Linear

trucks

worker

$$\hat{y} = m_1 x_1 + m_2 x_2 + m_3 x_3 \; - - - - - \; m_n x_n$$

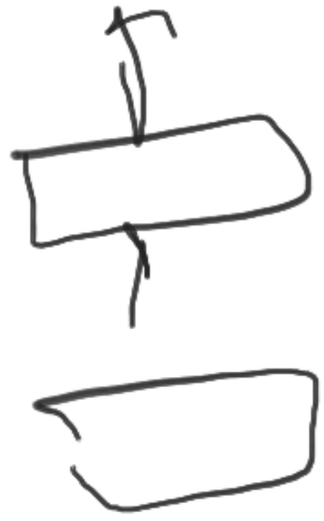
IVS → 400 crores

SIP

1 lakh

why?

important variable

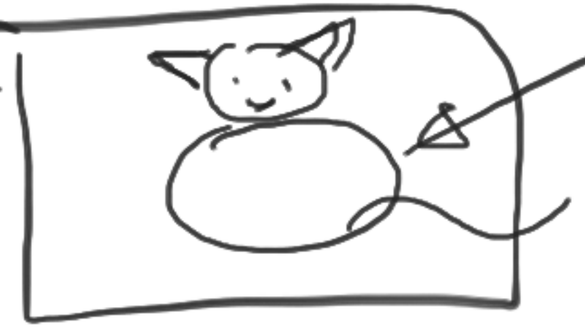


SVM
NN

explainability

Interpret

complex



cat

~~explain~~

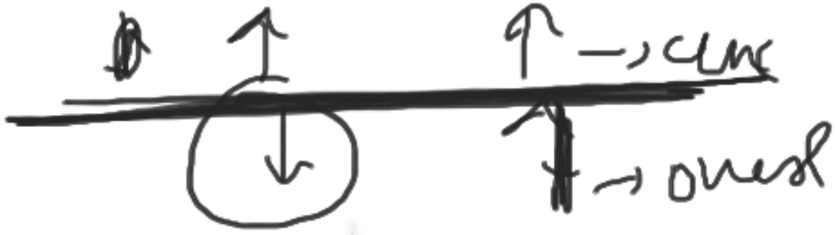
~~dog~~

Niece

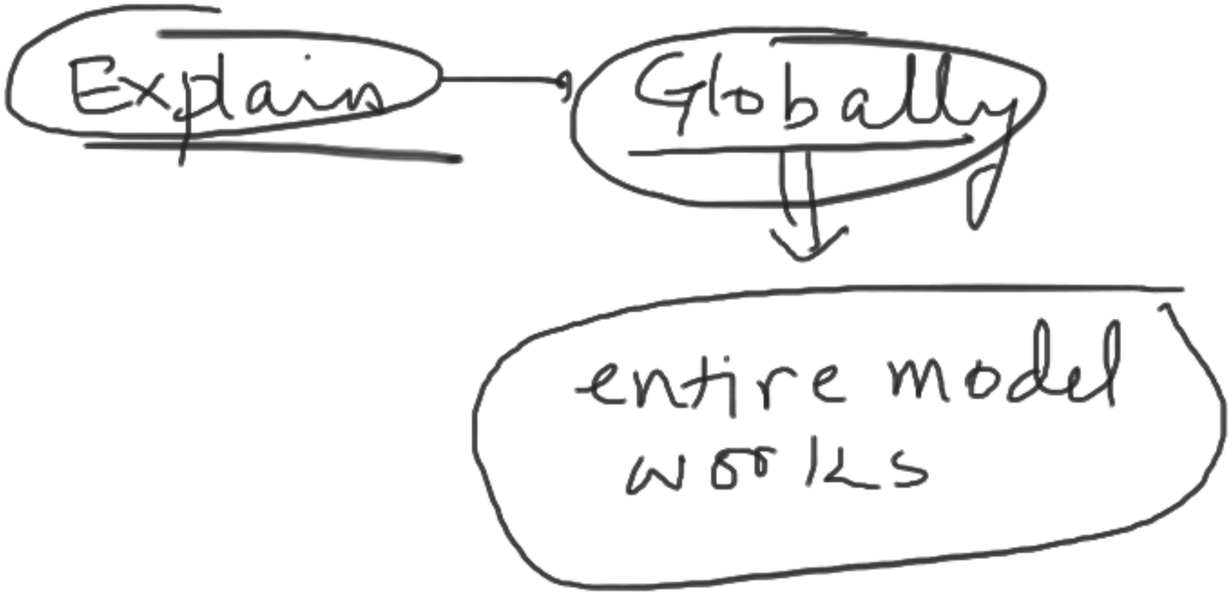
4yrs

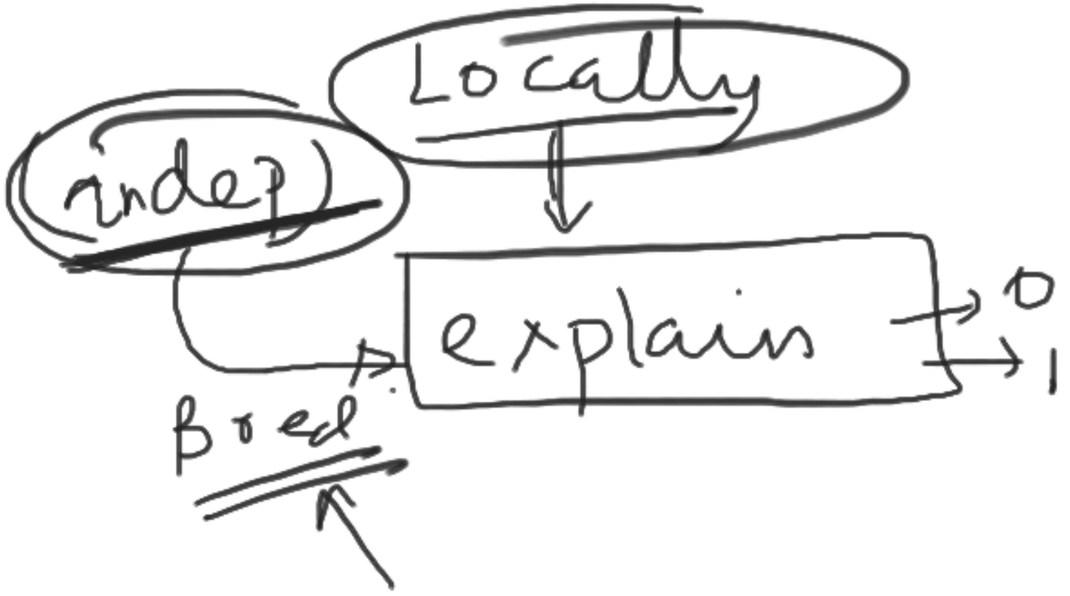
V-shape

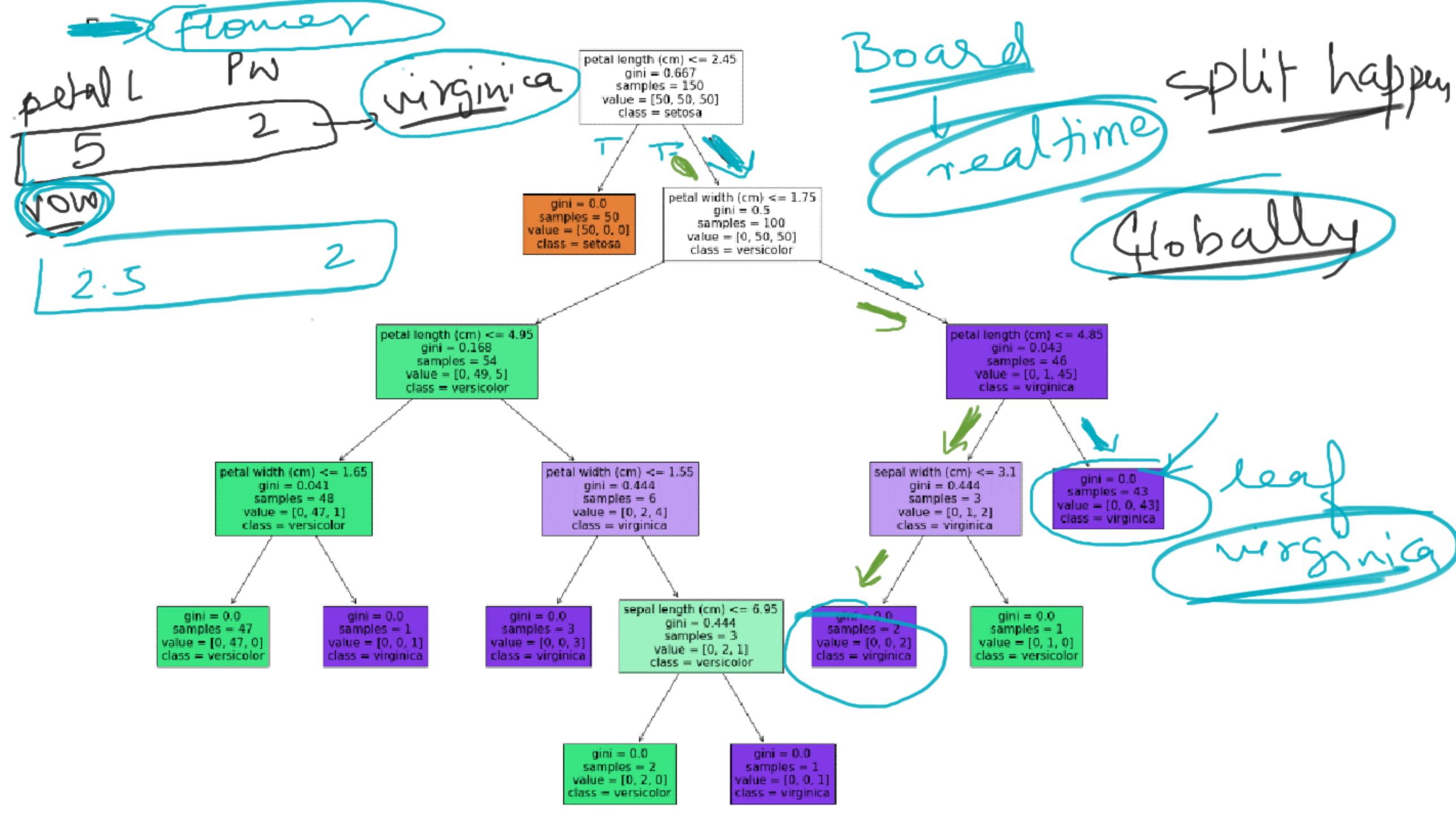
Let us recap on what we did



Data pre-process, clean, Feature Engineer	Model Development &Train	Model Tune	Model Evaluate	Experiment	Brainstorm to improve
Data pre-process • Clubbing data from different data source • Transforms are reproducible, deterministic and idempotent • Create and monitor data pipeline for consistency across data sets **Data Cleaning** • Duplicates removals • Data validity check • Missing values imputation • Outliers handling **Feature Engineering** • Feature Creation & Transformations • Feature Extraction • Feature Selection	**EDA** • Uni-variate analysis • Bi-variate/Multi-variate analysis • Pivots • Visualization and Data Insights **ML model assumption checks** **Data Preparation for Modelling** • Creating