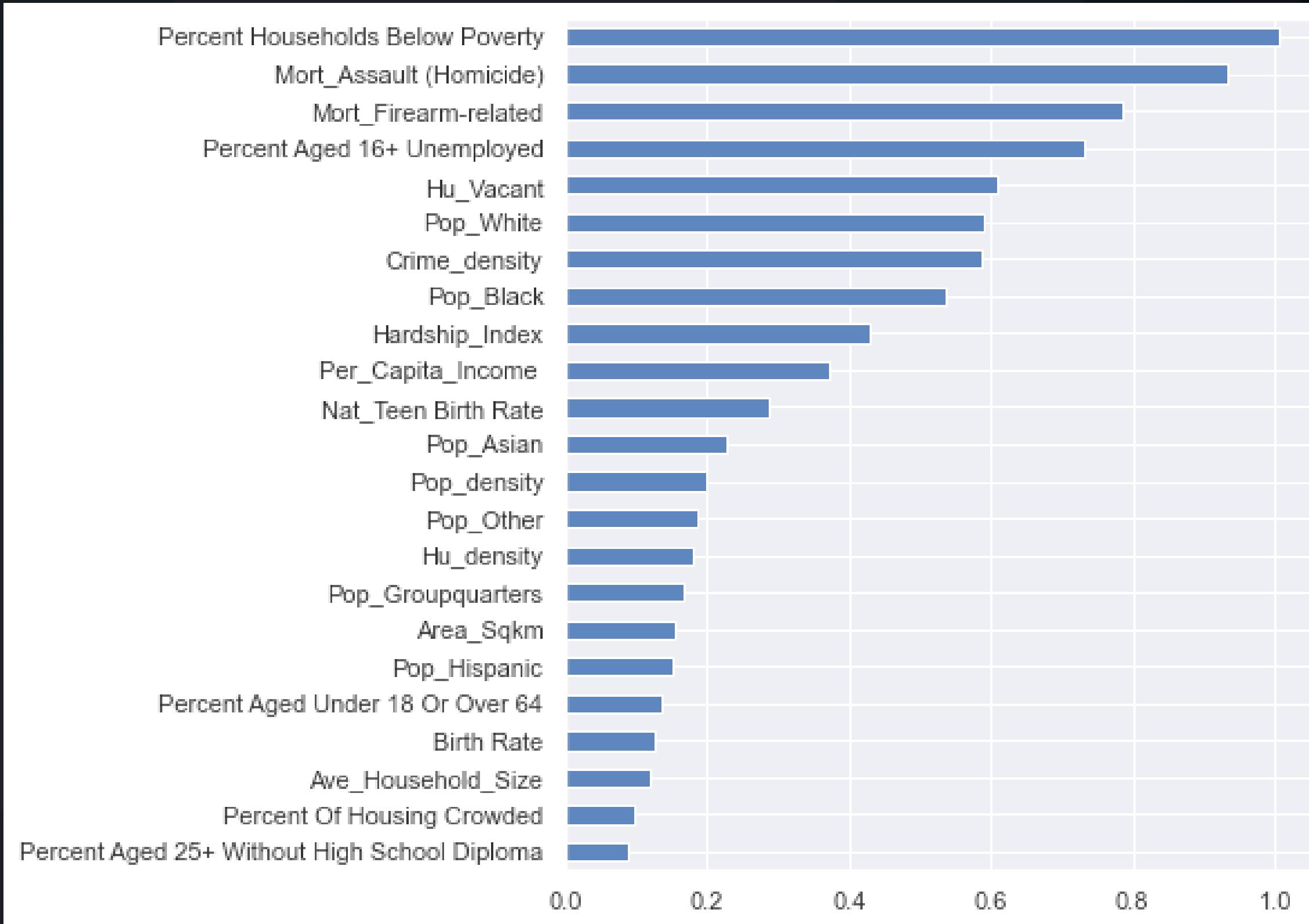
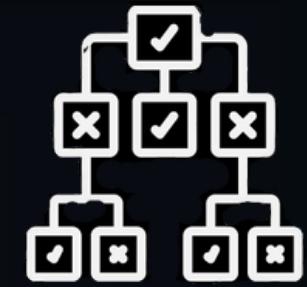
Then the prediction of the forest will equal the average of the bias terms plus the average contribution of each feature.

$$F(\mathbf{x}) = \frac{1}{J} \sum_{j=1}^J c_{jfull} + \sum_{k=1}^K \left(\frac{1}{J} \sum_{j=1}^J contrib_j(\mathbf{x}, k) \right)$$

