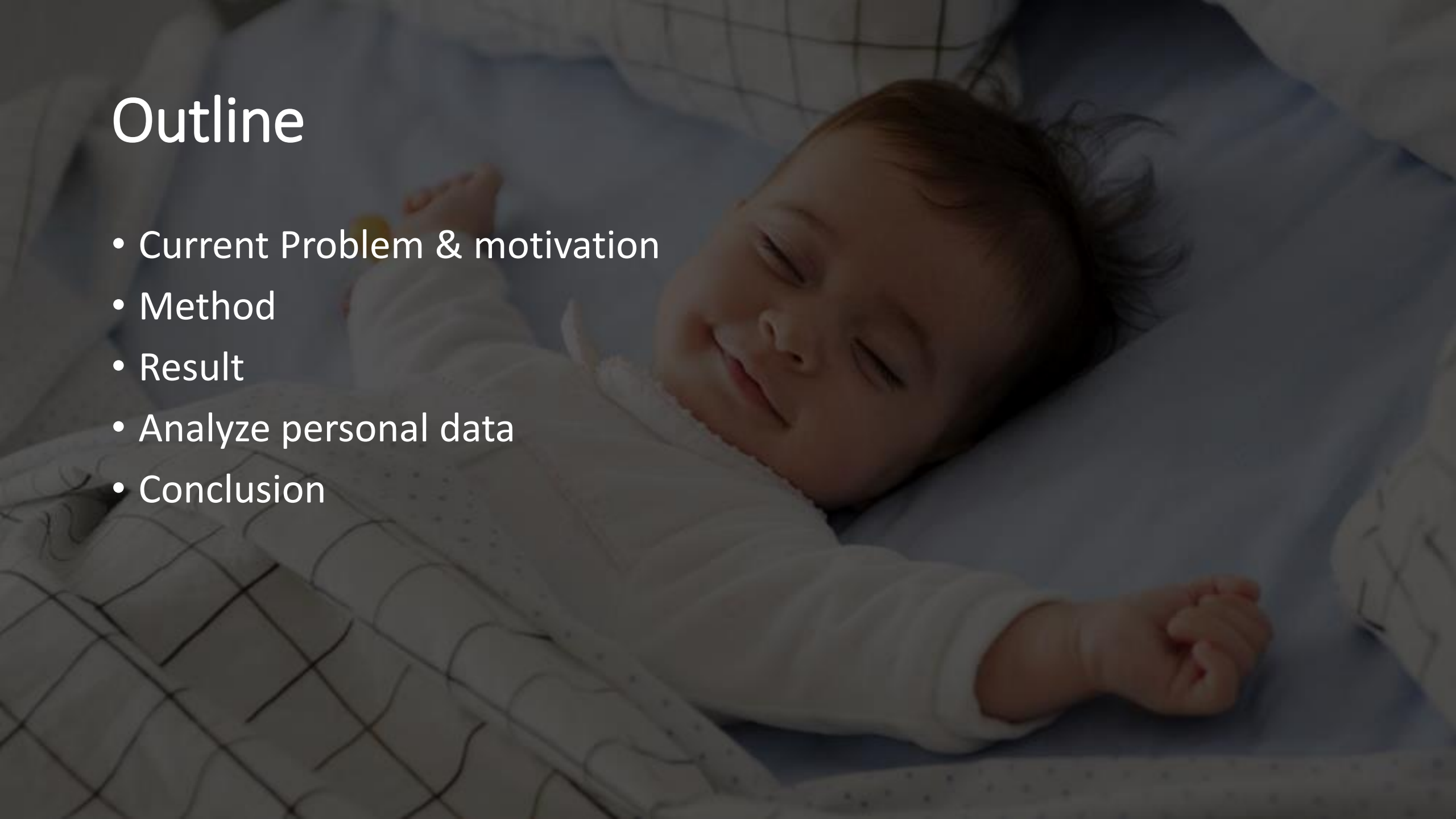
B04901068 劉力仁
B04901069 林志皓
B04901104 吳添聚

Did you sleep well?



Outline

- Current Problem & motivation
- Method
- Result
- Analyze personal data
- Conclusion

Current Problem & Motivation

- Sleep disorder is a common disease among modern people.
- The factor is complicated
  - Psychiatric disorders
  - Medical issues
  - Physical disturbances
- Complication
  - Depression
  - inflammation

- Find a good way to know the reason behind sleep disorder!

Dataset : NHANES

- A program designed to assess the health and nutritional status of adults and children in the United States.
- It combines interviews and physical examinations, uses a complex, multistage, probability sampling design.
- Findings from this survey will be used to determine the prevalence of major diseases and risk factors for diseases.

Experiment Design

- Train model for each group of data of certain category, and evaluate the performance (e.g. "Mental Health")
- Select categories with good performance, and train model to determine which feature is important (to be discuss later)
- Analyze the impact of single feature
- Analyze individual cases
- Compare the results with reference data