dummy variables • Over and Under Sampling (if data is imbalanced) • Split data into train & test **Model building from list of models available**	**Model tuning can be done by:** • Cover the Basics of what parameters are involved in the algorithm that is used. • Find Your Score Metric • Obtain Accurate Forecasting Score by Cross Validation • [redacted] • K-fold cross validation • Diagnose Best Parameter Value Using Validation Curves specially for taking care of overfitting and underfitting	**Evaluate models by holistically looking at** • **Test Harness**: The train test split • **Performance Measure** • Specific metrics for regression and classification • **Cross-validation**: Solving the case of underfitting and overfitting • **Testing Algorithms**: Business test logics that needs to looked into	**Experiment before launch** • **AB Testing:** • [redacted] • Bayesian • [redacted]	**Experiment before launch** • On the basis of causation that led to the failure of Experiment, corresponding methods are employed and the process restarts from model development and train.


Explain
Globally
entire model
works


Locally
indep
explain
0
1
Bred


petal length (cm) <= 2.45
gini = 0.667
samples = 150
value = [50, 50, 50]
class = setosa
gini = 0.0
samples = 50
value = [50, 0, 0]
class = setosa
petal width (cm) <= 1.75
gini = 0.5
samples = 100
value = [0, 50, 50]
class = versicolor
petal length (cm) <= 4.95
gini = 0.168
samples = 54
value = [0, 49, 5]
class = versicolor
petal length (cm) <= 4.85
gini = 0.043
samples = 46
value = [0, 1, 45]
class = virginica
petal width (cm) <= 1.65
gini = 0.041
samples = 48
value = [0, 47, 1]
class = versicolor
petal width (cm) <= 1.55
gini = 0.444
samples = 6
value = [0, 2, 4]
class = virginica
sepal width (cm) <= 3.1
gini = 0.444
samples = 3
value = [0, 1, 2]
class = virginica
gini = 0.0
samples = 43
value = [0, 0, 43]
class = virginica
gini = 0.0
samples = 47
value = [0, 47, 0]
class = versicolor
gini = 0.0
samples = 1
value = [0, 0, 1]
class = virginica
gini = 0.0
samples = 3
value = [0, 0, 3]
class = virginica
sepal length (cm) <= 6.95
gini = 0.444
samples = 3
value = [0, 2, 1]
class = versicolor
gini = 0.0
samples = 2
value = [0, 0, 2]
class = virginica
gini = 0.0
samples = 1
value = [0, 1, 0]
class = versicolor
gini = 0.0
samples = 2
value = [0, 2, 0]
class = versicolor
gini = 0.0
samples = 1
value = [0, 0, 1]
class = virginica

Ind
one single
LM
LM
LM
NLM
error
cost
QP

- LIME(XGB)
- LIME(LGB)
- LIME(RF)

