.predict_explain(X)

The Objective

- A machine learning model:
  - *model = …* → *fit(X,y)*
- *A prediction:*
  - *model.predict(X)* → *„Iris-virginica“*
- *A prediction that explains itself:*
  - *model.predict_explain(X)* → *„The prediction is Iris-virginica, because …“*

Start with KNeighborsClassifier

- An inherited machine learning model:
  - *my_knn = my_KneighborsClassifier.fit(X,y)*

- *Add prediction method that explains itself:*
  - *my_knn.predict_explain(X)* → *Prediction*
    Confidence
    Explanation
    Features_Distribution

Nursery Dataset

- Target values

  not_recom
  (recommend)
  very_recom
  priority
  spec_priority

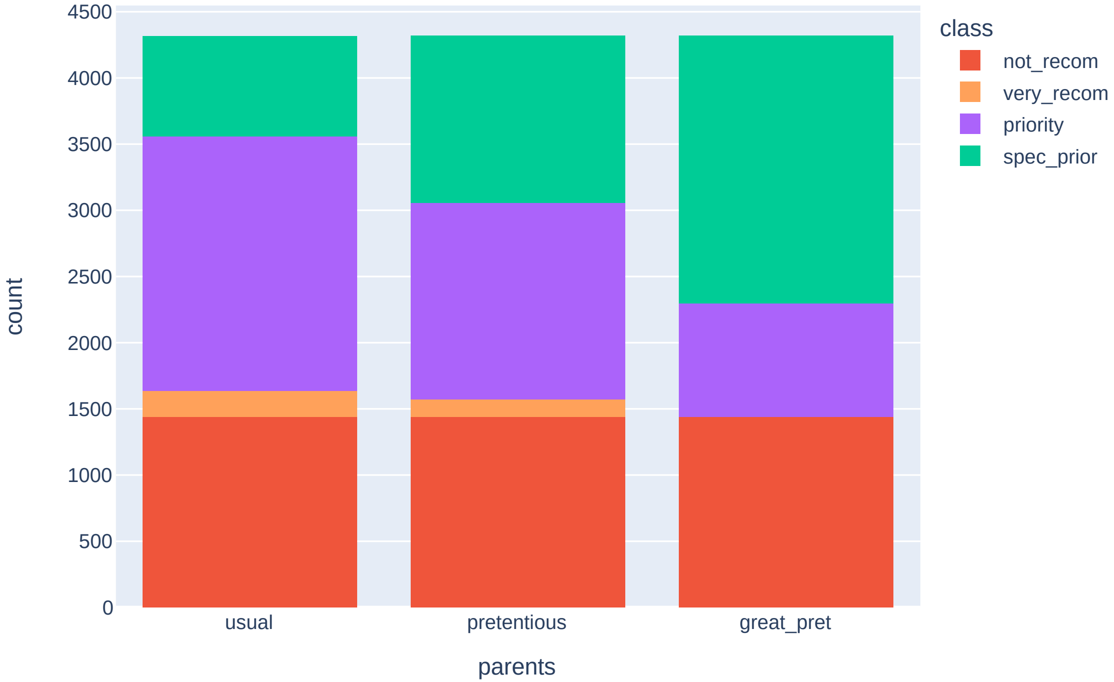
- *Counts:*
  - *12960*
  - *all feature combinations - one time*

- A structured dataset of children conditions

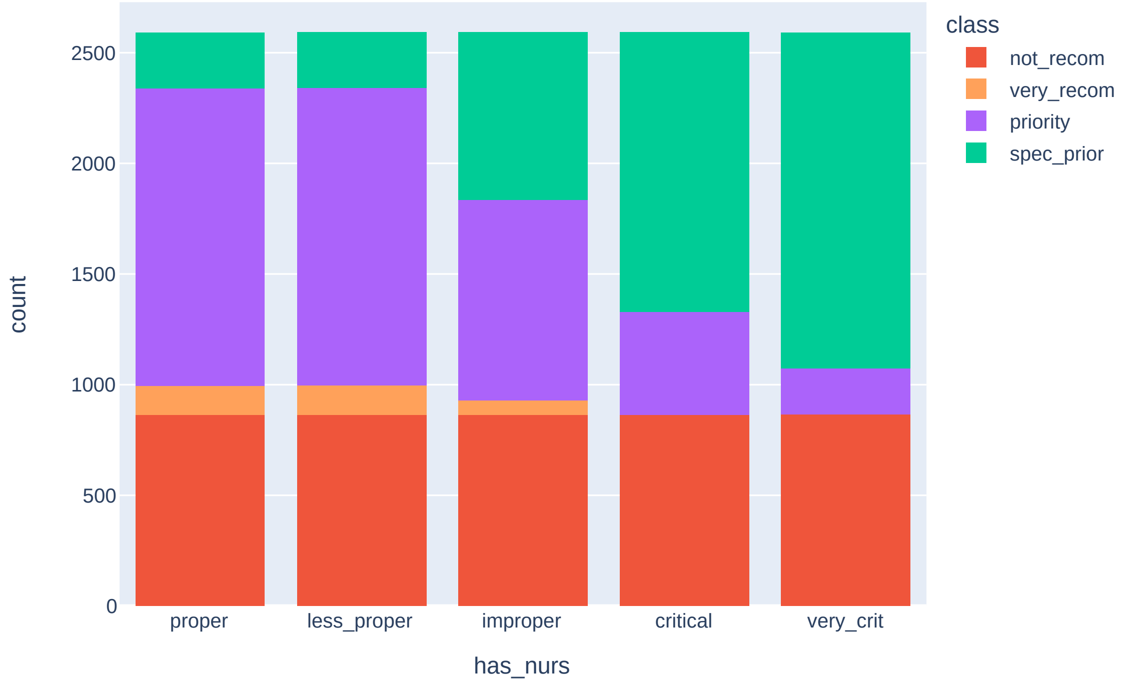
1	*parents*	*Parents' occupation*
2	*has_nurs*	*Child's nursery*
3	*form*	*Form of the family*
4	*children*	*Number of children*
5	*housing*	*Housing conditions*
6	*finance*	*Financial standing of the family*
7	*social*	*Social conditions*
8	*health*	*Health conditions (as veto feature)*

Bohanec, M., Rajkovic, V. (1987). An Expert System Approach to Multi-Attribute Decision Making.