Data preprocess

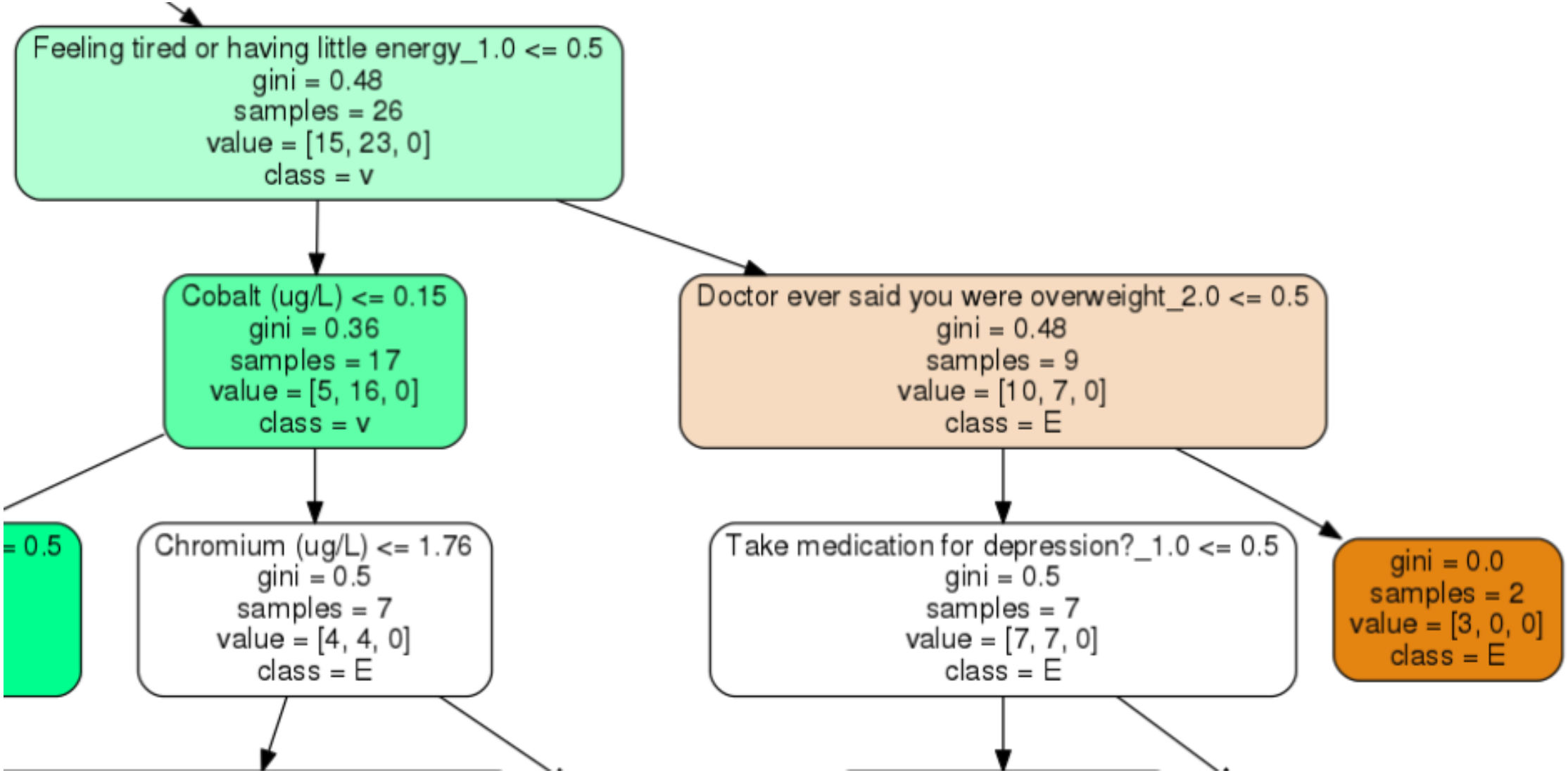
Problem	Process
Too much missing data	Drop feature/individule that has more than 50% are missing
Categorical feature (e.g. "yes" / "no")	One-Hot encoding (e.g. 01 / 10)
Missing/Unknown data	Replace with median value
Multiclass in target data	Split into 0/1 (1 = has sleep disorder)

Model : Decision Tree

- Ask questions and make diagnosis, just like doctors
- Good performance
- Easy to use
- Explainable to humans
- Convenient for analysis

- Many Trees --> **"Random Forest"**

A small part of a Decision Tree

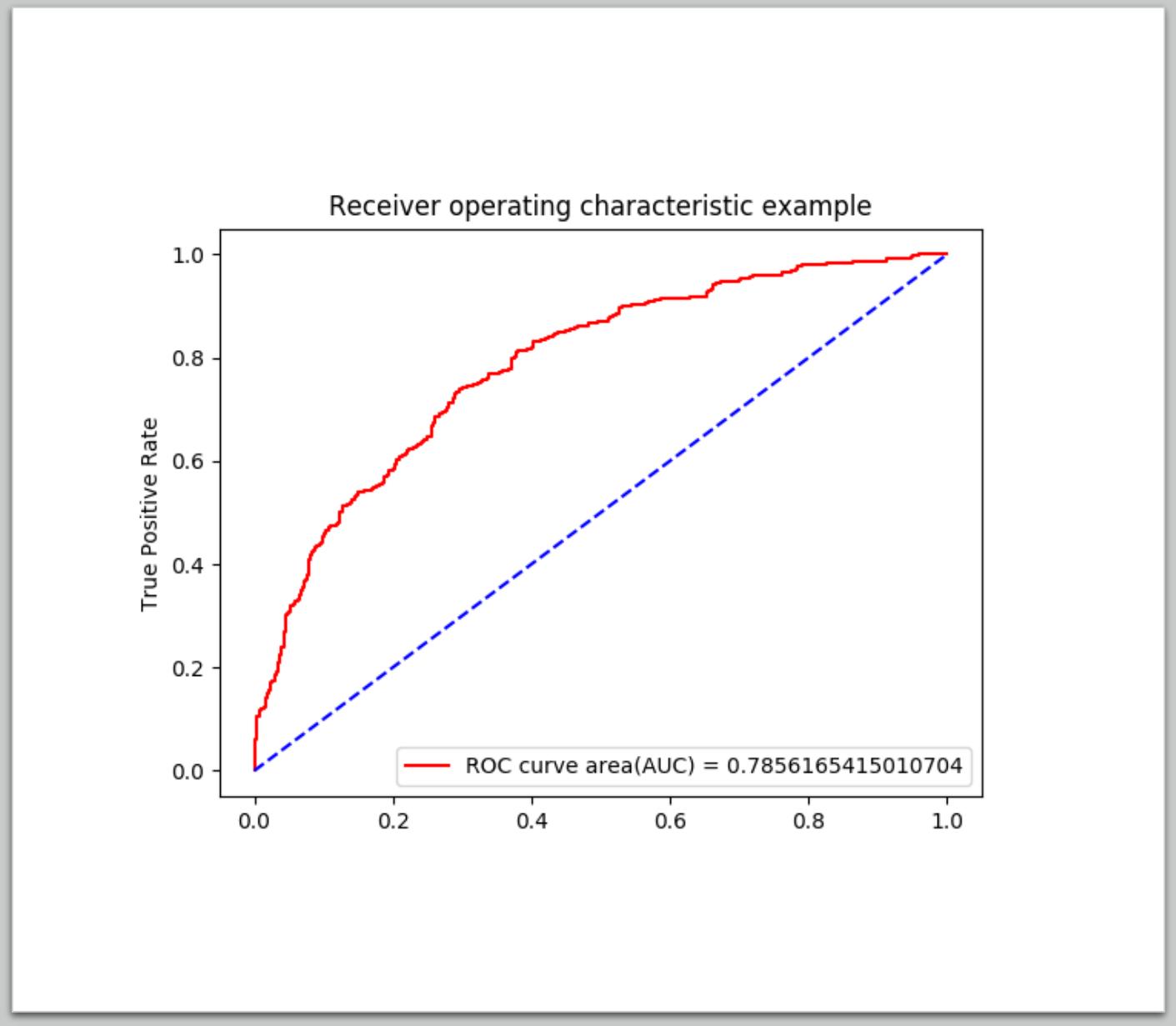


Evaluation

- Accuracy : 0.79

(target: "*Ever told doctor had trouble sleeping?* ")