same 6 step

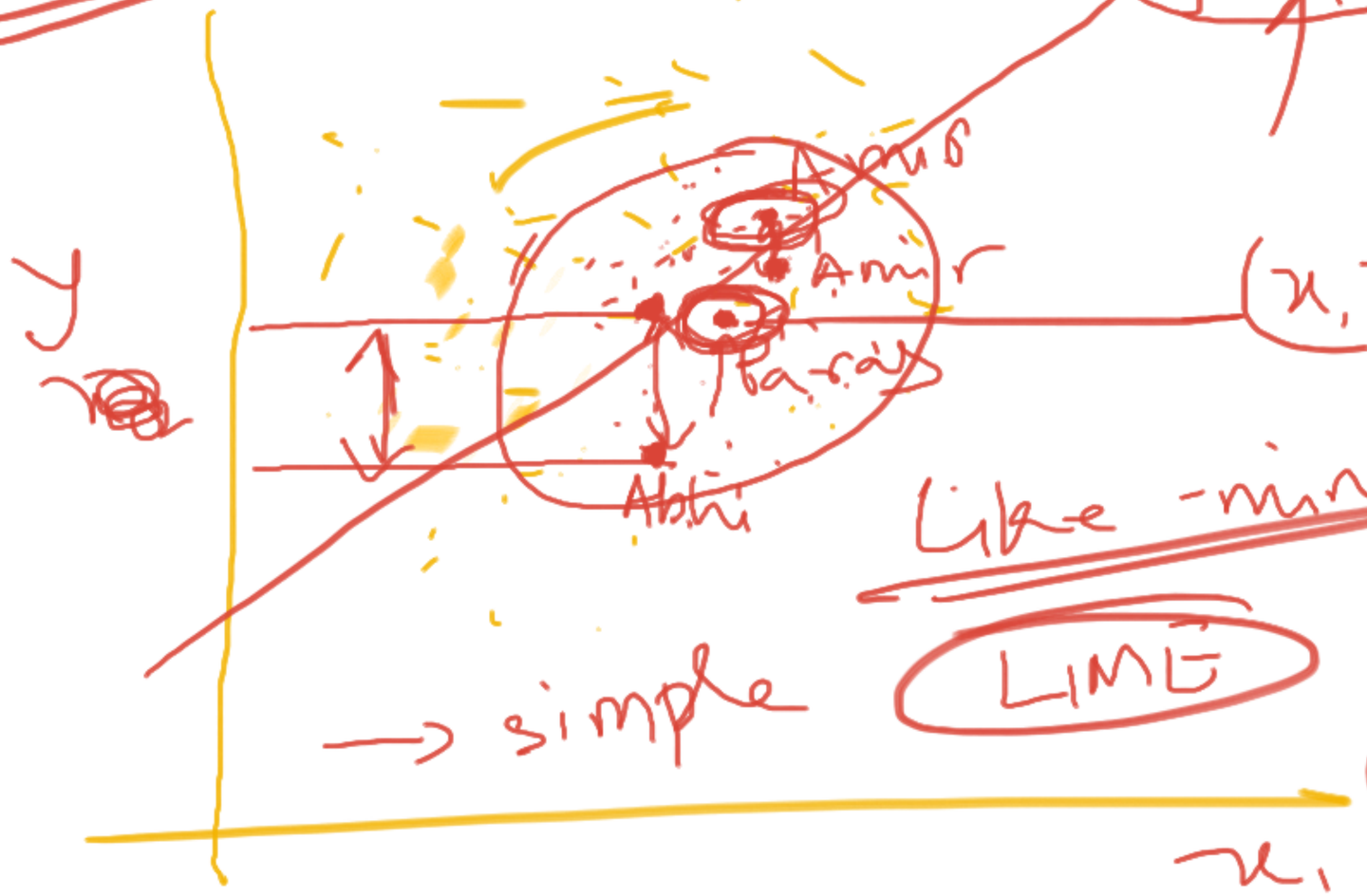
explain

LIME

XGB

$f(x) \rightarrow$ distribution

y : w.a.