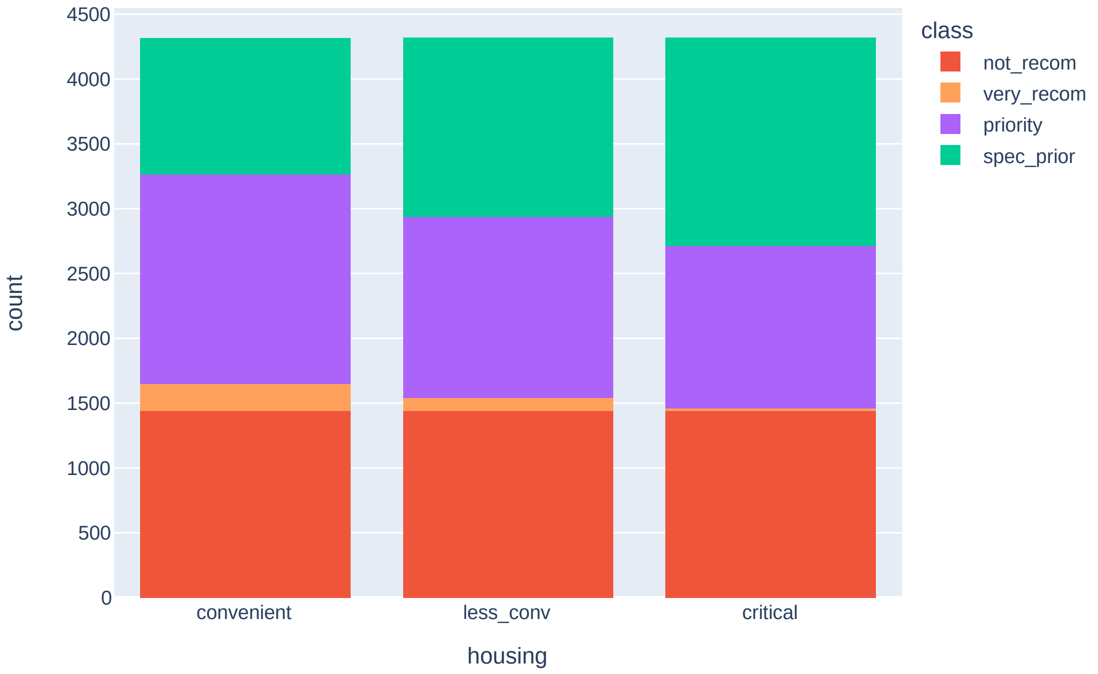
Nursery Dataset



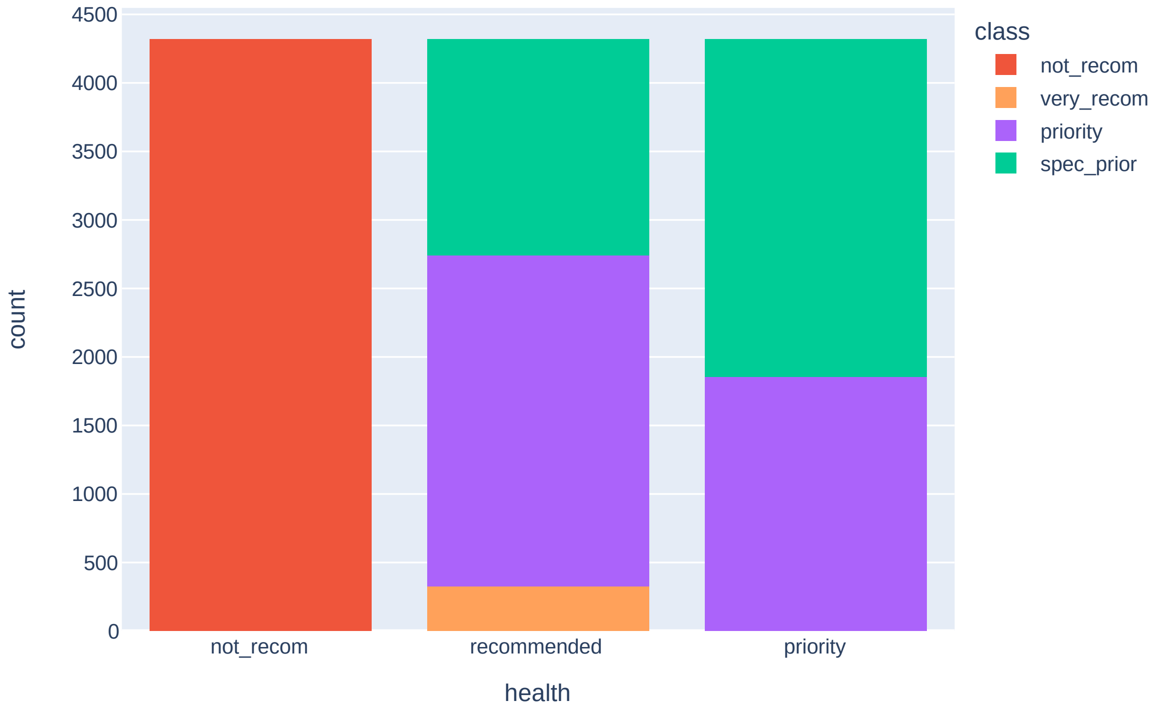
Nursery Dataset



Nursery Dataset



Nursery Dataset



my_knn.predict_explain(child_n)

- Should explain its prediction method
  - *„not_recom“, because […]*
  - *„very_recom“, because […]*
  - *„priority“, because […]*
  - *„spec_priority“, because […]*

- Should recognize features structure of child

1. *parents*
2. *has_nurs*
3. *form*
4. *children*
5. *housing*
6. *finance*
7. *social*
8. *health (as veto)*

my_knn.predict_explain(child_1)

- Prediction: '*not_recom*'

- *Confidence: True*

my_knn.predict_explain(child_1)

- *Explanation:*

  "The prediction 'not_recom' is quite sure:

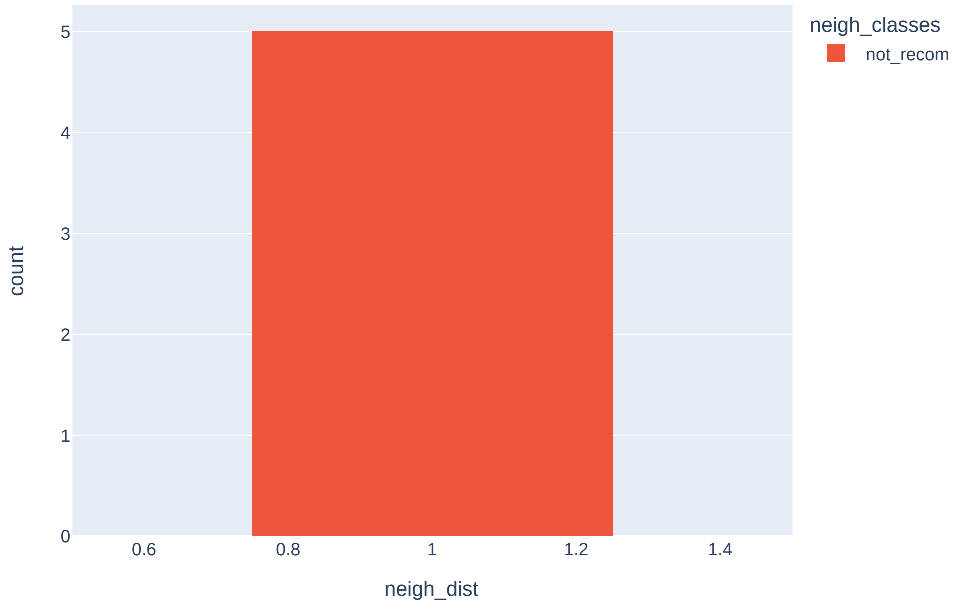
my_knn.predict_explain(child_1)