- ROC curve
- AUC (Area Under Curve): 0.78



Model : **XGBoost**

- Advantage:
  - Execution Speed
  - Sparse Aware
  - Regularization

- 5-fold cross-validation
- Feature preprocessing, selecting feature
- AUC results: 0.811 (std: 0.002)

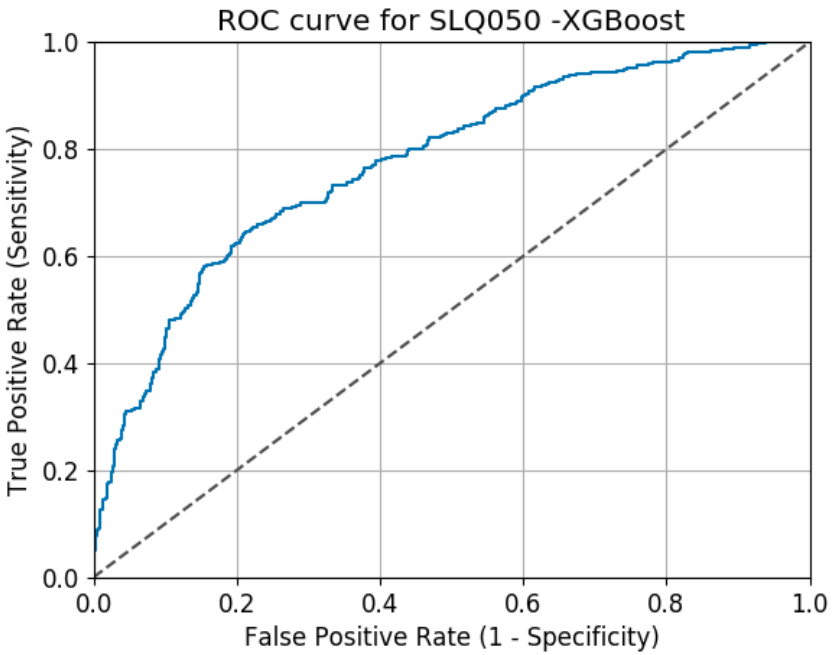
- Model: assuming we have K trees

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

- Objective

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Training loss — Complexity of the Trees



Experiment : important features

Trouble sleeping or sleeping too much v.s. Disability

Feature	Score - | ACC: 0.665 | AUC: 0.684
How often do you feel depressed?_5.0	0.10954090260671583
How often do you feel worried, anxious?_1.0	0.0588800138733321
How often do you feel worried, anxious?_5.0	0.04643514354830654
Have serious difficulty concentrating?_1.0	0.04564936758500549
How worried or anxious were you?_2.0	0.044185356559170275
How worried or anxious were you?_3.0	0.04248165708140591
Take medication for depression?_1.0	0.04124358435719114
Have serious difficulty concentrating?_2.0	0.03782379446127237
How often do you feel depressed?_4.0	0.03675607354008972
How worried or anxious were you?_1.0	0.03508081731306644

Trouble sleeping or sleeping too much v.s. Mental Health

Feature	Score - | ACC: 0.769 | AUC: 0.857
Difficulty these problems have caused_0.0	0.20698718269160268
Feeling tired or having little energy_0.0	0.15206240424680934
Poor appetite or overeating_0.0	0.07171344328207549
Feeling down, depressed, or hopeless_0.0	0.0632427358931295
Have little interest in doing things_0.0	0.05927120394872957
Difficulty these problems have caused_1.0	0.04394641364642713
Trouble concentrating on things_0.0	0.042760810420631006
Feeling bad about yourself_0.0	0.041545507047908874
Moving or speaking slowly or too fast_0.0	0.03431298755813071
Feeling tired or having little energy_1.0	0.03229898205893977

Ever told doctor had trouble sleeping?
v.s. Mental Health

Feature	Score - | ACC: 0.743 | AUC: 0.691
Feeling tired or having little energy_0.0	0.09039885180201236
Feeling down, depressed, or hopeless_0.0	0.06795848478846003
Difficulty these problems have caused_0.0	0.06290747078455021
Difficulty these problems have caused_1.0	0.055187663237221005
Poor appetite or overeating_0.0	0.05443016326253826
Trouble concentrating on things_0.0	0.04894055180713151
Moving or speaking slowly or too fast_0.0	0.04559488919852508
Have little interest in doing things_0.0	0.04431287129248116
Feeling bad about yourself_0.0	0.03381097243862631
Feeling tired or having little energy_1.0	0.032291412686477165

Ever told doctor had trouble sleeping?
v.s. Medical Conditions

Feature	Score - | ACC: 0.748 | AUC: 0.734
Doctor ever said you had arthritis_1.0	0.0858841391042585
Doctor ever said you had arthritis_2.0	0.07963682408108218
Doctor told you to exercise_2.0	0.034283899980709014
Doctor told you to exercise_1.0	0.03165026690769339
Any metal objects inside your body?_1.0	0.026355439423879592
Doctor told you to lose weight_2.0	0.02225883296313052
Doctor ever said you were overweight_2.0	0.022058191298710512
Doctor ever said you were overweight_1.0	0.021346834209423102
Ever told you had COPD?_2.0	0.020482862585220703
Close relative had heart attack?_2.0	0.019133417631625443

Ever told doctor had trouble sleeping?
v.s. Blood pressure & Cholesterol

Feature	Score - | ACC: 0.740 | AUC: 0.646
Ever told you had high blood pressure_1.0	0.20427240186370973
Ever told you had high blood pressure_2.0	0.19278955550834379
When blood cholesterol last checked_1.0	0.109539704699632
Doctor told you - high cholesterol level_2.0	0.09804259903072303
Told to take prescriptn for cholesterol_1.0	0.08932713210360729
Ever had blood cholesterol checked_2.0	0.08135306654906115
Doctor told you - high cholesterol level_1.0	0.05397184263532155
Told to take prescriptn for cholesterol_2.0	0.04988581655285744
Ever had blood cholesterol checked_1.0	0.036263491579037285
When blood cholesterol last checked_2.0	0.03224163176523854