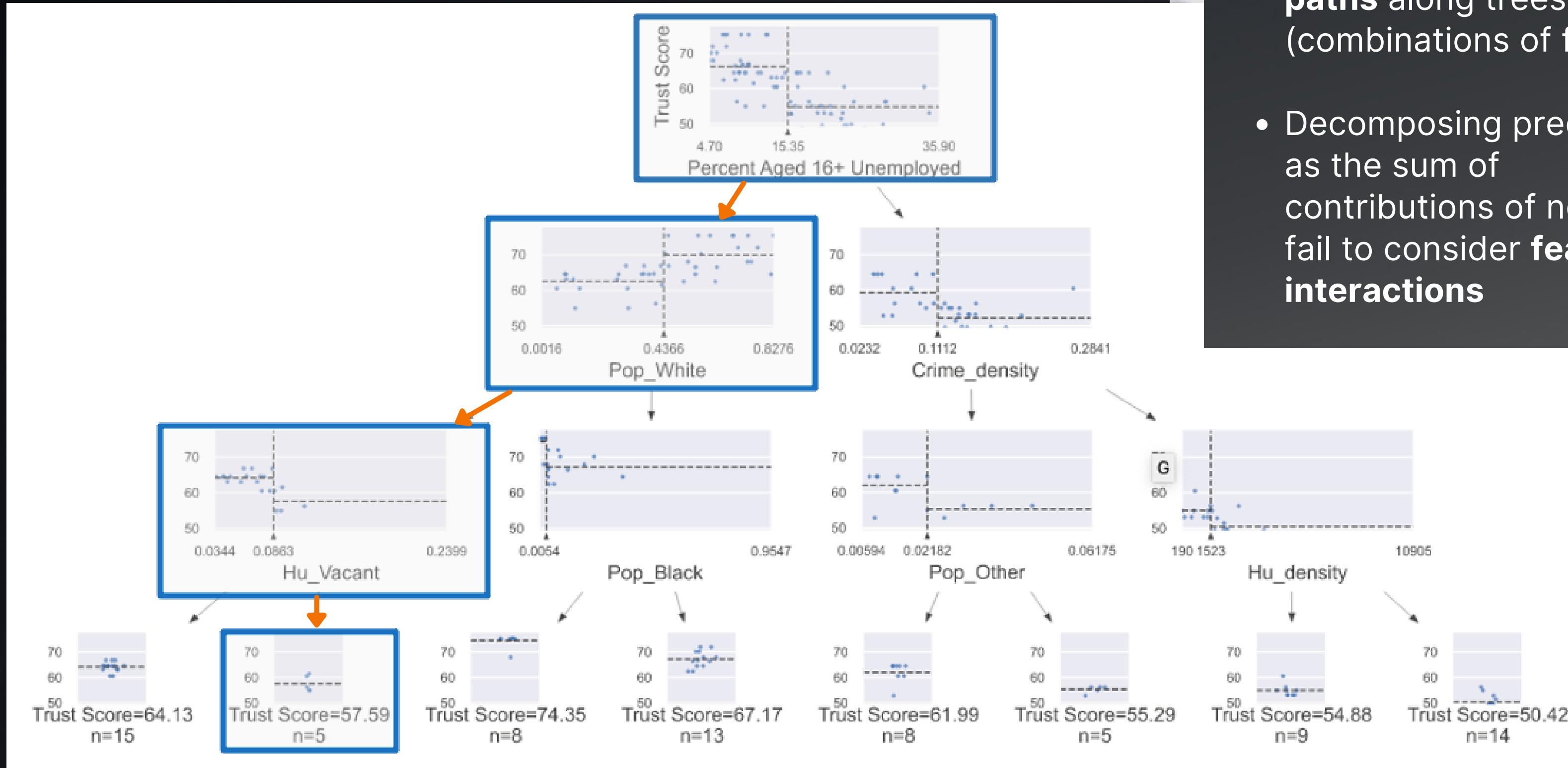
GLOBAL INTERPRETABILITY

Mean absolute contribution of each feature (similar to coeffecients in linear regression)

Tree Interpreter



Can we probe further?

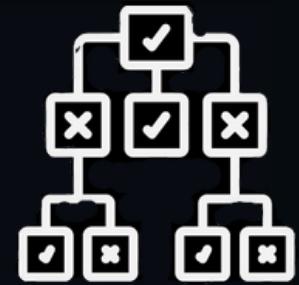


n_estimators of deep trees

- Predictions are made by **paths** along trees (combinations of features)
- Decomposing predictions as the sum of contributions of nodes will fail to consider **feature interactions**

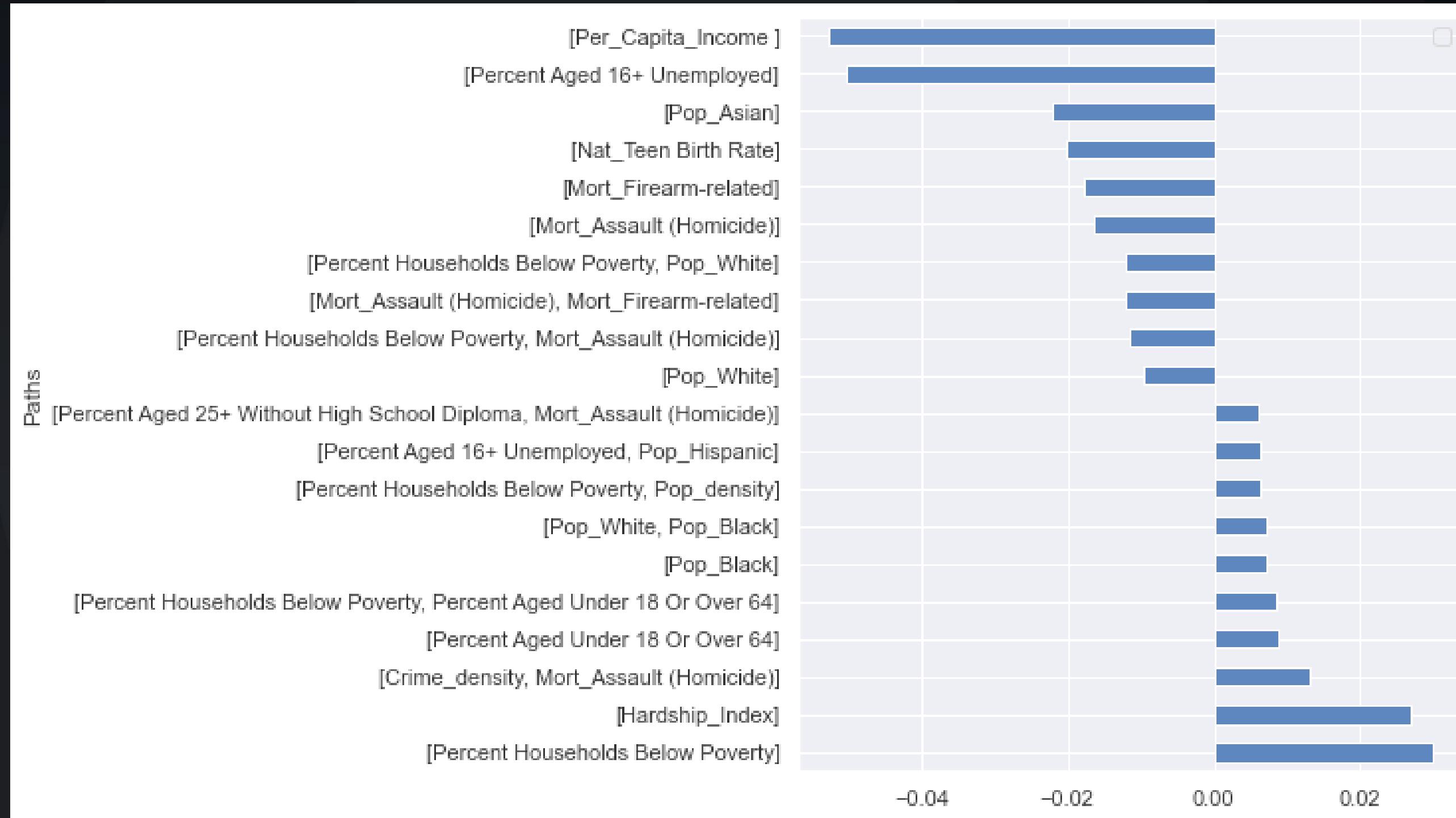
GLOBAL INTERPRETABILITY

Tree Interpreter



20 out of
2321
paths or
interactions

.....