$(x_1, x_2) \rightarrow$ class $\begin{cases} 0 \\ 1 \end{cases}$

Like-mined

x_1, x_2

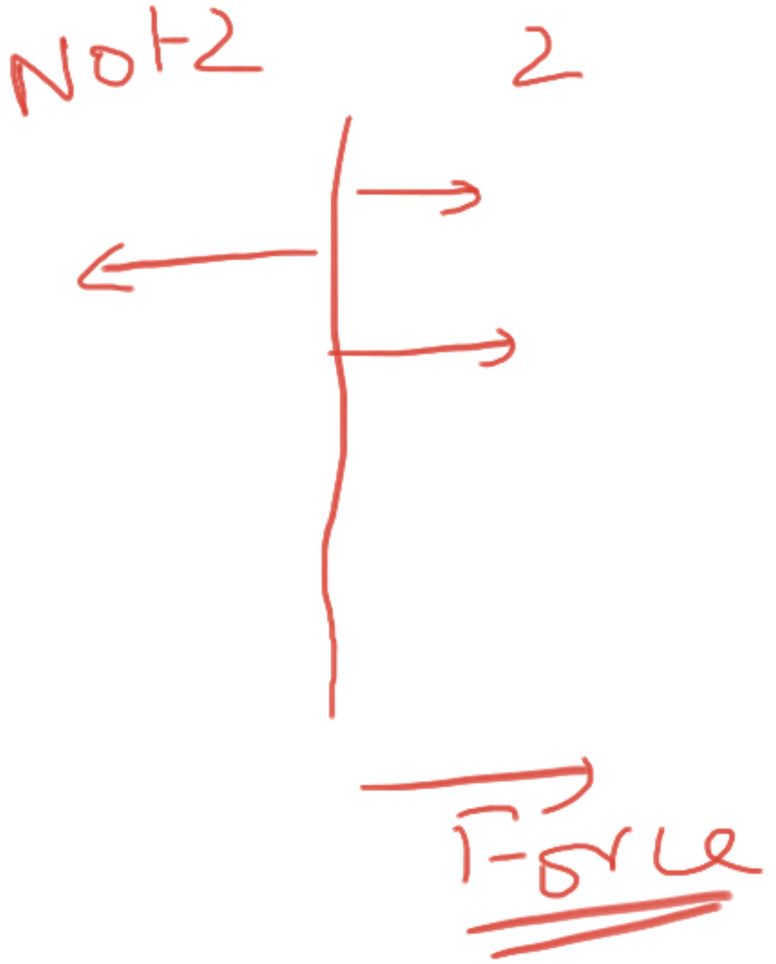
$\rightarrow$ simple

LIME

Paras $\rightarrow$ Amir

"m" $\rightarrow$

$$\boxed{y = m_1 x + m_2 x_2}$$


Not 2
2
Force