Ever told doctor had trouble sleeping?
v.s. Income

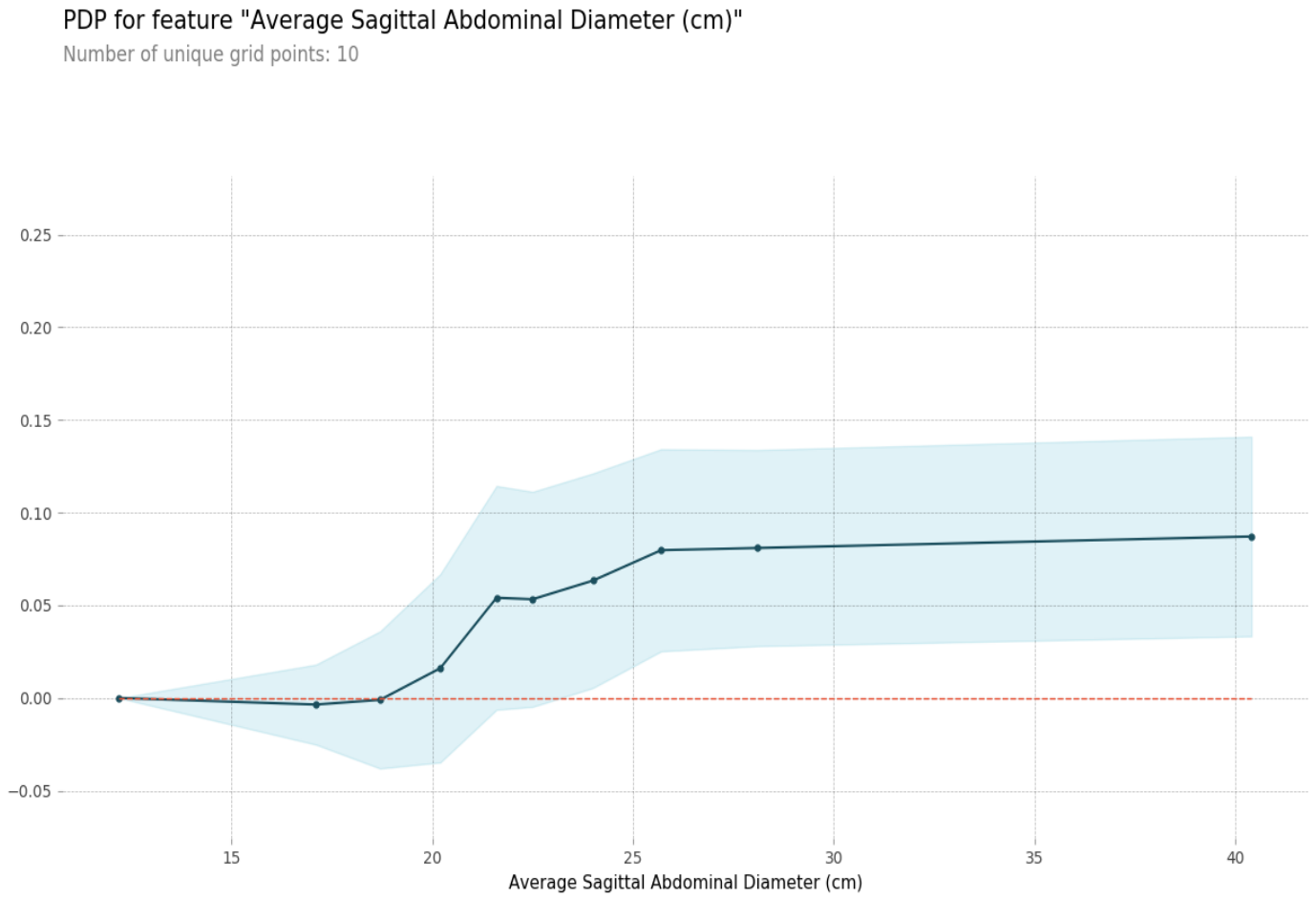
Feature	Score - | ACC: 0.739 | AUC: 0.636
Monthly family income	0.11679555613801275
Income from Social Security or RR_1.0	0.06503927716012675
How do you get to the grocery store?	0.06300139656400318
Income from Social Security or RR_2.0	0.05509341018769272
Income from other disability pension_1.0	0.04136942243001331
Income from other disability pension_2.0	0.04057791971357154
Income from wages/salaries_1.0	0.03434797897979966
Income from wages/salaries_2.0	0.033983271822012974
Income from Supplemental Security Income_1.0	0.028933966526148304
Income from Supplemental Security Income_2.0	0.027718687266971

