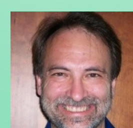
DATA+AI SUMMIT 2022

X-FIPE

e**X**tended **F**eature **I**mpact for **P**rediction **E**xplanation


Xingde Jiang
Machine Learning Engineer, Humana


Steve Brunner
Director of AI, Humana

ORGANIZED BY databricks

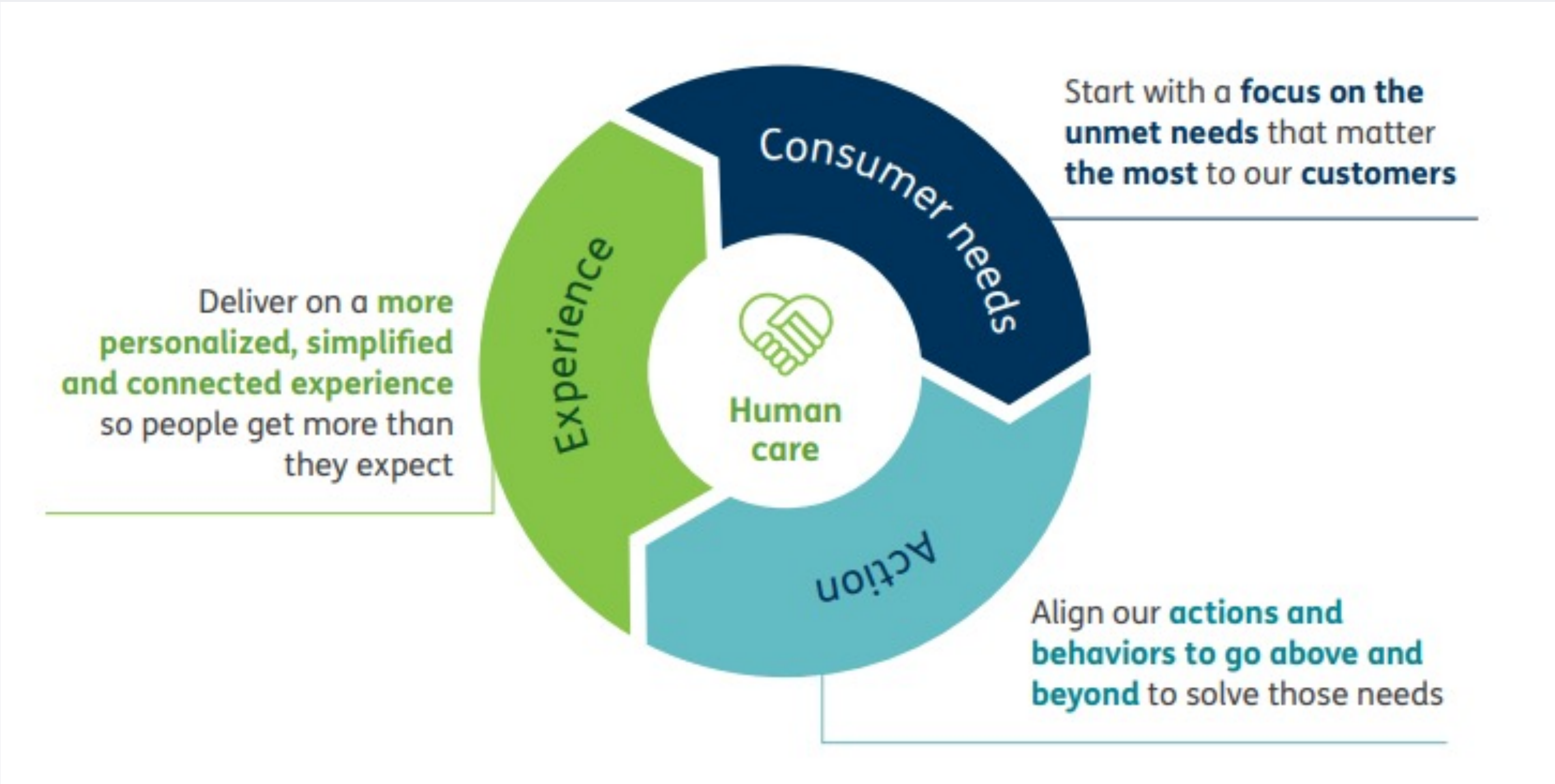
About Humana

Bringing you the human side of healthcare.