- *Explanation:*
  - *"The prediction 'not_recom' is quite sure:*
  - *On the one hand the 5 nearest neighbours have homogeneous target values (5x value 'not_recom').*

my_knn.predict_explain(child_1)

- *Explanation:*
  - *"The prediction 'not_recom' is quite sure:*
  - *On the one hand the 5 nearest neighbours have homogeneous target values (5x value 'not_recom').*
  - *And on the other hand the nearest neighbour has the same target value too.",*

my_knn.predict_explain(child_1)



`my_knn.predict_explain(child_2)`

- Prediction: '*very_recom*'
- *Confidence: False*

`my_knn.predict_explain(child_2)`

- *Explanation:*

  "The prediction 'very_recom' is rather unsure:

my_knn.predict_explain(child_2)

- *Explanation:*

  "The prediction 'very_recom' is rather unsure:

  - *On the one hand the 5 nearest neighbours have diverse target values (2x value 'priority', 3x value 'very_recom').*

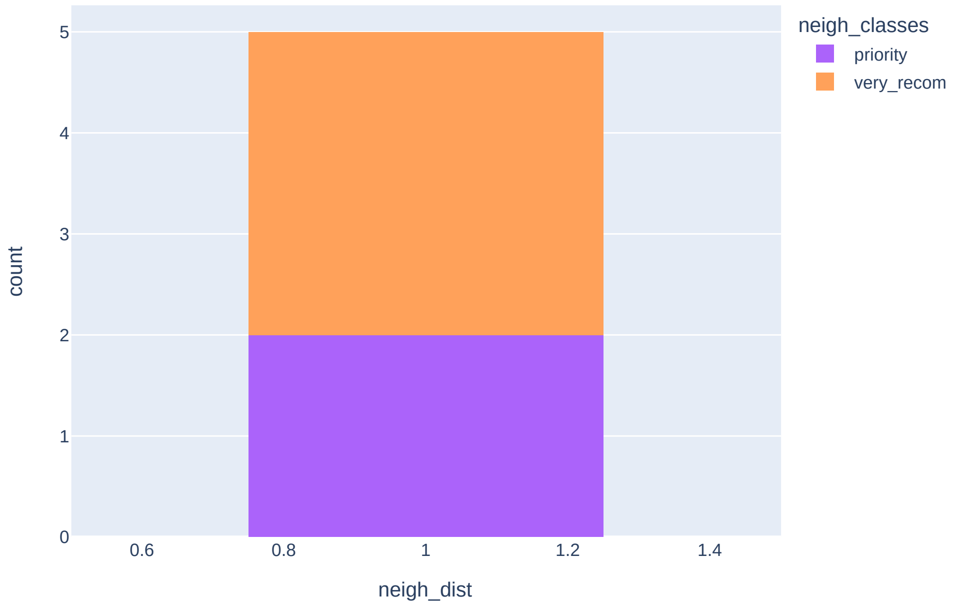
my_knn.predict_explain(child_2)

- *Explanation:*

  "The prediction 'very_recom' is rather unsure:

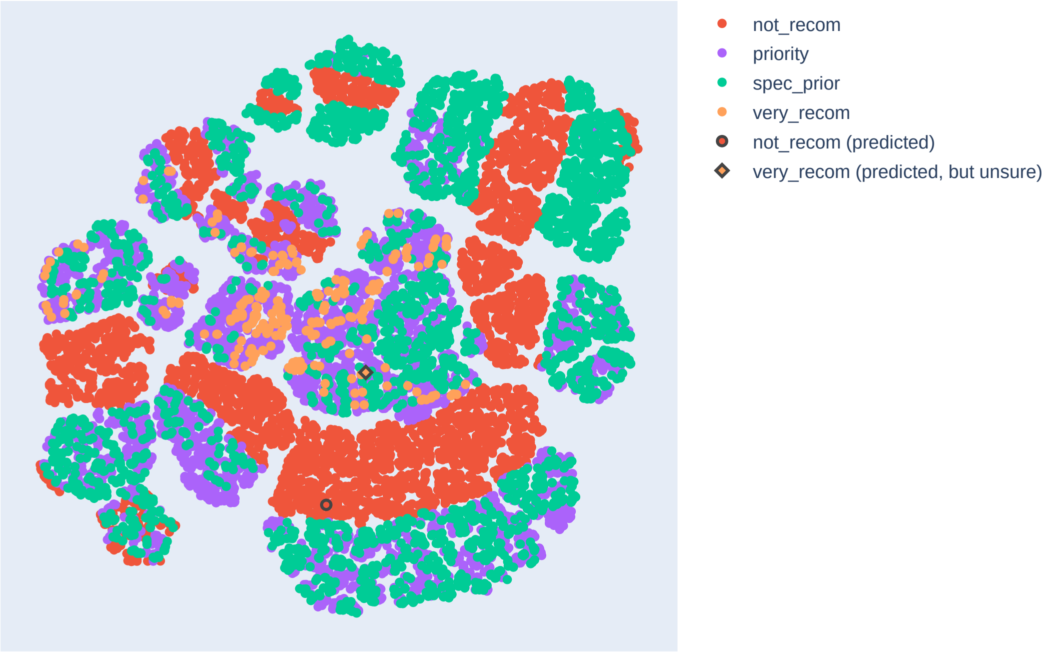
  - *On the one hand the 5 nearest neighbours have diverse target values (2x value 'priority', 3x value 'very_recom').*
  - *And on the other hand the nearest neighbour has another target value ('priority') as the prediction."*

my_knn.predict_explain(child_2)