Uses aggregated **joint contributions**, or contributions of **entire decision paths**.

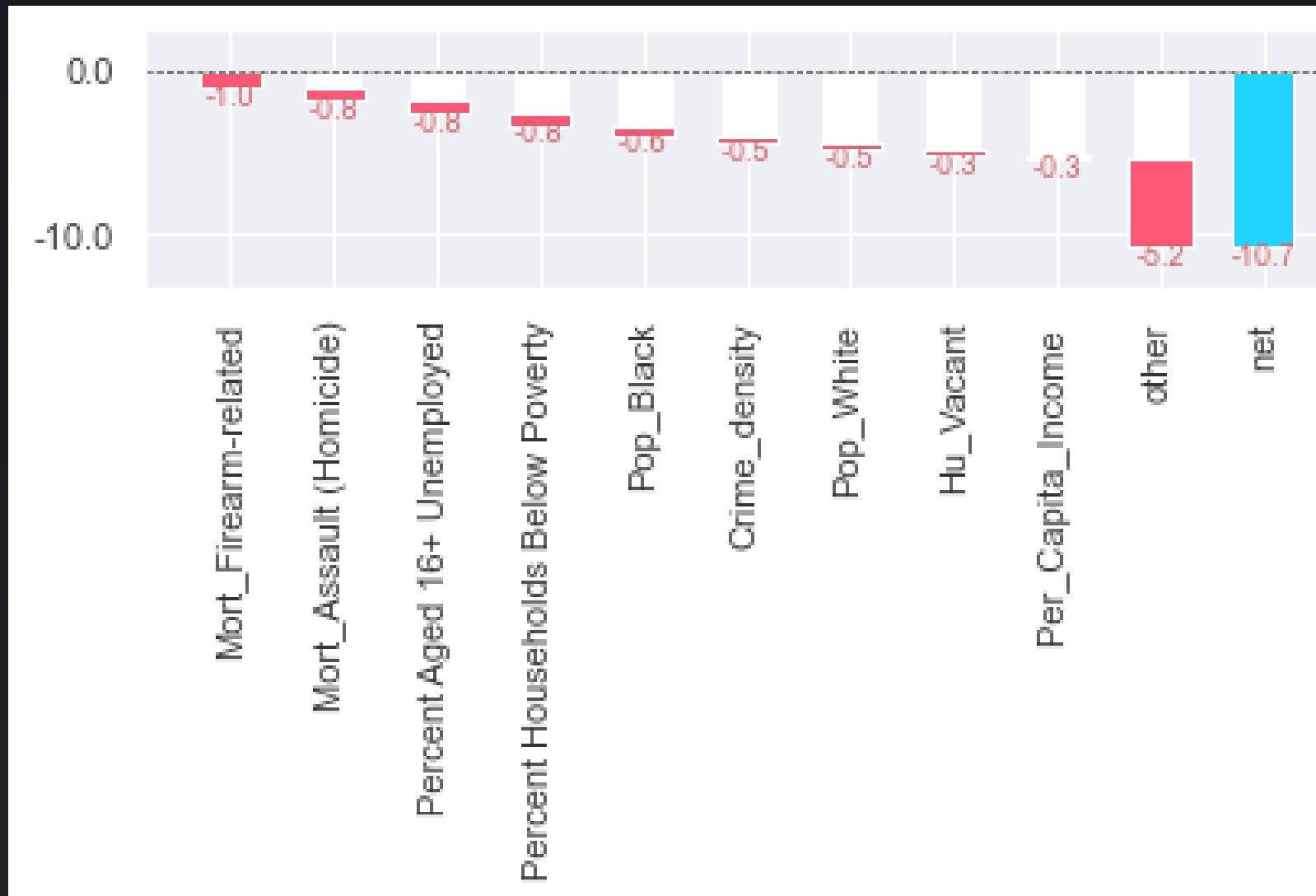
LOCAL INTERPRETABILITY

Tree Interpreter



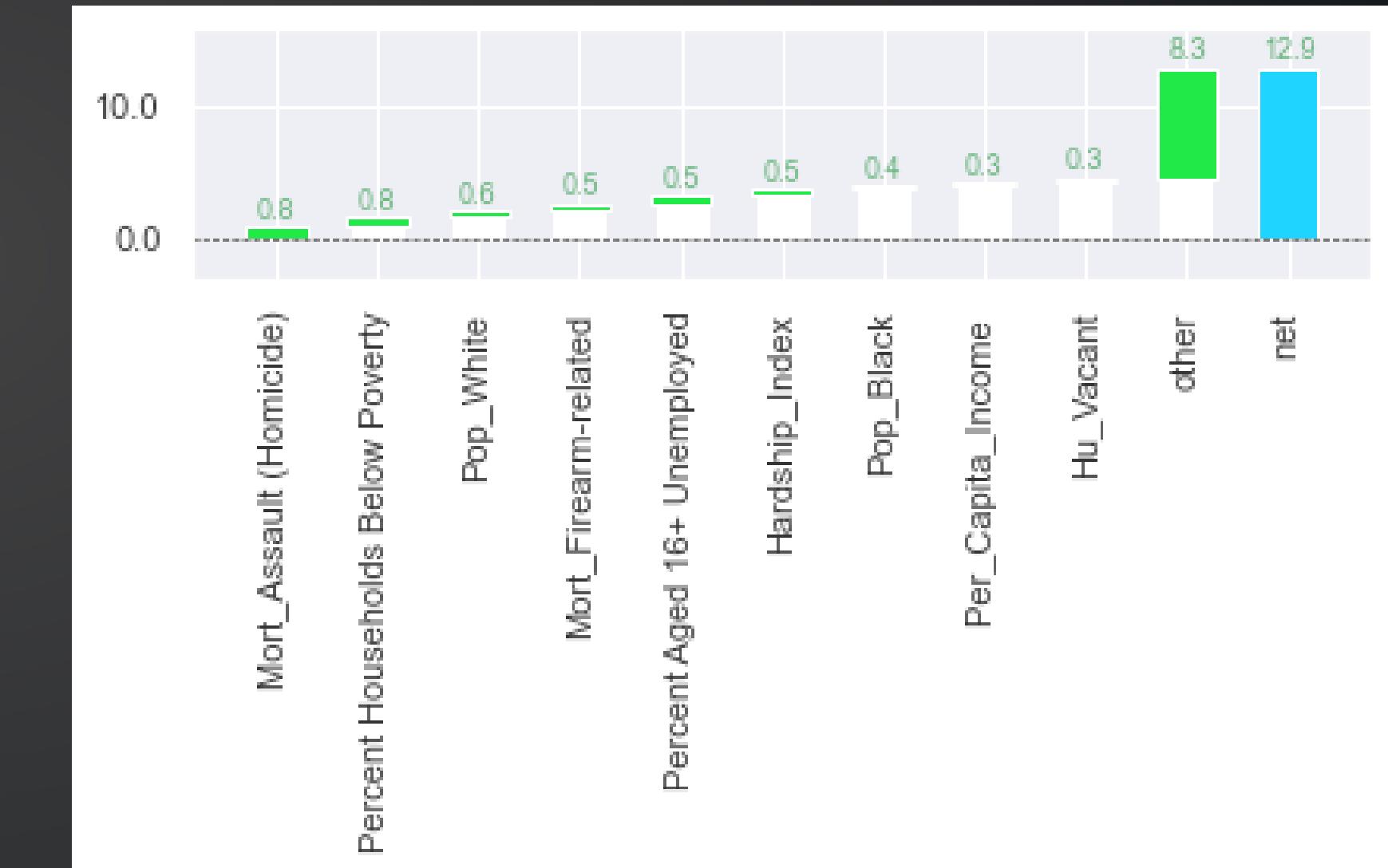
treeinterpreter + waterfall_chart

Englewood



413 paths/interactions

Norwood Park

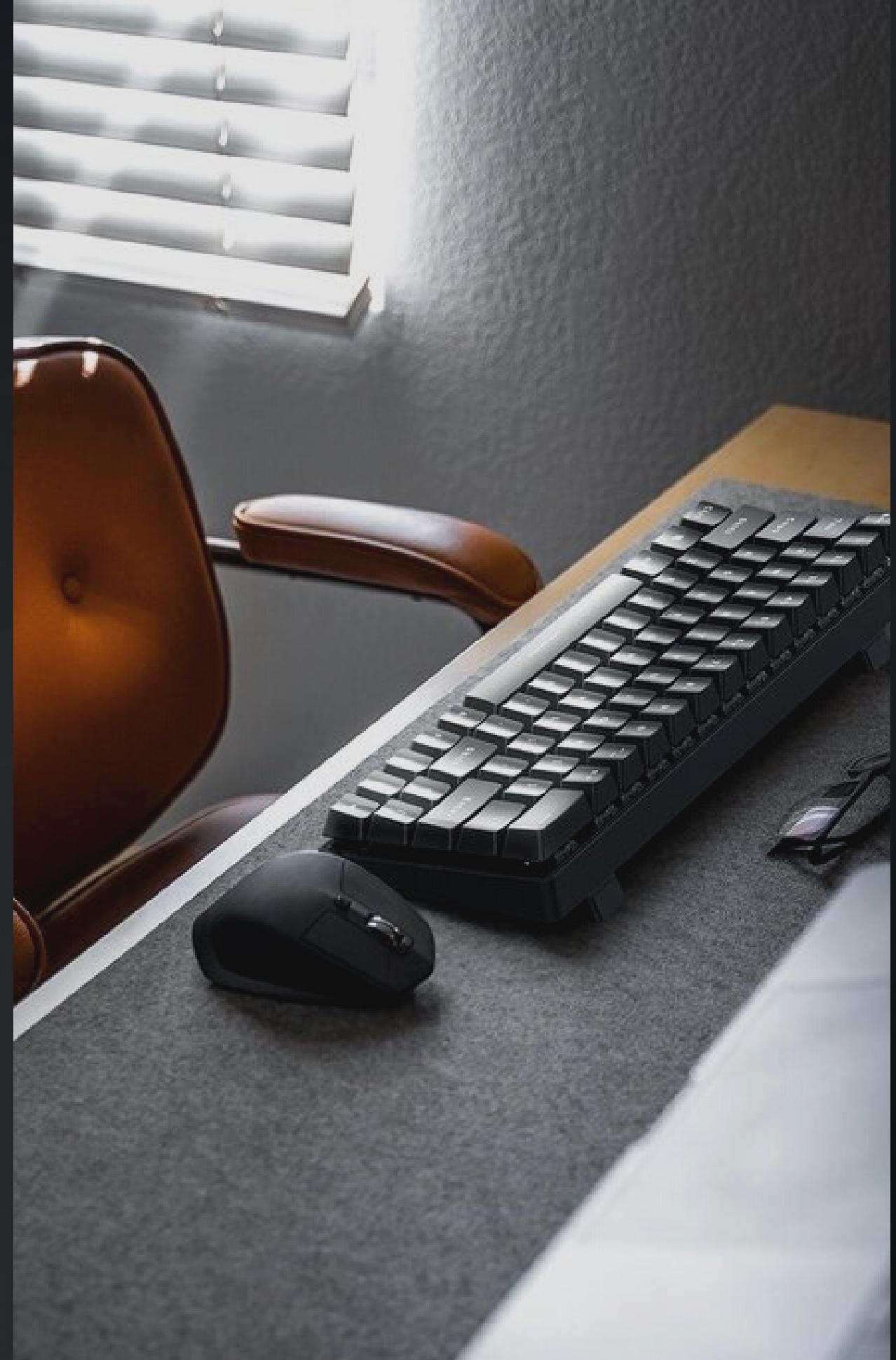


285 paths/interactions

CASEFILE # ML2