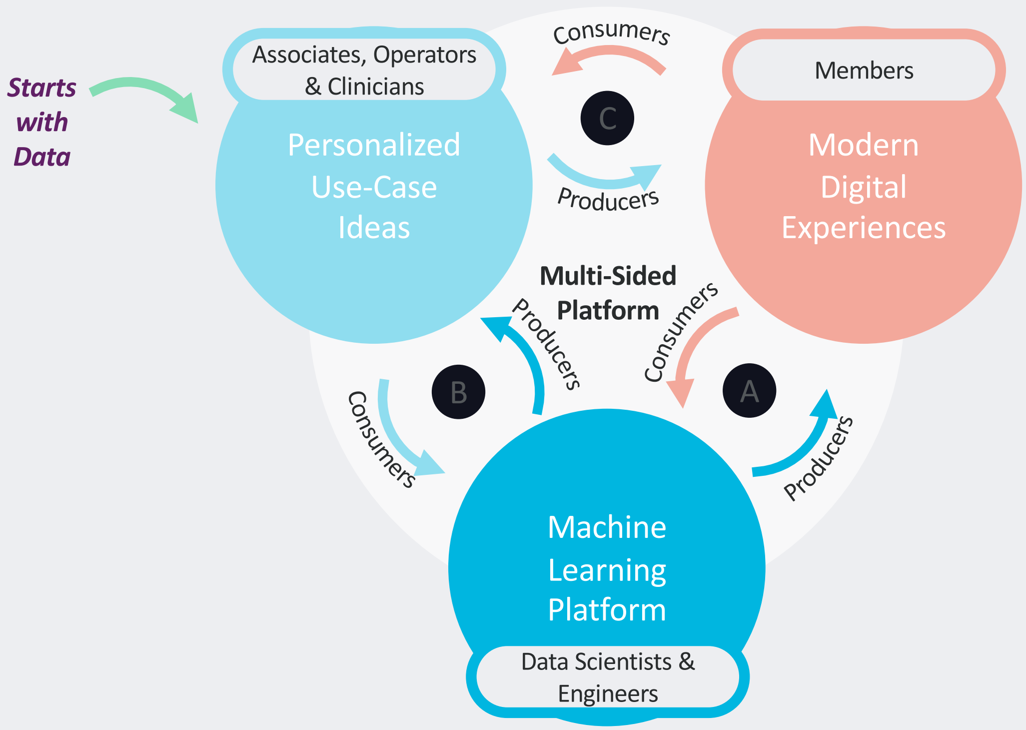
At Humana, we're dedicated to improving the health outcomes of every person we serve. We believe everyone should have equitable access to the tools and support they need to be as healthy as possible -- support that's personalized and easy to use.

Humana®

Delivering human care relies on aligning our actions and behaviors to build trust with our members



MLP - Leveraging multi-sided platform ecosystem to accelerate value creation



Enable Personalized Experiences & Improve Feedback Loop

A — **Data Scientists** produce personalization **models** enabling seamless **member** experience; Member interactions captured as **data**, feeding back to data scientists enabling increased model accuracy

Improve Analytic Strength & Increase Analytics Availability

B — **Data Scientists** create **models** for new use-cases for **associates** to use to help member achieve their best health.