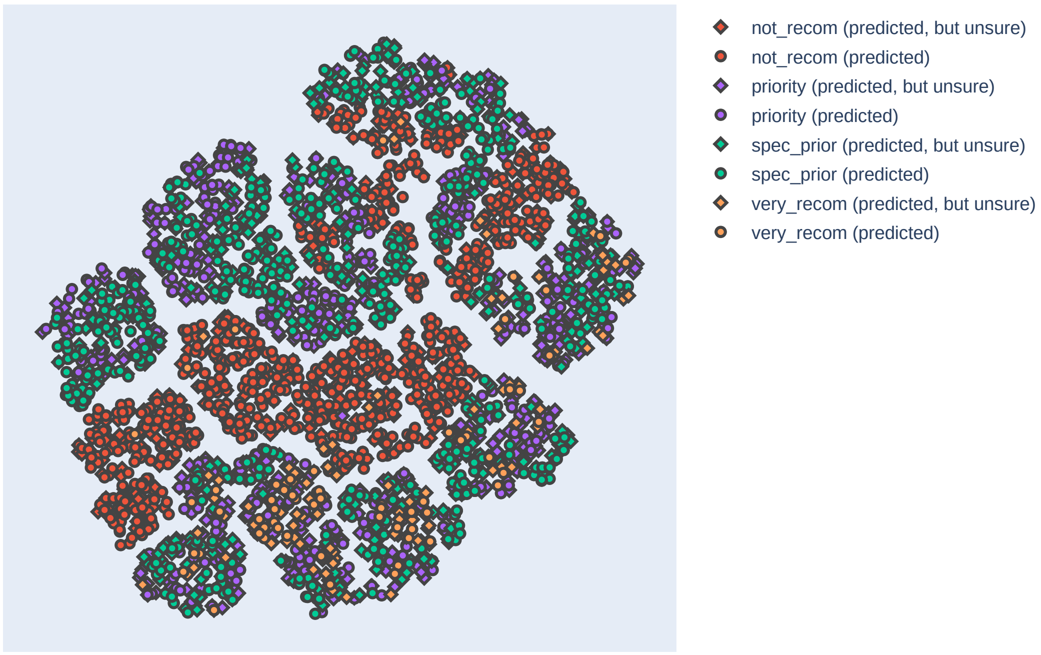


my_knn.predict_explain(child_1,2)

Dimensionality reduction for y_predict_explain: TSNE visualization

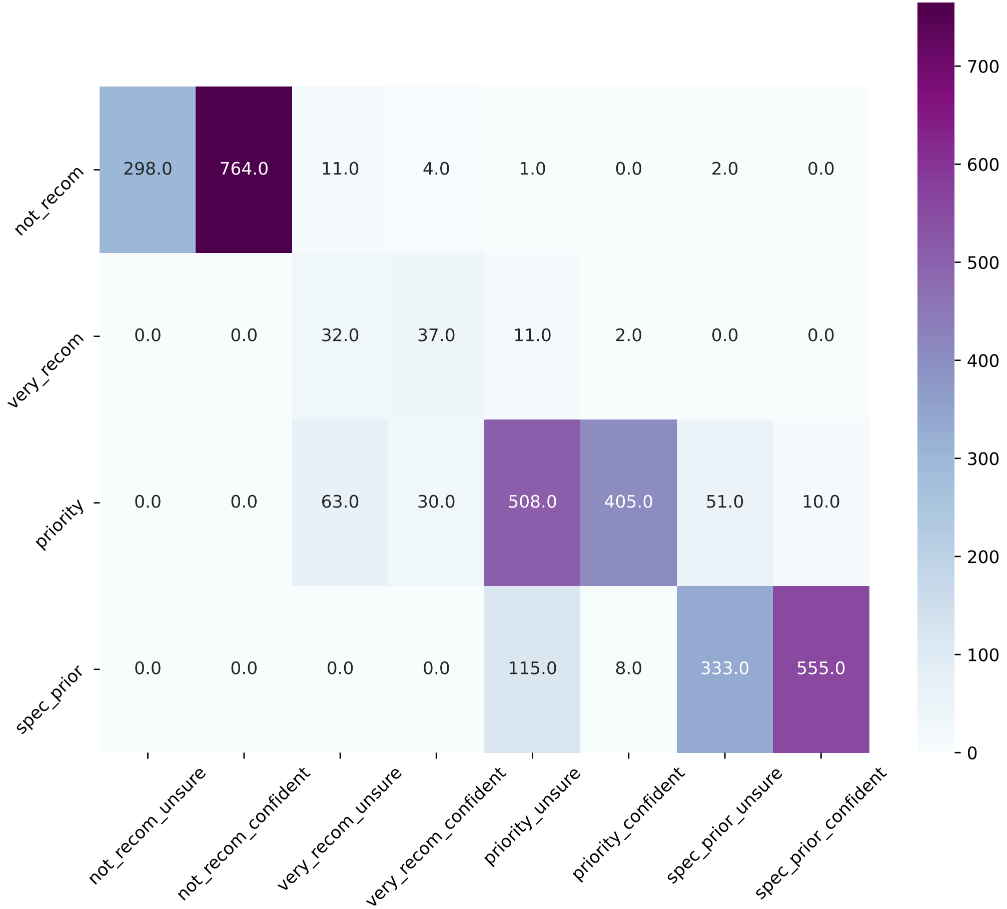


my_knn.predict_explain(all)



Dimensionality reduction for y_predict_explain: TSNE visualization

my_knn.predict_explain(all)



	not_recom_unsure	not_recom_confident	very_recom_unsure	very_recom_confident	priority_unsure	priority_confident	spec_prior_unsure	spec_prior_confident
not_recom	298.0	764.0	11.0	4.0	1.0	0.0	2.0	0.0
very_recom	0.0	0.0	32.0	37.0	11.0	2.0	0.0	0.0
priority	0.0	0.0	63.0	30.0	508.0	405.0	51.0	10.0
spec_prior	0.0	0.0	0.0	0.0	115.0	8.0	333.0	555.0