Game theory → oversimplified

Cab → 100/-

$d_1 : d_2 : d_3$

Me

x_1, d_1

Suj 1-

x_2, d_2

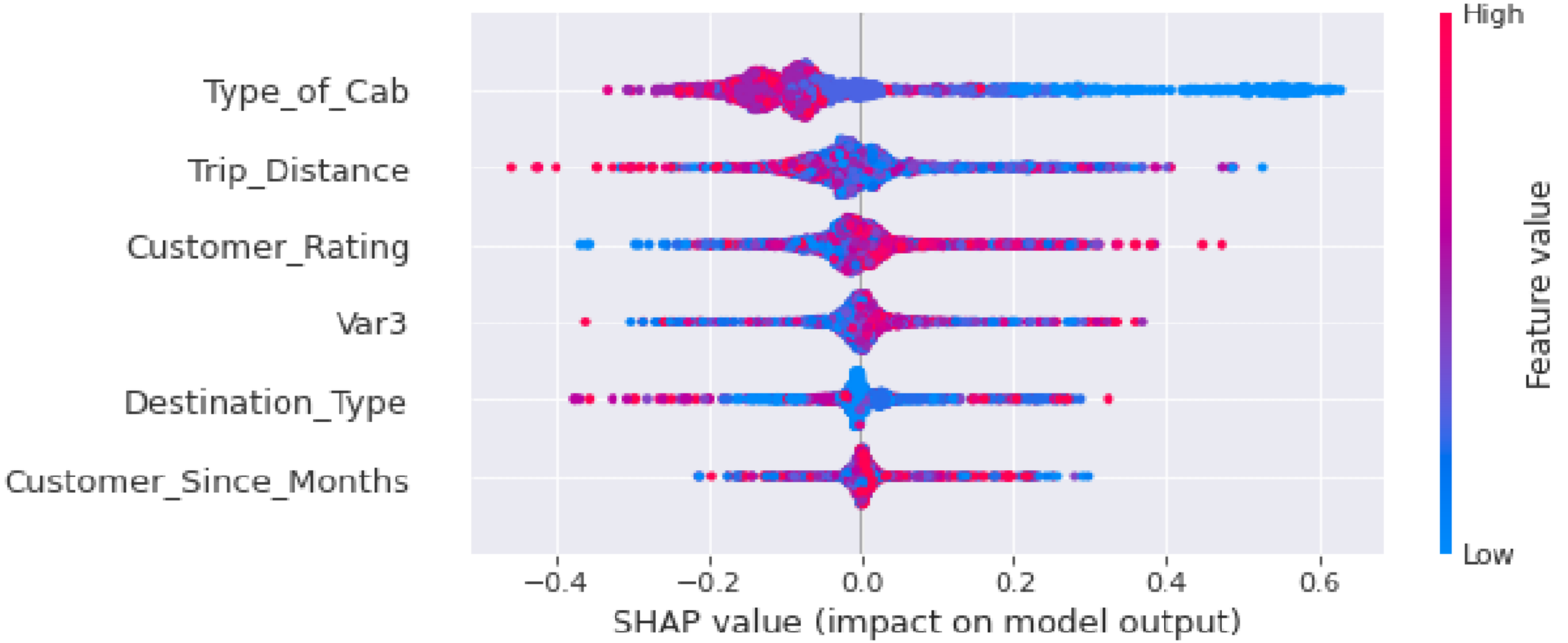
Raja

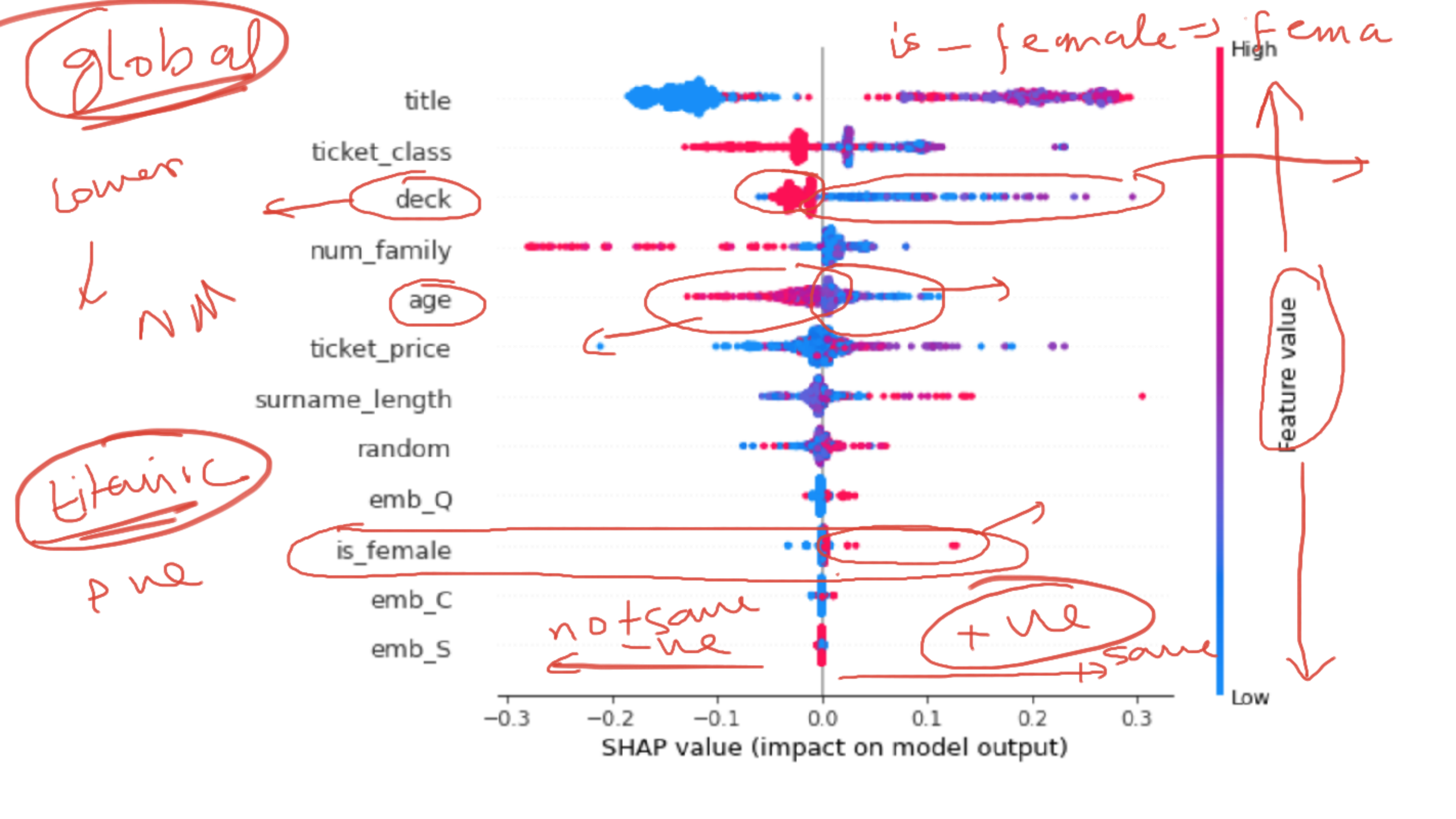
x_3, d_3

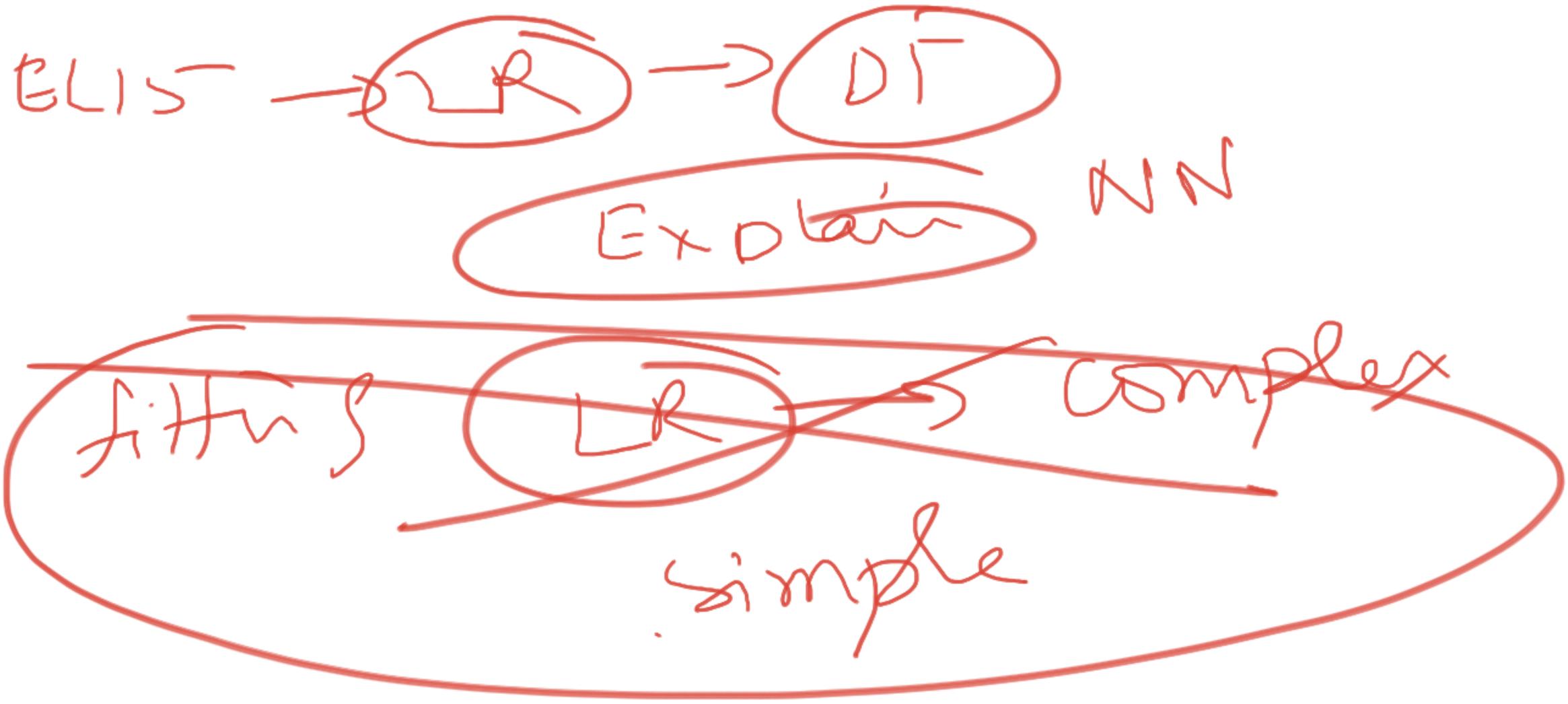
$$f(x_1, d_1) : f(x_2, d_2) : f(x_3, d_3)$$