How often do you snore?
v.s. All features

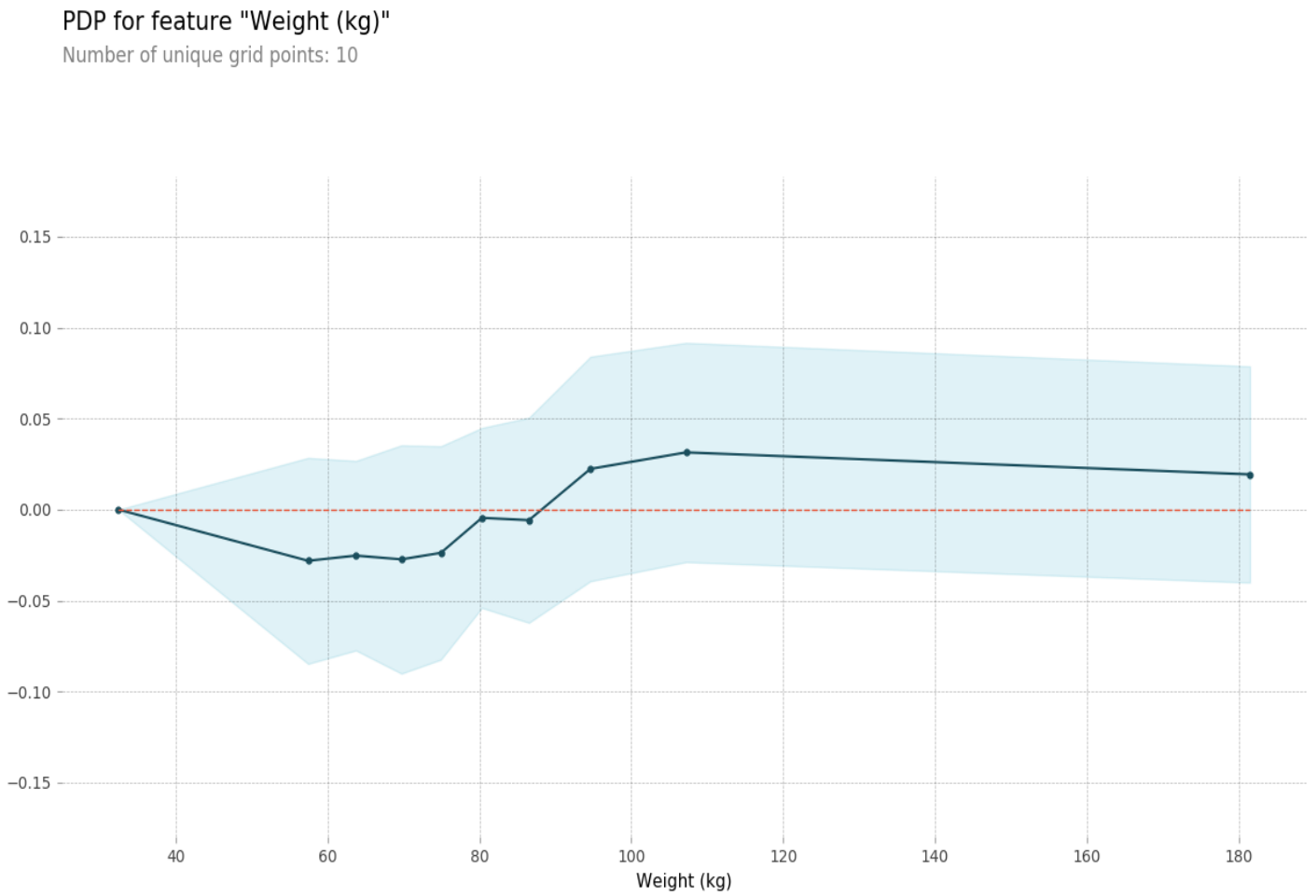
Feature	Score - | ACC: 0.660 | AUC: 0.715
Average Sagittal Abdominal Diameter (cm)	0.017109416872987 55
Weight (kg)	0.015077170192779847
Arm Circumference (cm)	0.014205861460742248
Self-reported weight - 1 yr ago (pounds)	0.014187743619511785
Sagittal Abdominal Diameter 2nd (cm)	0.013451427105262719
Body Mass Index (kg/m**2)	0.013166091832220537
Current self-reported weight (pounds)	0.012793357189743598
Sagittal Abdominal Diameter 1st (cm)	0.012601497290695249
Waist Circumference (cm)	0.011210460983543964

Experiment : Features' impact

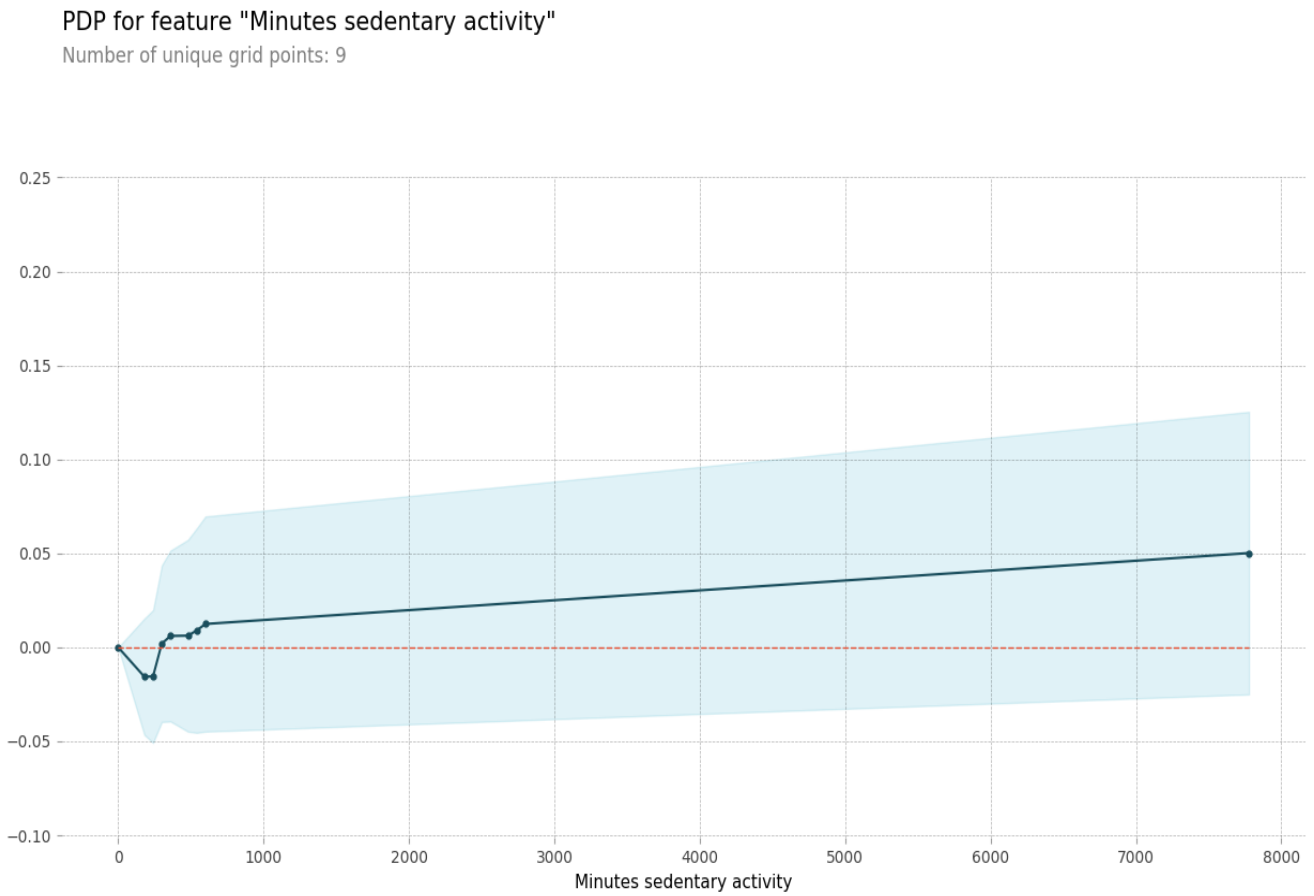
How often do you snore?
v.s. Average Sagittal Abdominal Diameter