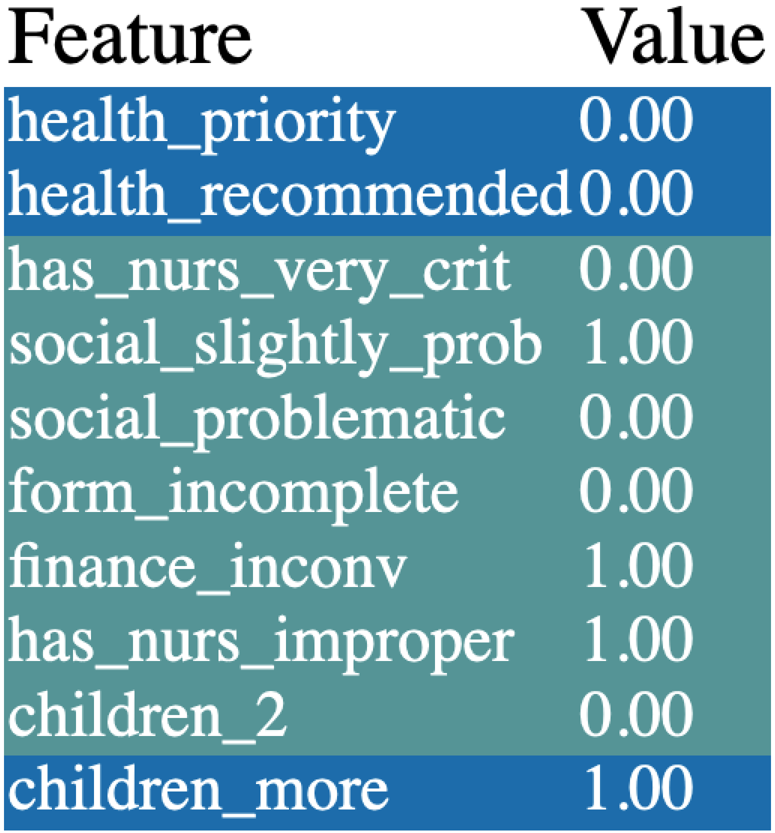
my_knn.predict_explain(child_n)

- Should explain its prediction method
  - *„not_recom“, because […]*
  - *„very_recom“, because […]*
  - *„priority“, because […]*
  - *„spec_priority“, because […]*

- Should recognize features structure of child

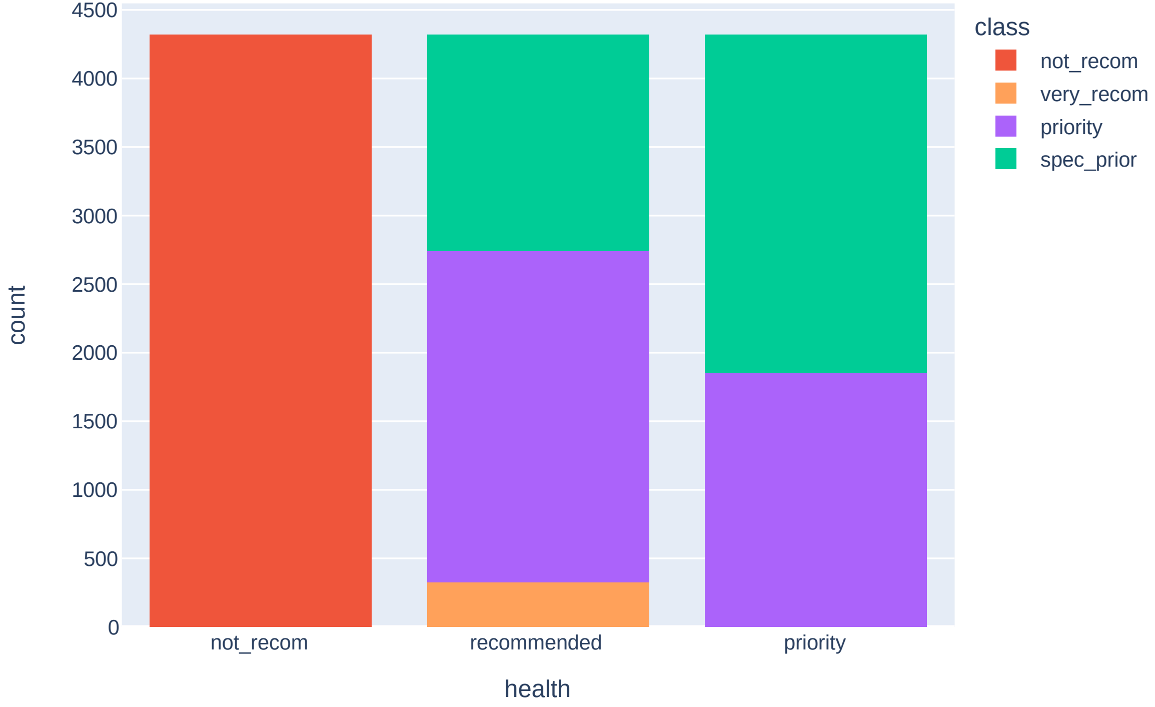
1. *parents*
2. *has_nurs*
3. *form*
4. *children*
5. *housing*
6. *finance*
7. *social*
8. *health (as veto)*

Features(child_1)



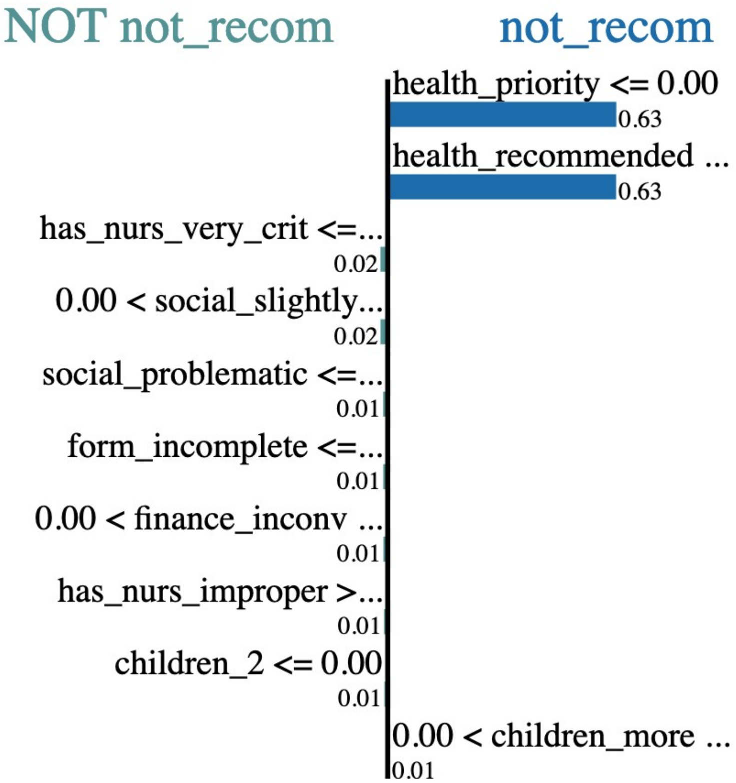
Feature	Value
health_priority	0.00
health_recommended	0.00
has_nurs_very_crit	0.00
social_slightly_prob	1.00
social_problematic	0.00
form_incomplete	0.00
finance_inconv	1.00
has_nurs_improper	1.00
children_2	0.00
children_more	1.00