Attract More Members & Generate More Data

C — **Operators and Data Scientists** co-create **models**, driving test and learns, better and more personalized experiences for members, which increases consumer trust, drives more member acquisition, improves retention and health outcomes

The Network Effect: The more consumers and producers on the platform, the more interactions and data that can be collected and leveraged to improve health outcomes through advanced analytics and modern digital experience

Ethical AI Governance

Industry practice that **peaks interest of key investors, and regulators**

AI Principles set forward a high level collection of philosophies that should be guiding the development of advanced analytic systems

Designed to be released to the general public to **increase the awareness of how carefully customer's data is used**

Principles are the result of a **extensive collaboration** across Enterprise Data and Analytics, Enterprise Risk Management, Legal, HR, Clinical, and Marketing

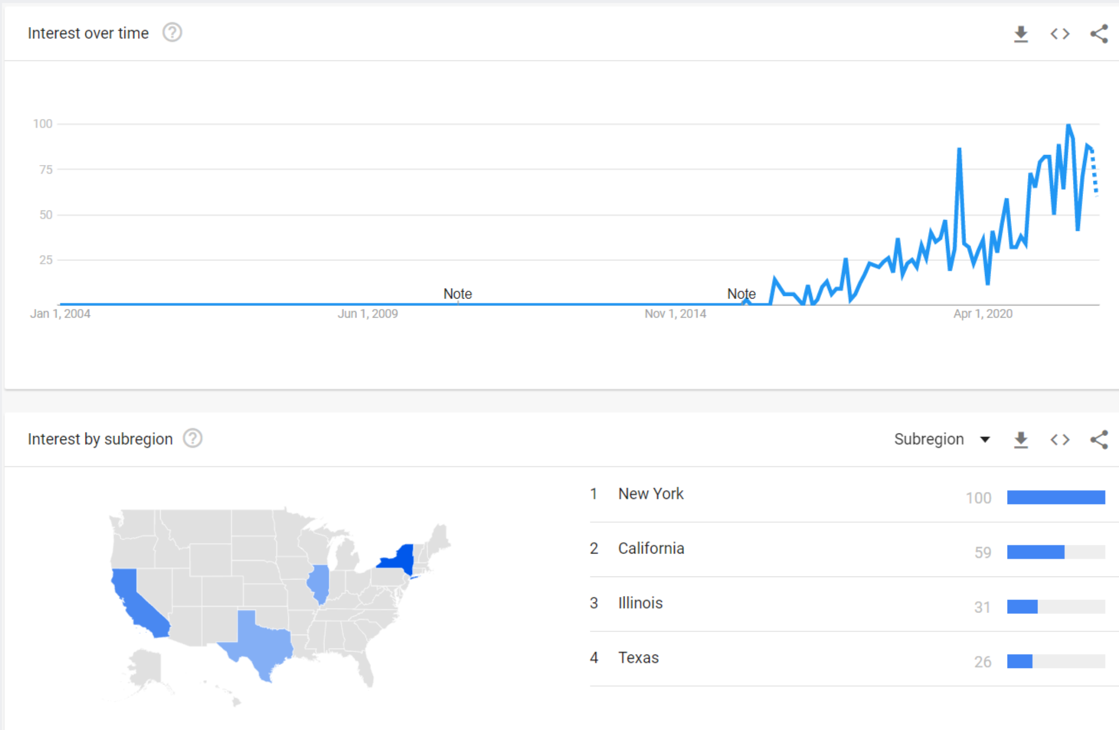
Agenda

- Why does explainability matter?
- What are top drivers?
- What is the current landscape of model explainability?
- X-FIPE: computationally efficient and theoretically sound.

Why explainablity matters

Explainable AI Google Trends

Explainable AI starts to trend in the last 5 years



Interpretability is vital for business stakeholders to adopt ML to serve consumers in compliance with rules

Data scientists & AI researchers
Debug, improve, and compare models
Check for biases

Consumers
Trust the business
Ensure explanations are in its simplest form

Business teams
Trust the model
Adopt the model

Policymakers
Provide guidelines to organizations
Protect rights of consumers

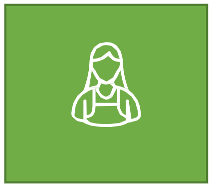
What are top drivers

Utilization driven

Score: 0.99

Number of **ER** visit: HIGH
Number of **ER (nonemergency)** visit: HIGH
ER visit **claim cost**: HIGH
Physician exam/consultation medical claim count: HIGH
ER visit claim in **past 3 months**: YES

Respiratory condition driven



Score: 0.85

Physician exam/consultation medical claim count: HIGH
Chronic pulmonary disease med claim count: HIGH
Patient/provider/equipment transportation claim percent: HIGH
Electrocardiogram medical claim count: HIGH
Respiratory disease (uncategorized) - medical claim count: HIGH