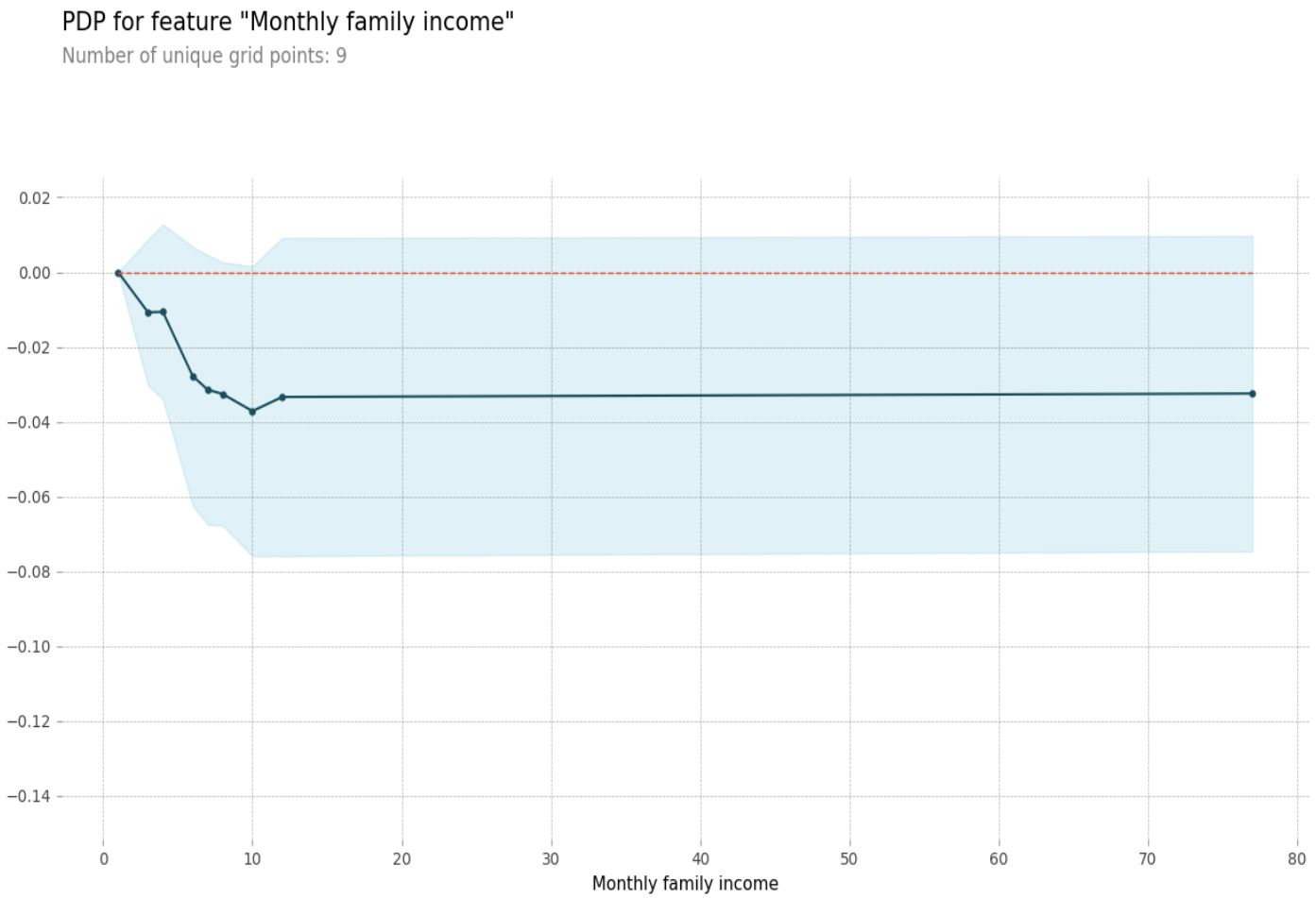
How often do you snore?
v.s. Weight



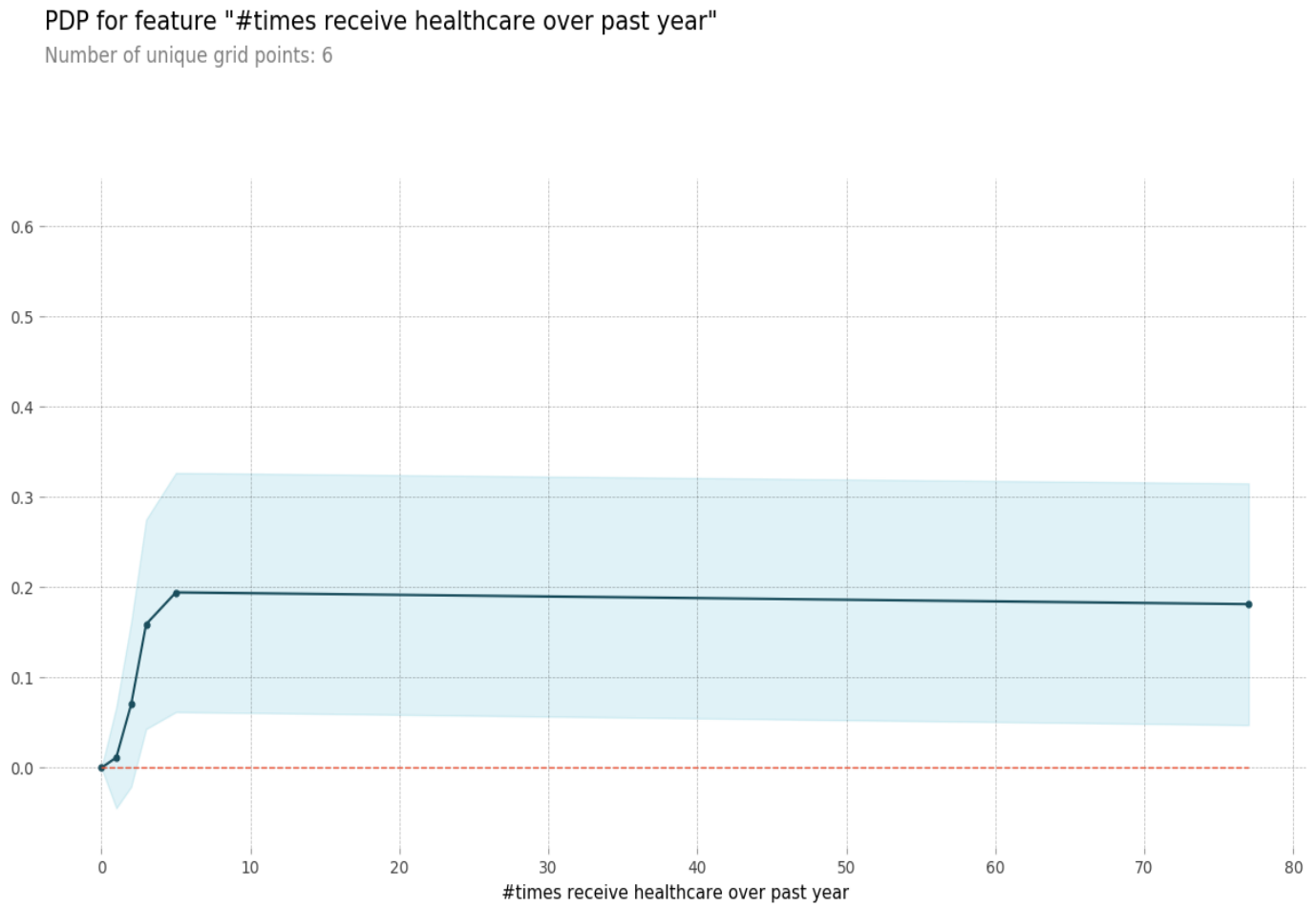
Ever told doctor had trouble sleeping?
v.s. Minutes sedentary activity