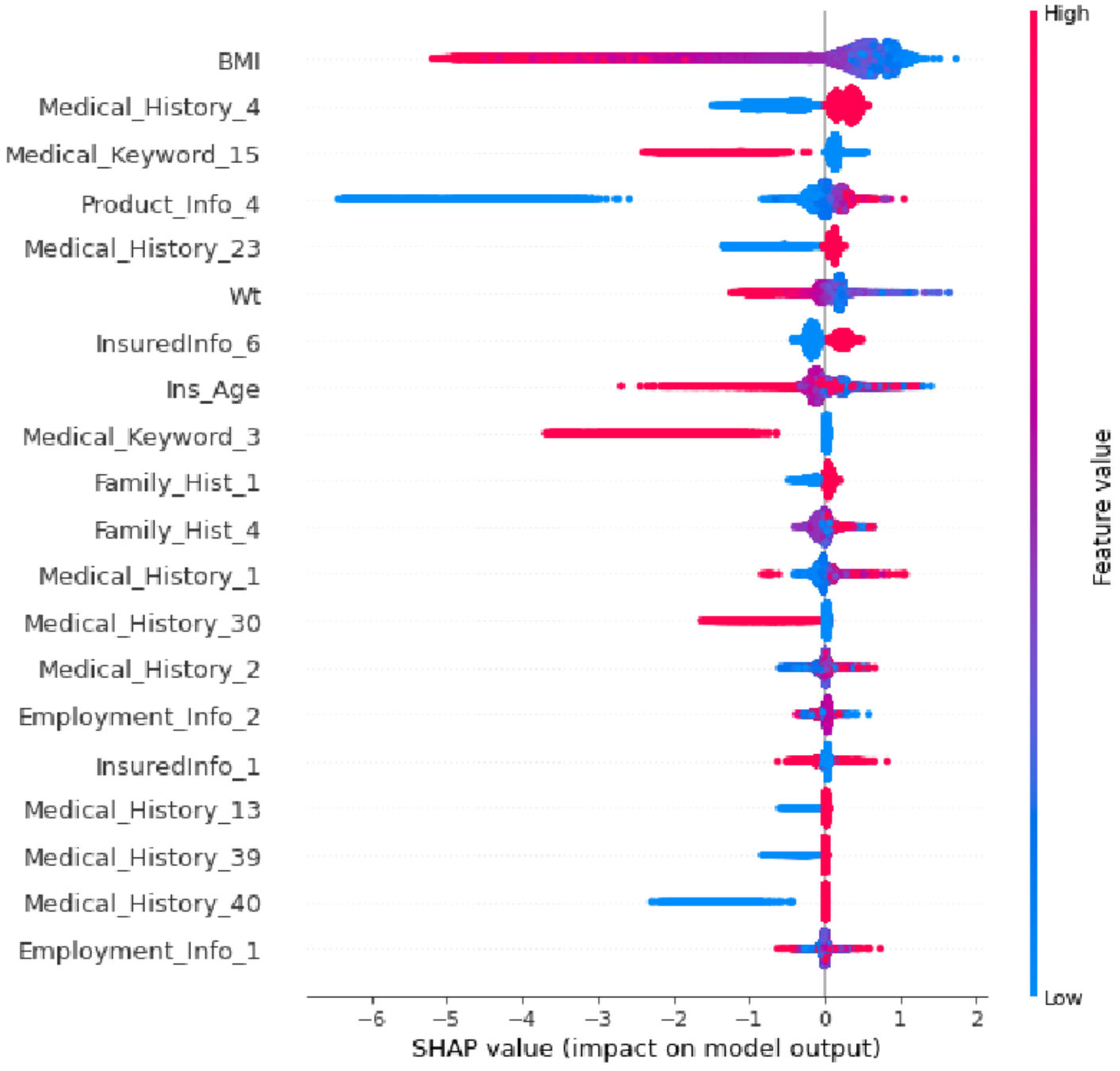
S1 + AP

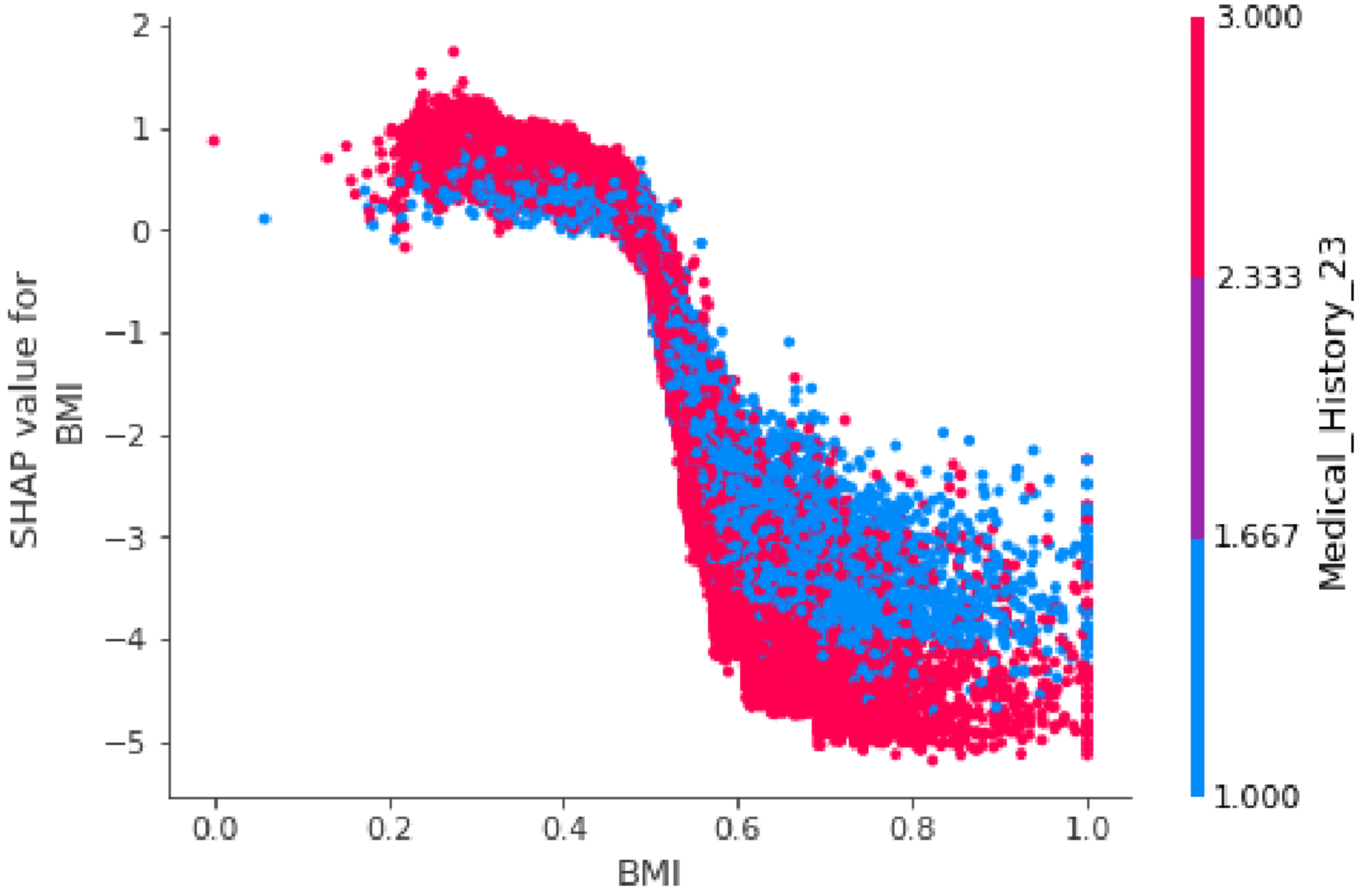
local interp
(SHAP)
higher ⇄ lower
base value
model output
14.34
16.34
18.34
20.34
22.34
24.41
26.34
28.34
30.34
PTRATIO = 15.3
LSTAT = 4.98
RM = 6.575
NOX = 0.538
AGE = 65.2
RAD = 1