Behavioral condition driven

Score: 0.40

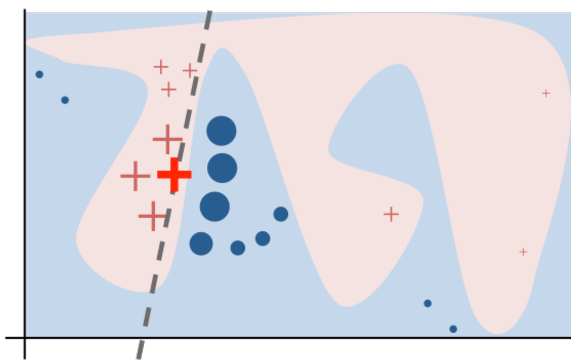
Physician exam/consultation medical claim count: HIGH
Rx claim count: HIGH
Mental condition (uncategorized) medical claim count: HIGH
Total med claim cost: HIGH
Anxiety medical claim count: HIGH

Model explainability algorithms

LIME and SHAP

Local Interpretable Model-Agnostic Explanations

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \ \mathcal{L}(f, g, \pi_x) + \Omega(g)$$



LIME: Ribeiro, et al. "Why should I trust you?: Explaining the predictions of any classifier" 2016.

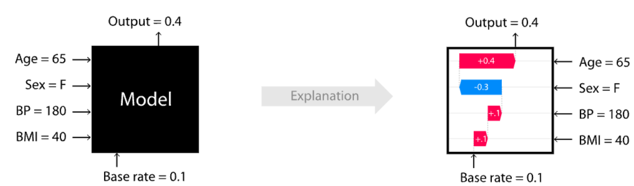
SHapley Additive exPlanations

Definition 1 **Additive feature attribution methods** *have an explanation model that is a linear function of binary variables:*

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i, \tag{1}$$

where $z' \in \{0,1\}^M$, M *is the number of simplified input features, and* $\phi_i \in \mathbb{R}$.

SHAP



SHAP: Lundberg, et al. “A Unified Approach to Interpreting Model Predictions”, 2017.

Comparison of interpretability methods

#	Comparison Feature	LIME	SHAP	FIPE*	X-FIPE
1	Computational efficiency (run time)	No (Medium/High)	No (High)	Yes (Medium)	Yes+ (Low)
2	Avoid heuristics/user settings	No	Yes	Yes	Yes
3	Handle correlated features	Yes	Yes	Yes	Yes
4	Avoid additional model training	No	Yes	Yes	Yes
5	Full features representation	No	Yes	Yes	Yes
6	Python and Spark model agnostic	Yes	No*	Yes	Yes
	Total positive "yes" features	2	4	6	6+