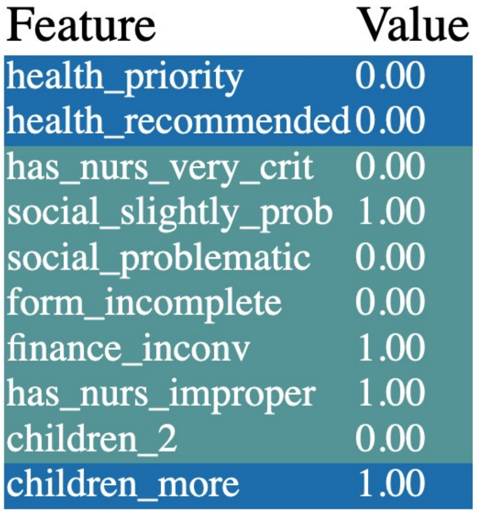
Features(child_1)



Features(child_1)

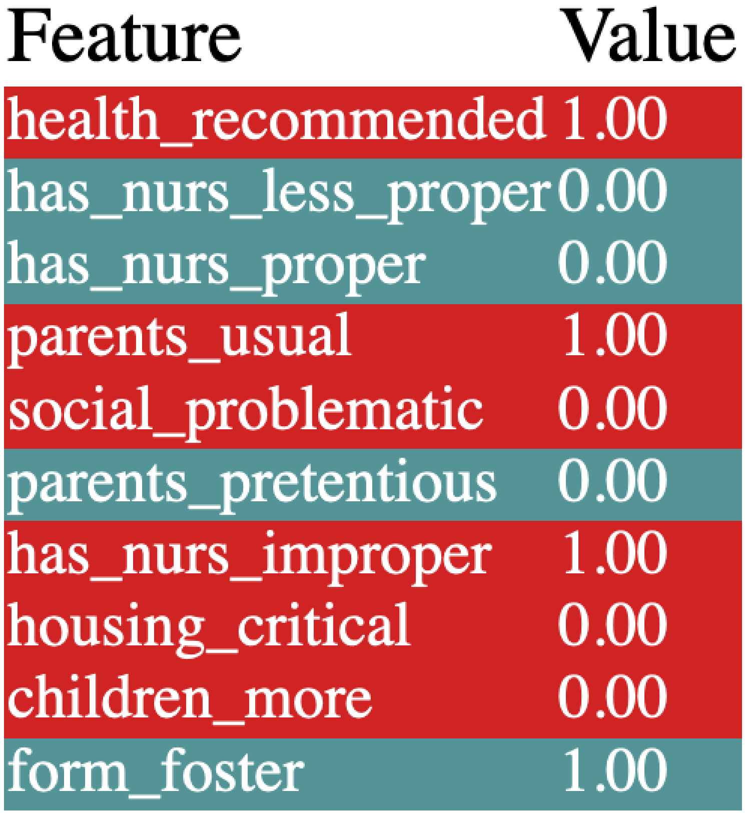
- LIME





Feature	Value
health_priority	0.00
health_recommended	0.00
has_nurs_very_crit	0.00
social_slightly_prob	1.00
social_problematic	0.00
form_incomplete	0.00
finance_inconv	1.00
has_nurs_improper	1.00
children_2	0.00
children_more	1.00

Features(child_2)



Feature	Value
health_recommended	1.00
has_nurs_less_proper	0.00
has_nurs_proper	0.00
parents_usual	1.00
social_problematic	0.00
parents_pretentious	0.00
has_nurs_improper	1.00
housing_critical	0.00
children_more	0.00
form_foster	1.00

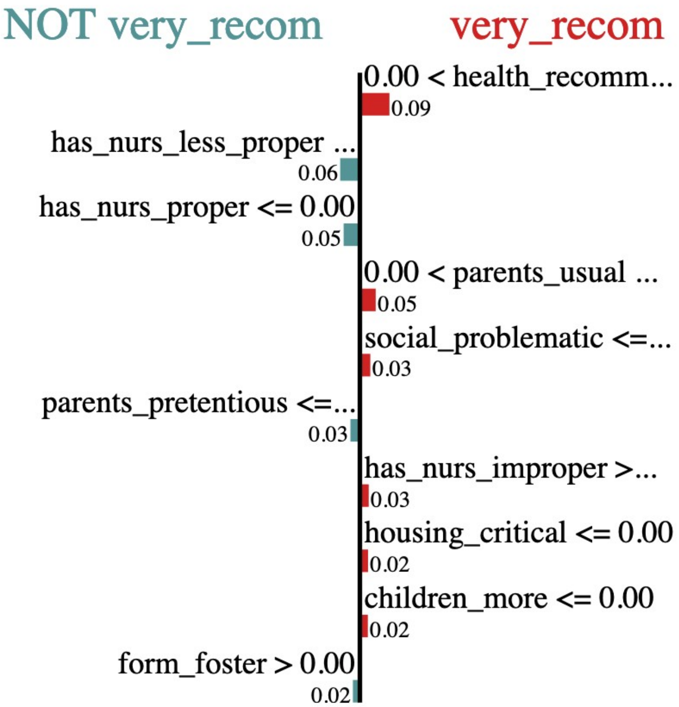
Features(child_2)

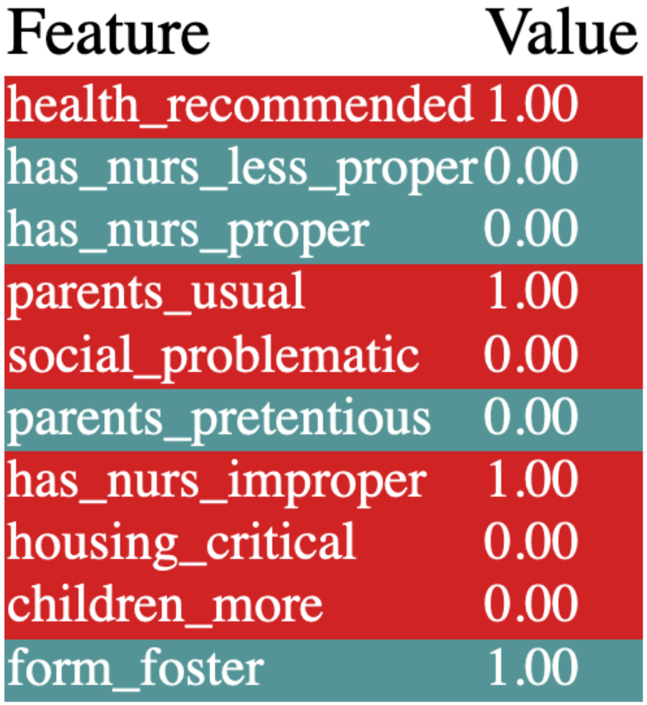
- A structured dataset of children conditions