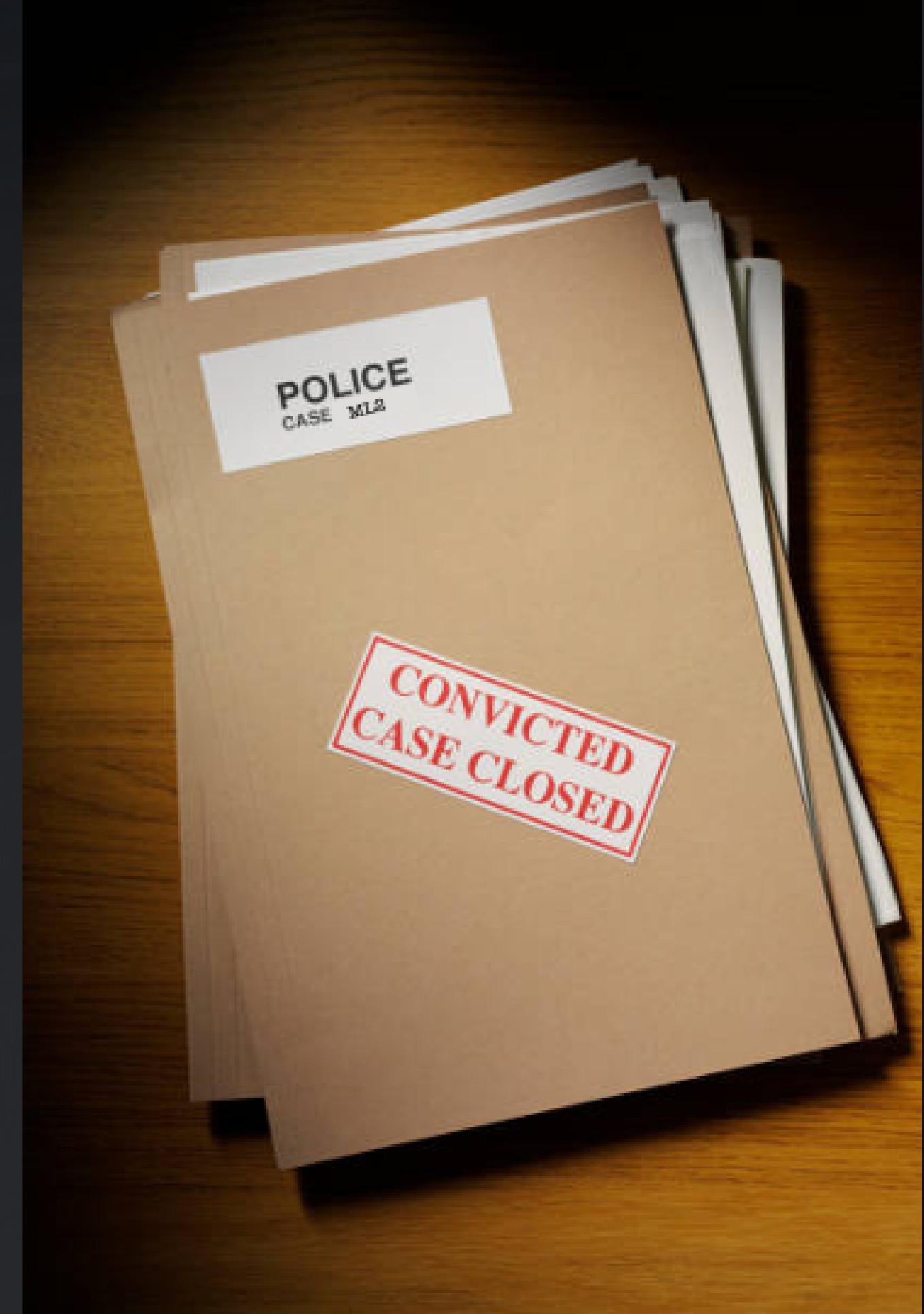
Investigation Findings:

- For sensitive social topics, **interpretability is key**
- **Tree Interpreter** can be used to explain the **contributions** of **interactions** of features in a **Random Forest model**
- Like SHAP, it can be used for both **local** and **global** interpretability



C A S E F I L E # M L 2

Thank you.