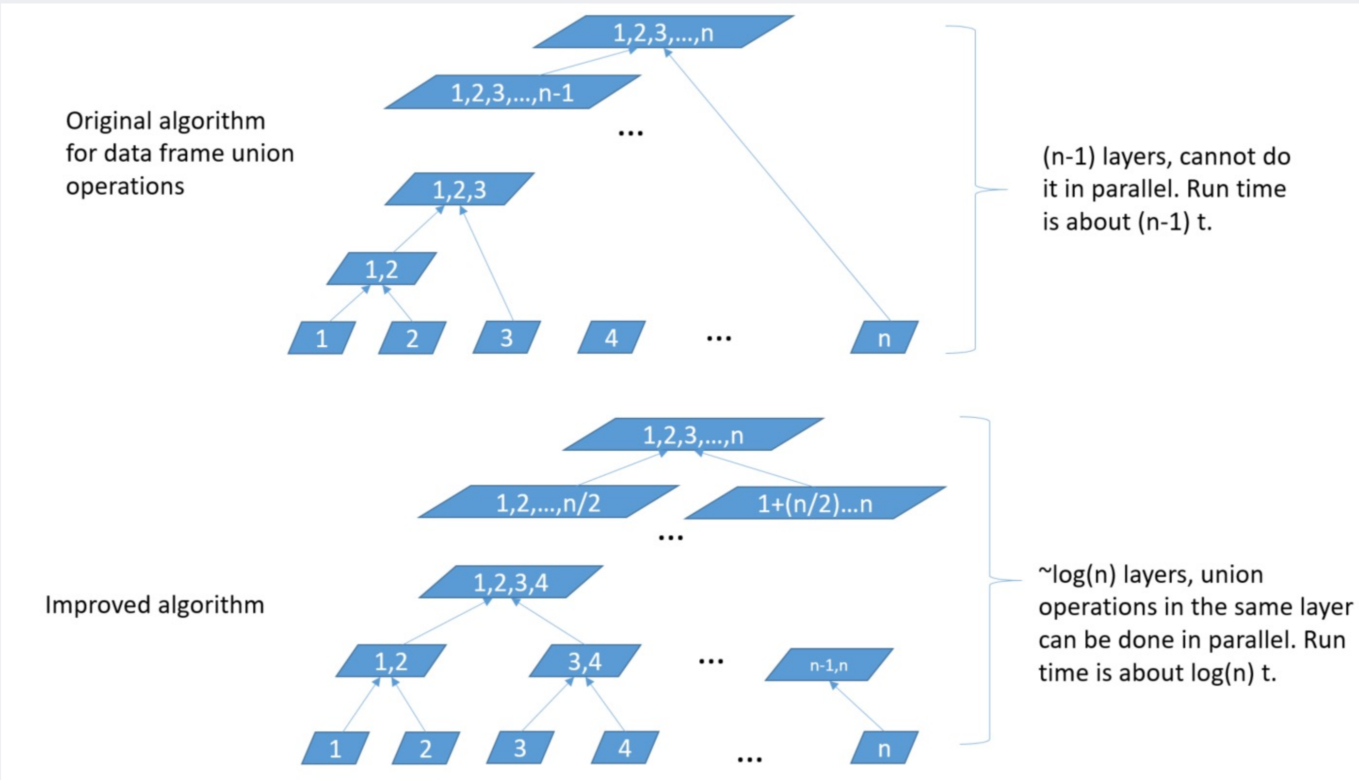
* FIPE: https://www.researchgate.net/publication/335189270_Feature_Impact_for_Prediction_Explanation
* Open Issue: A Spark version in plan? https://github.com/slundberg/shap/issues/38

X-FIPE

What is novel with X-FIPE (1)

- Computationally efficient
  - Time complexity: O(n) -> O(log n)

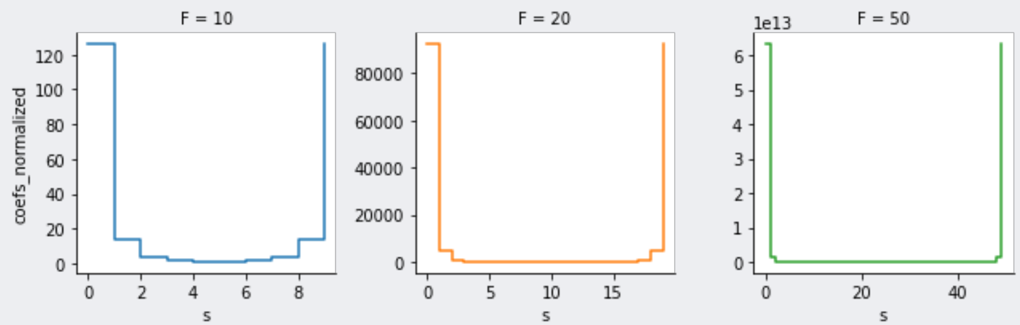
O(log n) algorithm for combining dataframes