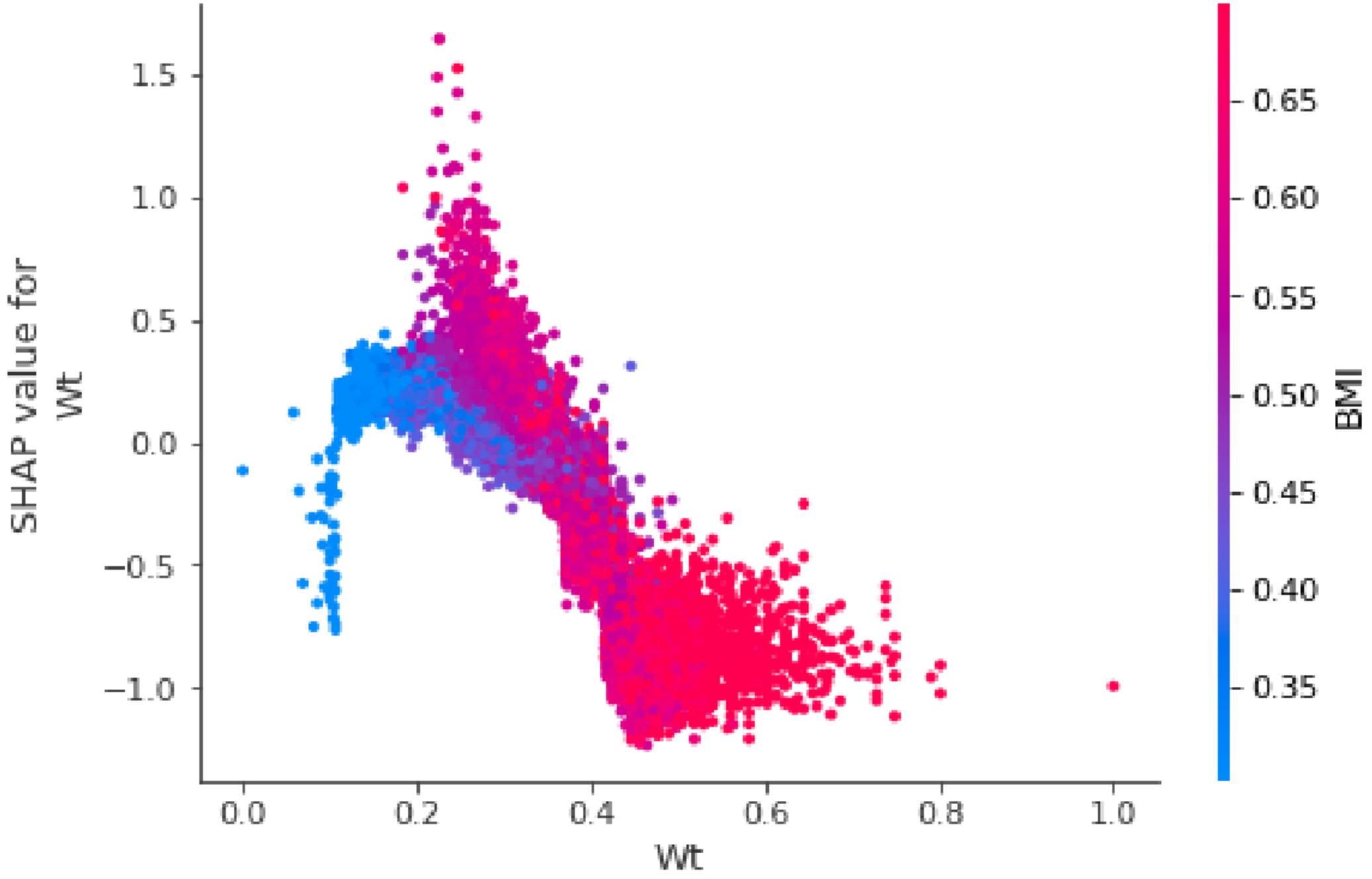

Type_of_Cab
Trip_Distance
Customer_Rating
Var3
Destination_Type
Customer_Since_Months
−0.4
−0.2
0.0
0.2
0.4
0.6
SHAP value (impact on model output)
High
Feature value
Low


global
is_female → fema
title
ticket_class
deck
num_family
age
ticket_price
surname_length
random
emb_Q
is_female
emb_C
emb_S
High
Low
Feature value
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
SHAP value (impact on model output)


ELIS
LR
DT
Explain
NN
fitting
LR
complex
simple


BMI
Medical_History_4
Medical_Keyword_15
Product_Info_4
Medical_History_23
Wt
InsuredInfo_6
Ins_Age
Medical_Keyword_3
Family_Hist_1
Family_Hist_4
Medical_History_1
Medical_History_30
Medical_History_2
Employment_Info_2
InsuredInfo_1
Medical_History_13
Medical_History_39
Medical_History_40
Employment_Info_1
−6
−5
−4
−3
−2
−1
0
1
2
SHAP value (impact on model output)
High
Low
Feature value


SHAP value for BMI
2
1
0
−1
−2
−3
−4
−5
0.0
0.2
0.4
0.6
0.8
1.0
BMI
3.000
2.333
1.667
1.000
Medical_History_23


SHAP value for Wt
1.5
1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0
Wt
BMI
0.65
0.60
0.55
0.50
0.45
0.40
0.35