APPENDIX



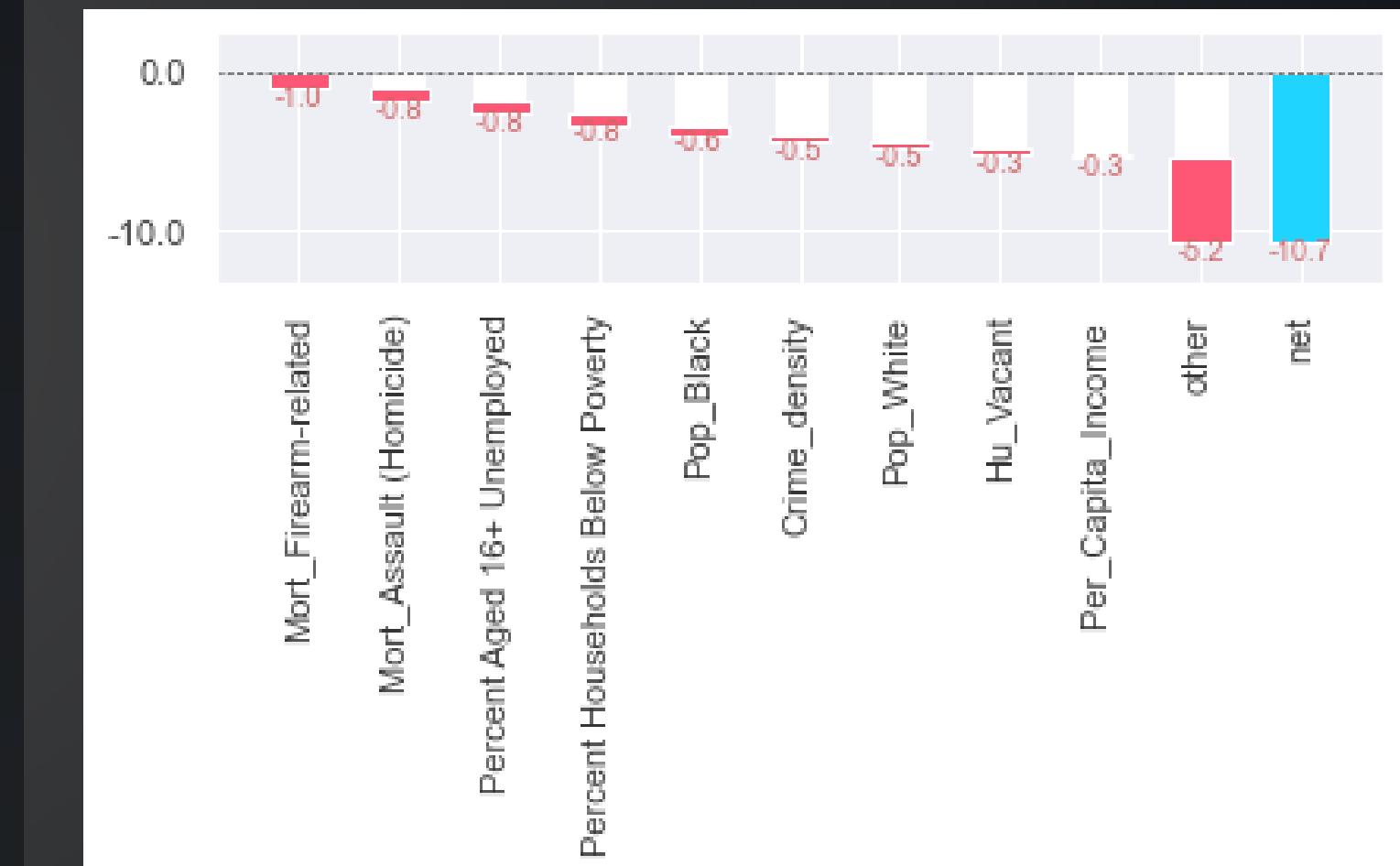
LOCAL INTERPRETABILITY

Englewood

treeinterpreter + waterfall_chart

	Contribution
	Paths
[Mort_Firearm-related]	-0.981614
[Mort_Assault (Homicide)]	-0.805818
[Percent Aged 16+ Unemployed]	-0.788044
[Percent Households Below Poverty]	-0.783304
[Pop_Black]	-0.585894
...	...
[Per_Capita_Income , Hardship_Index, Crime_density, Pop_White, Hu_Vacant]	0.047200
[Percent Households Below Poverty, Percent Aged 16+ Unemployed, Pop_Groupquarters, Pop_Other, Nat_Teen Birth Rate]	0.049660
[Hardship_Index, Hu_Vacant, Pop_density, Mort_Firearm-related]	0.057254
[Percent Households Below Poverty, Hardship_Index, Crime_density, Pop_Groupquarters]	0.069295
[Per_Capita_Income , Hu_Vacant, Mort_Firearm-related]	0.089437