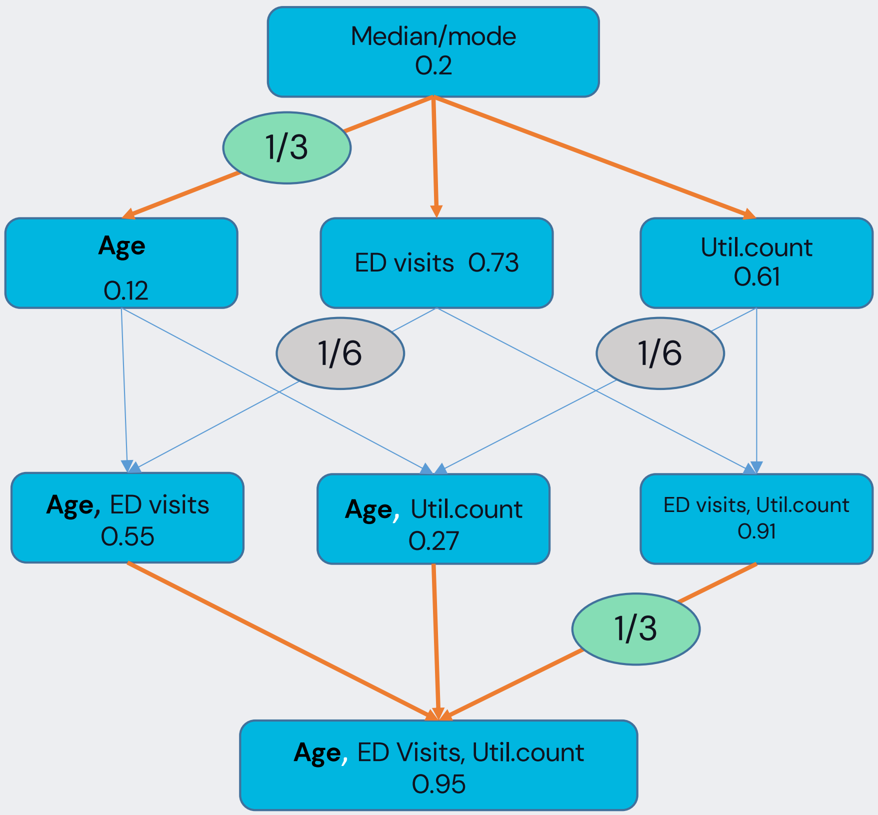
What is novel with X-FIPE (2)

- Theoretically sound
  - First order approximation of Shapley value
  - The coefficients are the largest when |S| = 0 or F-1
  - As F increases, coefficients decrease much faster for the other sizes of subsets

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \left[f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S) \right]. \qquad (4)$$



X-FIPE is a first order approximation of Shapley



X-FIPE take two most important layers:
First Layer and Last Layer

In X-FIPE formula

Not in X-FIPE formula

1/3 Included coefficients for **Age**

1/6 Excluded coefficients for **Age**

X-FIPE walk-through

1. Model Scores and statistics are obtained as the follow:

Sample of three IDs with their inputs values along with their model predictions

ID	Age	ED Visits	Utilization count	Model Prediction (S_o)
1	62	4	7	0.95
2	67	1	2	0.3
3	68	0	3	0.43

The median values for the model three inputs and the resulting prediction

Age	ED Visits	Utilization count	Model Prediction (S_m)
66	1	2	0.2

2. For ID-1, X-FIPE produces the following new observations, which correspond to different settings and lead to new scores.

ID	Age	ED Visits	Utilization count	Model Prediction	Feature Name	Score Tag
1	66	4	7	0.91	Age	S_x
1	62	1	2	0.12	Age	S_c
1	62	1	7	0.27	ED Visits	S_x
1	66	4	2	0.73	ED Visits	S_c
1	62	4	2	0.55	Utilization count	S_x
1	66	1	7	0.61	Utilization count	S_c

3. The impact is calculated for each feature

Shaply value for X-fipe : $S_c - S_m + S_0 - S_x$

$$Y_x(Age) = 0.12 - 0.2 + 0.95 - 0.91 = -0.04$$

$$Y_x(ED_{Visits}) = 0.73 - 0.2 + 0.95 - 0.27 = 1.21$$

$$Y_x(Util) = 0.61 - 0.2 + 0.95 - 0.55 = 0.81$$

4. Rank the impact of each feature.

ID	Feature Name	Impact (Yx)	Rank
1	ED Visits	1.21	1
1	Utilization Count	0.81	2
1	Age	-0.04	3