Ever told doctor had trouble sleeping?
v.s. Monthly family income



Ever told doctor had trouble sleeping?
v.s. times receive healthcare over past year



Experiment : Shap Values

Shapley value

Given a coalitional game (N, v), the Shapley Value divides payoffs among players according to:

$$\phi_i(N,v) = \frac{1}{N!} \sum_{S \subseteq N \setminus \{i\}} |S|!(|N| - |S| - 1)! \Big[v(S \cup \{i\}) - v(S) \Big].$$

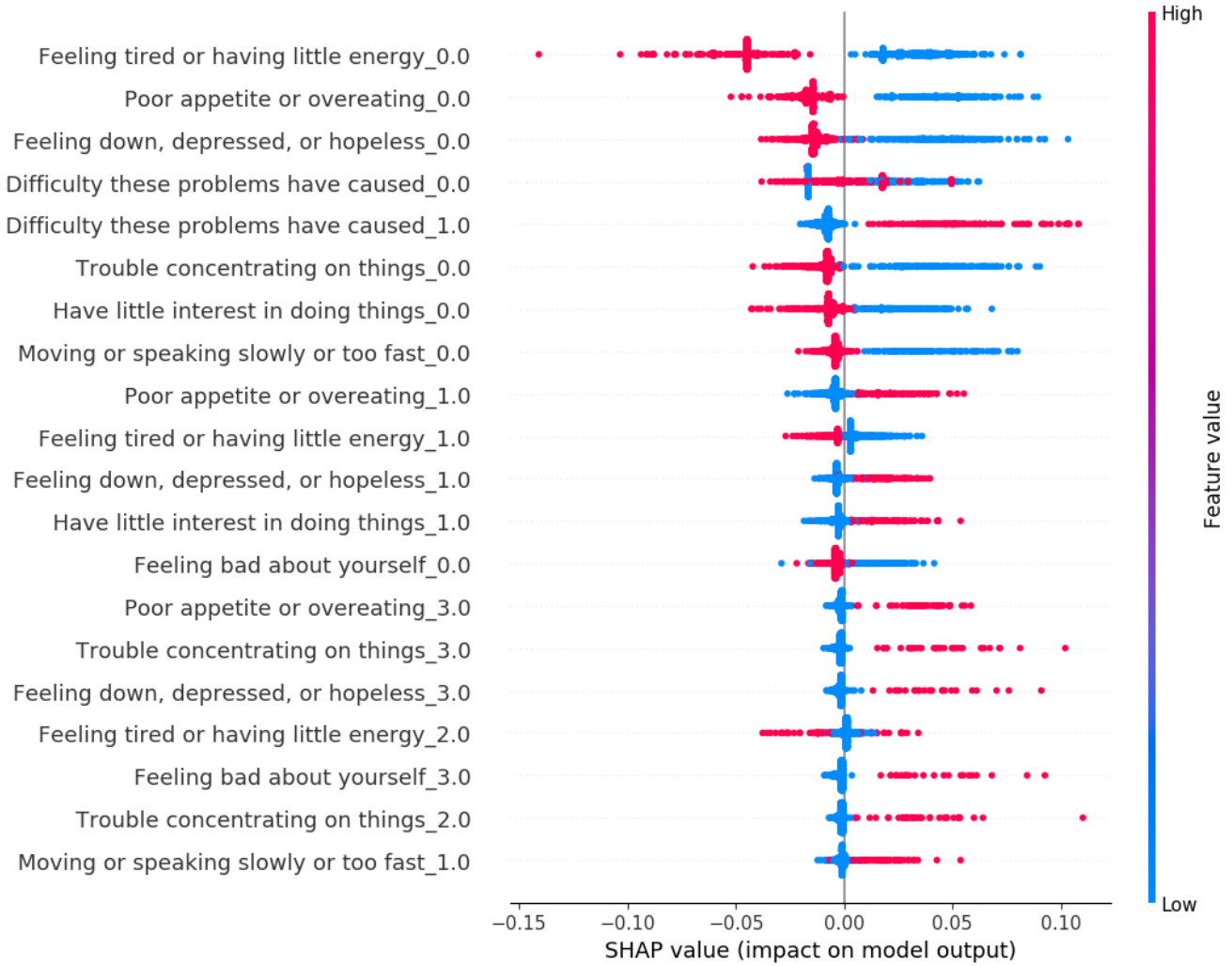
for each player i.

	1	2	3
$1 \leftarrow 2 \leftarrow 3$	0	90	30
$1 \leftarrow 3 \leftarrow 2$	0	40	80
$2 \leftarrow 1 \leftarrow 3$	90	0	30
$2 \leftarrow 3 \leftarrow 1$	50	0	70
$3 \leftarrow 1 \leftarrow 2$	80	40	0
$3 \leftarrow 2 \leftarrow 1$	50	70	0
total	270	240	210
Shapley value	45	40	35

SHAP plot : Mental Health – Depression

- Some Features that have an **important** effect on output.