413 rows × 1 columns



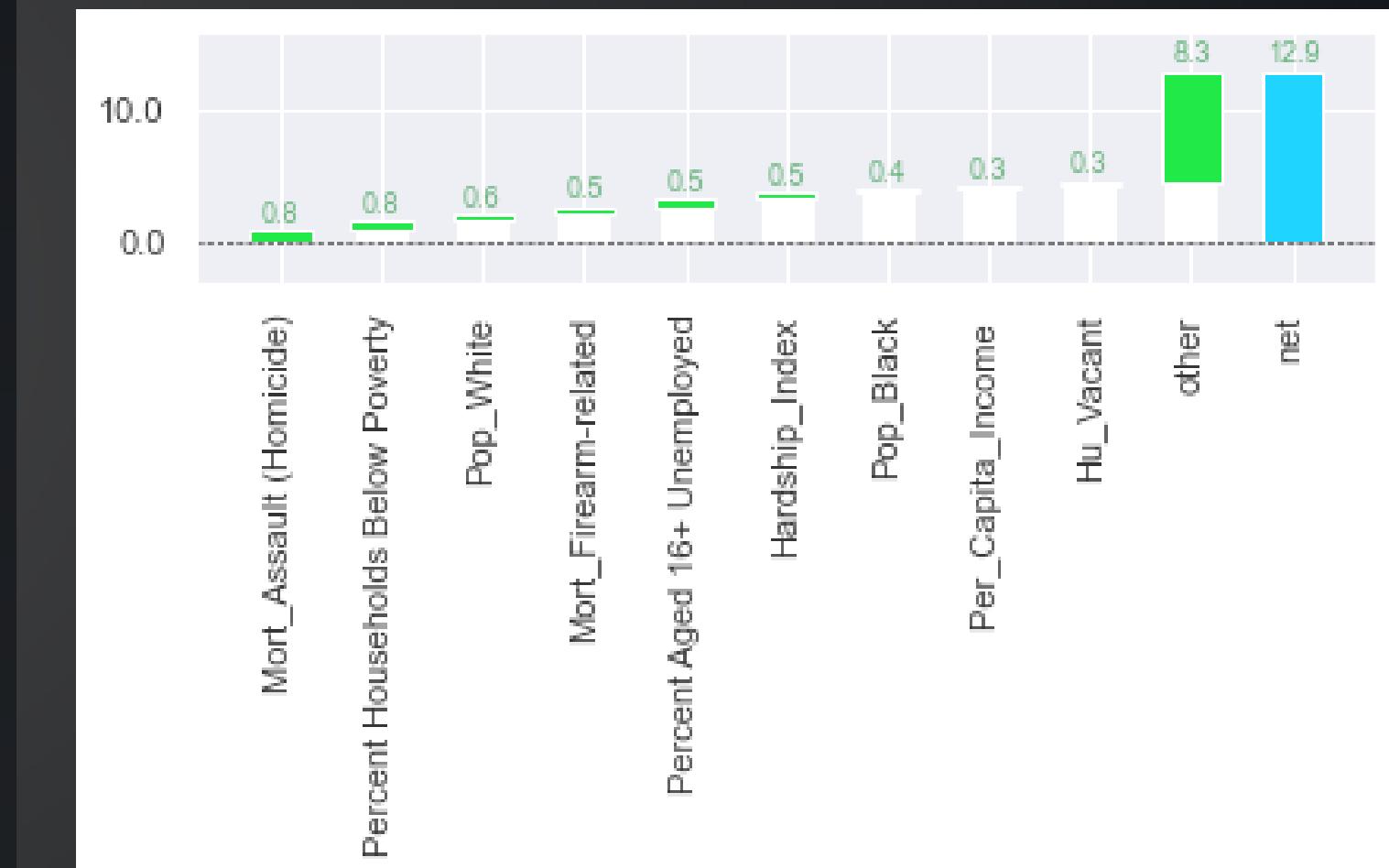
LOCAL INTERPRETABILITY

Norwood Park

treeinterpreter + waterfall_chart

	Contribution
	Paths
[Mort_Assault (Homicide)]	0.794125
[Percent Households Below Poverty]	0.769954
[Pop_White]	0.583119
[Mort_Firearm-related]	0.525672
[Percent Aged 16+ Unemployed]	0.512026
...	...
[Percent Aged Under 18 Or Over 64, Hu_Vacant]	-0.023214
[Percent Aged 16+ Unemployed, Pop_Groupquarters, Pop_White]	-0.024076
[Percent Households Below Poverty, Pop_Groupquarters, Pop_Other, Mort_Assault (Homicide), Mort_Firearm-related]	-0.038221
[Percent Households Below Poverty, Percent Aged Under 18 Or Over 64, Pop_White]	-0.041196
[Per_Capita_Income , Hardship_Index, Pop_Asian, Birth Rate]	-0.099282