X-FIPE in code

```python
# pip install xfipe
from xfipe import xfipe # import xfipe
xfipe_obj = xfipe(predict_proba, model_type, fipe_feature_list, score_col_name) # xfipe object
new_obs_nox_median = xfipe_obj.fipe_score_s_c(median_mode_list, test_data, id_col_name) # Sc
new_obs_x_median = xfipe_obj.fipe_score_s_x(median_mode_list, test_data, id_col_name) # Sx
score_original = xfipe_obj.fipe_score_s_o(test_data, id_col_name) # So
median_score = xfipe_obj.fipe_score_s_m(median_mode_list, id_col_name) # Sm
fipe_data_all = xfipe.fipe_calc_feature_impacts(new_obs_nox_median, new_obs_x_median, score_original, median_score, id_col_name) # xfipe values
```

To learn more

- Interpretable Machine Learning, Molnar 2022 https://christophm.github.io/interpretable-ml-book/
- A Survey of Methods for Explaining Black Box Models, GUIDOTTI, 2018 https://dl.acm.org/doi/10.1145/3236009
- Explainable Artificial Intelligence, Arrieta 2019 https://doi.org/10.1016/j.inffus.2019.12.012
- awesome-machine-learning-interpretability https://github.com/jphall663/awesome-machine-learning-interpretability
- Pitfalls of extracting causal insights from ML models, Lundberg 2021 https://tinyurl.com/notcausal

Meet the team

Xingyu Liu (Max)
Lead ML Engineer

- O(log(n)) DataFrame union algorithm
- X-FIPE algorithm design and implementation

Xingde Jiang (Jason)
Senior ML Engineer

- X-FIPE algorithm design and implementation
- Business evaluation

Steve Brunner
Director of AI

- Business evaluation
- O(log(n)) DataFrame union algorithm

Humana®

Appendix

X–FIPE: Feature Impact Calculation

Shaply value for X-fipe : $S_c - S_m + S_0 - S_x$

$$S_m = f(x_1^m, x_2^m, x_3^m, \ldots)$$

$$S_c = f(x_1, x_2^m, x_3^m, \ldots)$$

$$S_x = f(x_1^m, x_2, x_3, \ldots)$$

$$S_0 = f(x_1, x_2, x_3, \ldots)$$

<u>Where:</u>

S_o: Predicted score without any change for all features from their original value

S_x: Predicted score when only feature x is set to its median/mode value while all other input features are kept at their original values

S_m: Predicted score when all features are set their median/mode value

S_c: Predicted score when all input features except x are set to their median/mode values. Feature x is kept at its original value

Compare with SHAP

sdr_person	fipe_top_5	shap_top_5	x-fipe_top_5
1005369424	[{"fea":"betos_eval_pmpm_ct","sv":"-1.0"}, {"fea":"cnt_cp_webstatement_0","sv":"-0.32"}, {"fea":"total_bh_mbr_resp_pmpm_cost","sv":"0.325"}, {"fea":"rwjf_uninsured_pct","sv":"-0.22"}, {"fea":"cnt_cp_print_0","sv":"0.192"}]	[{"fea":"betos_eval_pmpm_ct","sv":"-0.602"}, {"fea":"total_bh_mbr_resp_pmpm_cost","sv":"0.3455"}, {"fea":"cnt_cp_print_0","sv":"0.1206"}, {"fea":"cmsd1_men_ind","sv":"0.0953"}, {"fea":"rwjf_uninsured_pct","sv":"-0.094"}]	[{"fea":"betos_eval_pmpm_ct","sv":"-1.0"}, {"fea":"total_bh_mbr_resp_pmpm_cost","sv":"0.325"}, {"fea":"rwjf_uninsured_pct","sv":"-0.22"}, {"fea":"cnt_cp_print_0","sv":"0.192"}, {"fea":"cmsd1_men_ind","sv":"0.187"}]

number of common feature	fipe & shap	x_fipe & shap
5	12752	16099
4	121287	132261
3	407644	417489
2	431752	421911
1	195501	185386
0	32446	28236