1	*parents*	*Parents' occupation*
2	*has_nurs*	*Child's nursery*
3	*form*	*Form of the family*
4	*children*	*Number of children*
5	*housing*	*Housing conditions*
6	*finance*	*Financial standing of the family*
7	*social*	*Social conditions*
8	*health*	*Health conditions (as veto feature)*

Features(child_2)

- LIME

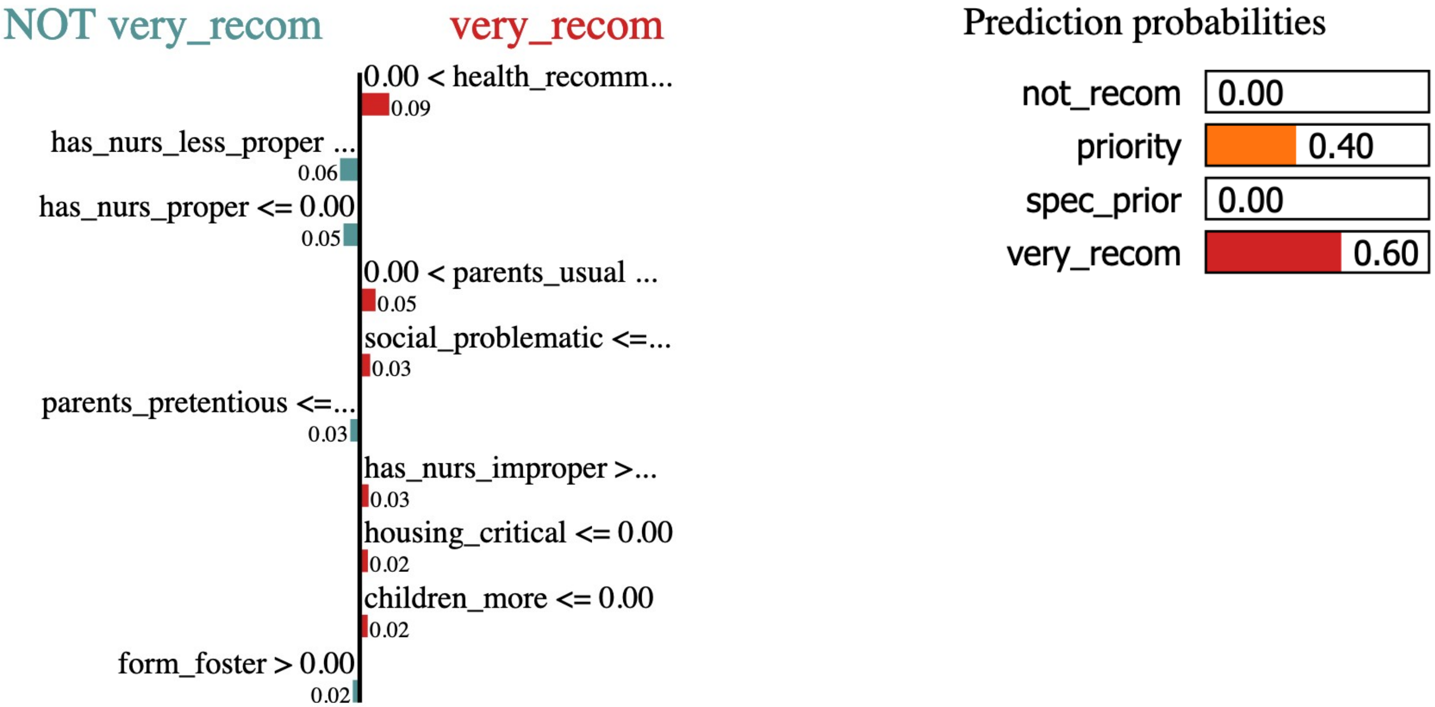




Feature	Value
health_recommended	1.00
has_nurs_less_proper	0.00
has_nurs_proper	0.00
parents_usual	1.00
social_problematic	0.00
parents_pretentious	0.00
has_nurs_improper	1.00
housing_critical	0.00
children_more	0.00
form_foster	1.00

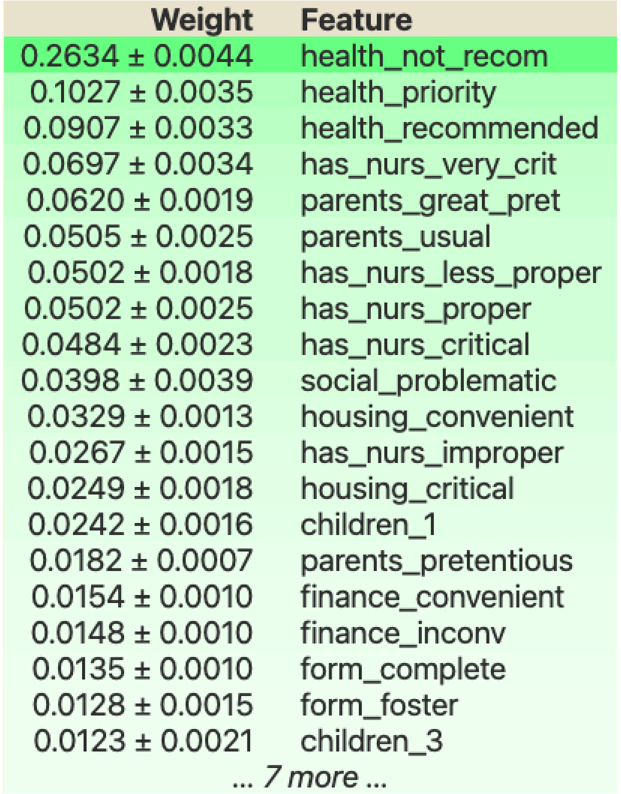
Features(child_2)

- LIME



Features(all)