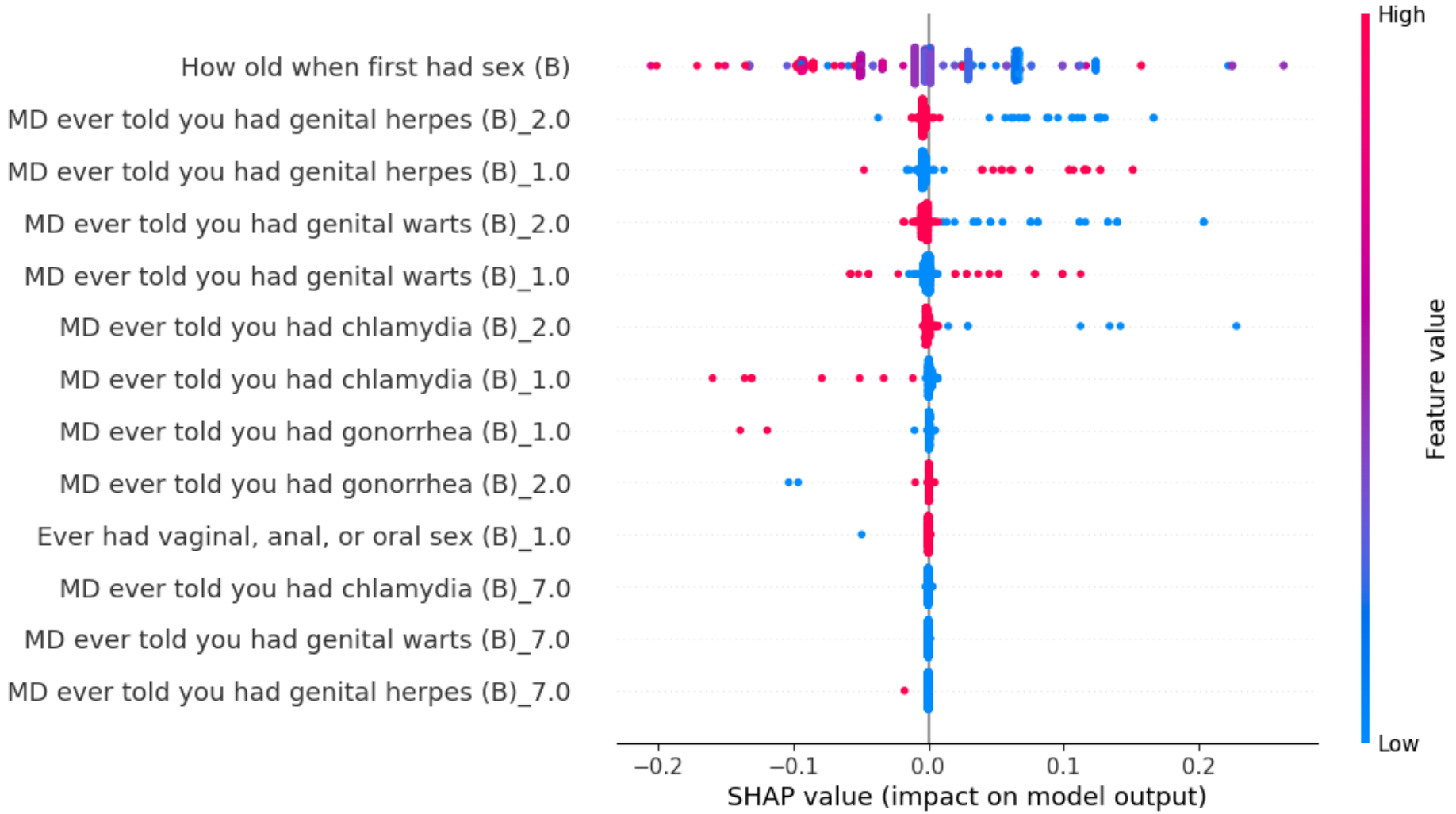
Target : Ever told doctor had trouble sleeping?



SHAP plot : Sexual Behavior

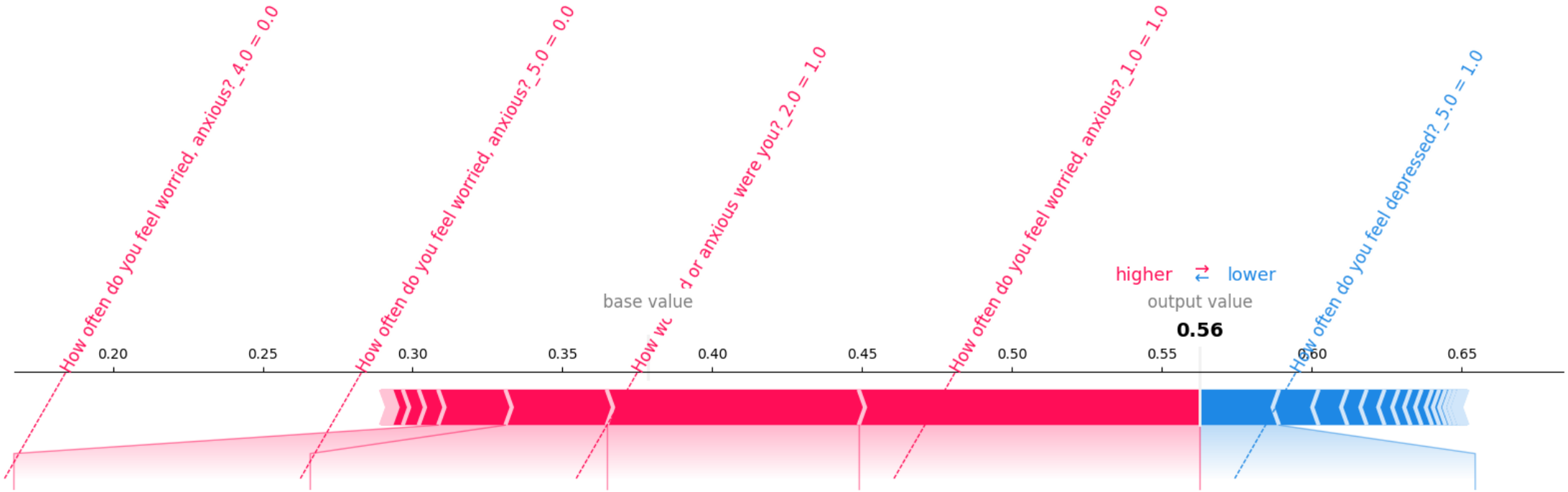
- Some Features that have a **trivial** effect on output.

Target : Ever told doctor had trouble sleeping?



SHAP plot :

- Choose one testing data
- Show how his answer impact on the output



We introduce a Dataset that is widely use around the world.

We present how to preprocess ,which tools we can use to analyze the data.

There are lots of interesting factors when we dig into the results.

Last, we use shapley value to evaluate the importance of features which is good at explaining the reasons of sleep disorder!

Conclusion

Thank you for listening, have a good night