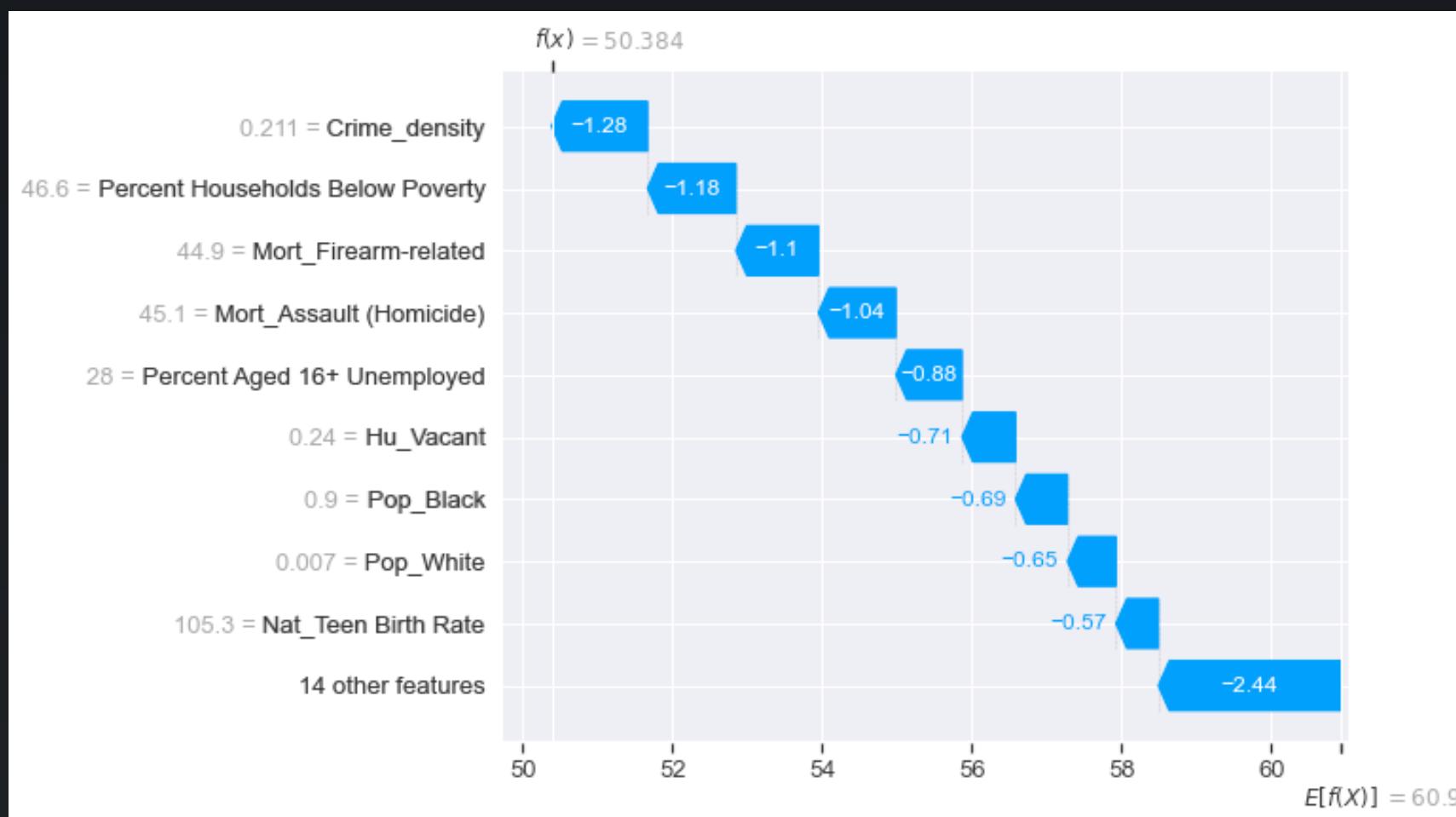
285 rows × 1 columns



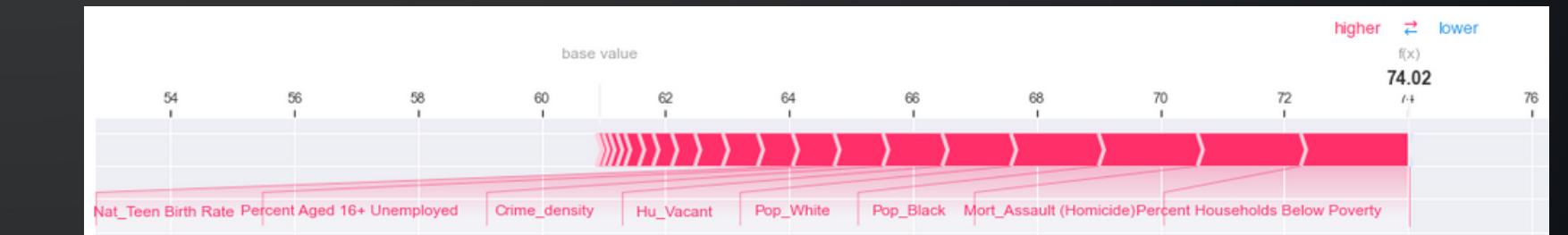
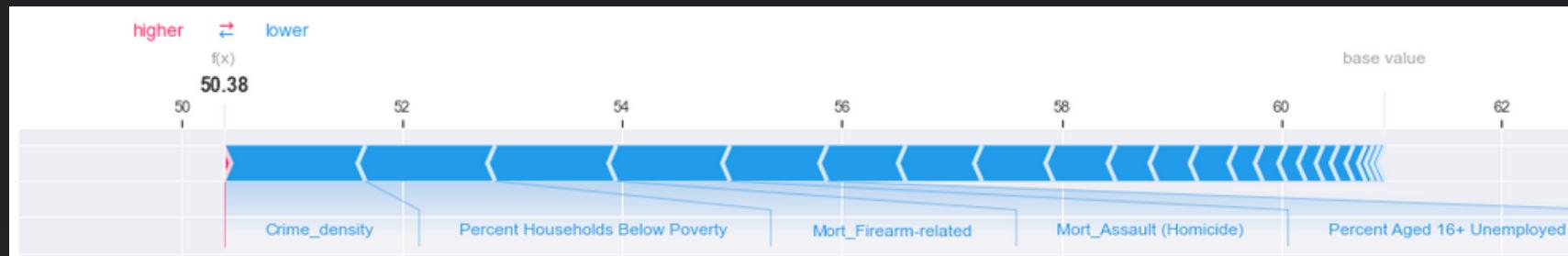
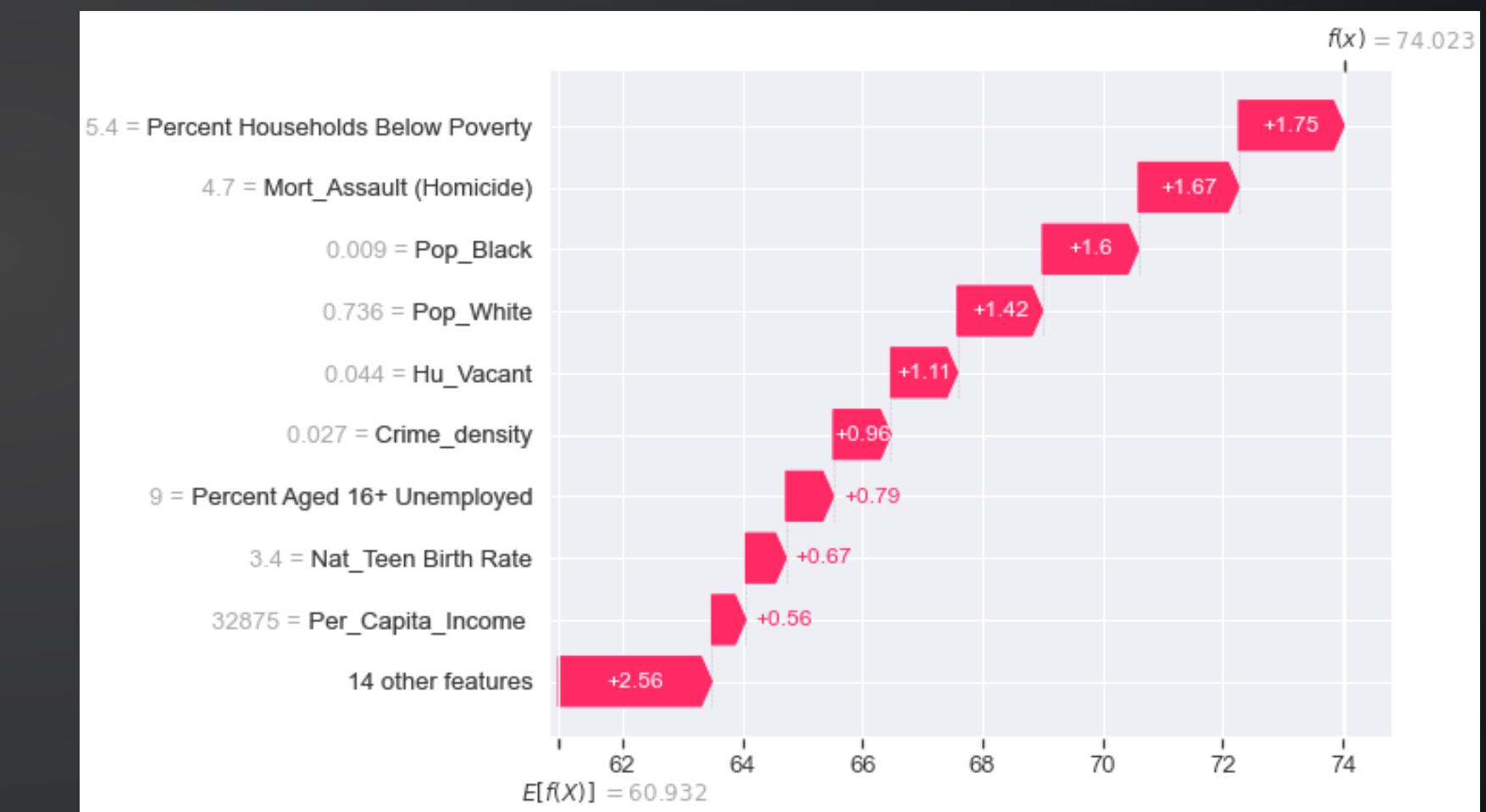
LOCAL INTERPRETABILITY

shap.waterfall_plot + shap.force_plot

Englewood



Norwood Park



NORMAN



CASE FILE # M L 2



2021

**TEAM
NINE-NINE**



CHICAGO P.D.

CASE FILE # M L 2

2021

NEW YORK P.D.



A Deeper Investigation into the Predictors for Police Sentiment Scores

Using TreeInterpreter + other Methods

NORMAN LAPID

.....