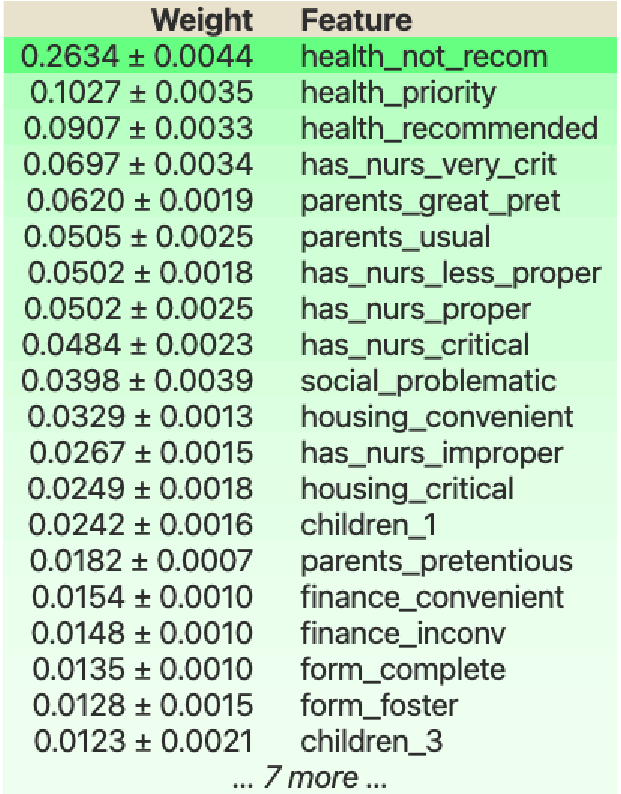
- eli5



Weight	Feature
0.2634 ± 0.0044	health_not_recom
0.1027 ± 0.0035	health_priority
0.0907 ± 0.0033	health_recommended
0.0697 ± 0.0034	has_nurs_very_crit
0.0620 ± 0.0019	parents_great_pret
0.0505 ± 0.0025	parents_usual
0.0502 ± 0.0018	has_nurs_less_proper
0.0502 ± 0.0025	has_nurs_proper
0.0484 ± 0.0023	has_nurs_critical
0.0398 ± 0.0039	social_problematic
0.0329 ± 0.0013	housing_convenient
0.0267 ± 0.0015	has_nurs_improper
0.0249 ± 0.0018	housing_critical
0.0242 ± 0.0016	children_1
0.0182 ± 0.0007	parents_pretentious
0.0154 ± 0.0010	finance_convenient
0.0148 ± 0.0010	finance_inconv
0.0135 ± 0.0010	form_complete
0.0128 ± 0.0015	form_foster
0.0123 ± 0.0021	children_3
… 7 more …	

Features(all)

- eli5



Weight	Feature
0.2634 ± 0.0044	health_not_recom
0.1027 ± 0.0035	health_priority
0.0907 ± 0.0033	health_recommended
0.0697 ± 0.0034	has_nurs_very_crit
0.0620 ± 0.0019	parents_great_pret
0.0505 ± 0.0025	parents_usual
0.0502 ± 0.0018	has_nurs_less_proper
0.0502 ± 0.0025	has_nurs_proper
0.0484 ± 0.0023	has_nurs_critical
0.0398 ± 0.0039	social_problematic
0.0329 ± 0.0013	housing_convenient
0.0267 ± 0.0015	has_nurs_improper
0.0249 ± 0.0018	housing_critical
0.0242 ± 0.0016	children_1
0.0182 ± 0.0007	parents_pretentious
0.0154 ± 0.0010	finance_convenient
0.0148 ± 0.0010	finance_inconv
0.0135 ± 0.0010	form_complete
0.0128 ± 0.0015	form_foster
0.0123 ± 0.0021	children_3

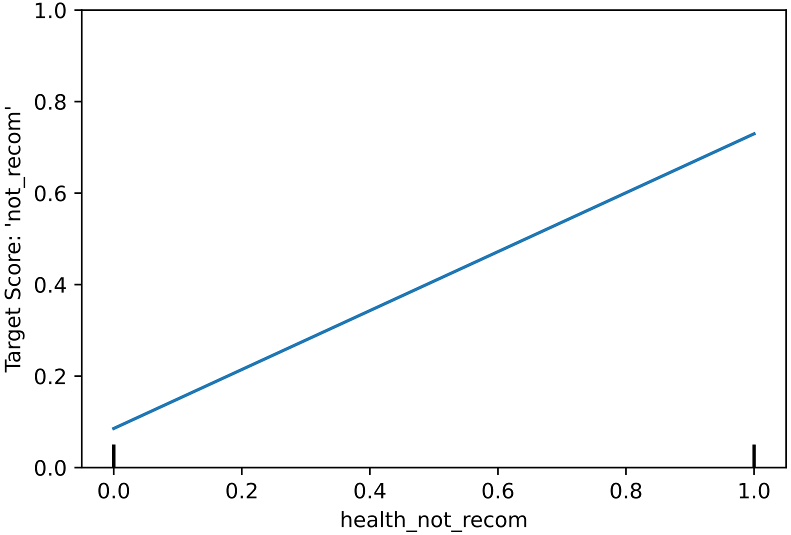
... 7 more ...

- A structured dataset of children conditions

1	*parents*	*Parents' occupation*
2	*has_nurs*	*Child's nursery*
3	*form*	*Form of the family*
4	*children*	*Number of children*
5	*housing*	*Housing conditions*
6	*finance*	*Financial standing of the family*
7	*social*	*Social conditions*
8	*health*	*Health conditions (as veto feature)*

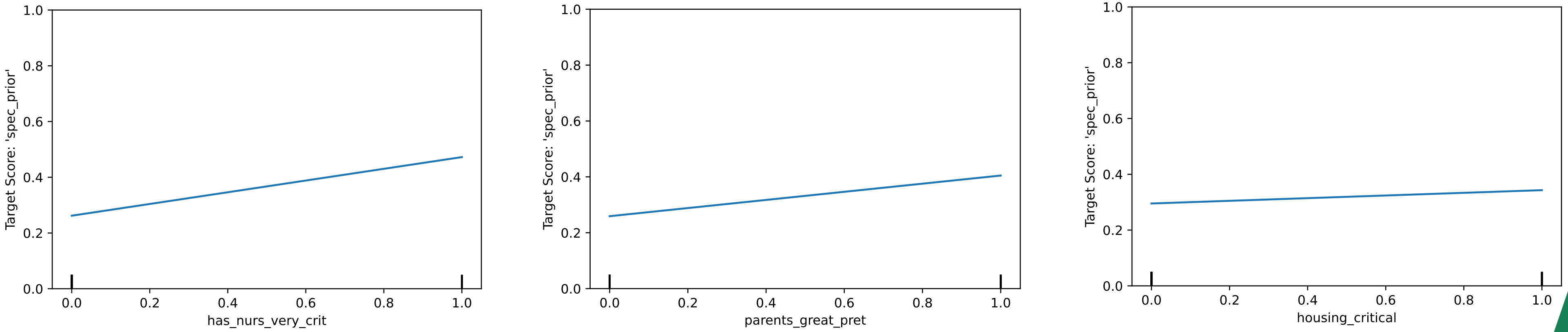
Features(all)

- PDP



Features(all)

- PDP



Merci!