Florian Schauer, Franca Speth

DBMS Praktikum

Comparison of Interpretability Tools for a Human Resource Dataset
Summer Term 2020

Agenda

1. Problem Definition
2. Data Preparation
3. Model Training
4. Model Evaluation
5. Interpret Model with InterpretML
6. Interpret Model with AIX360
7. Comparison of Tools

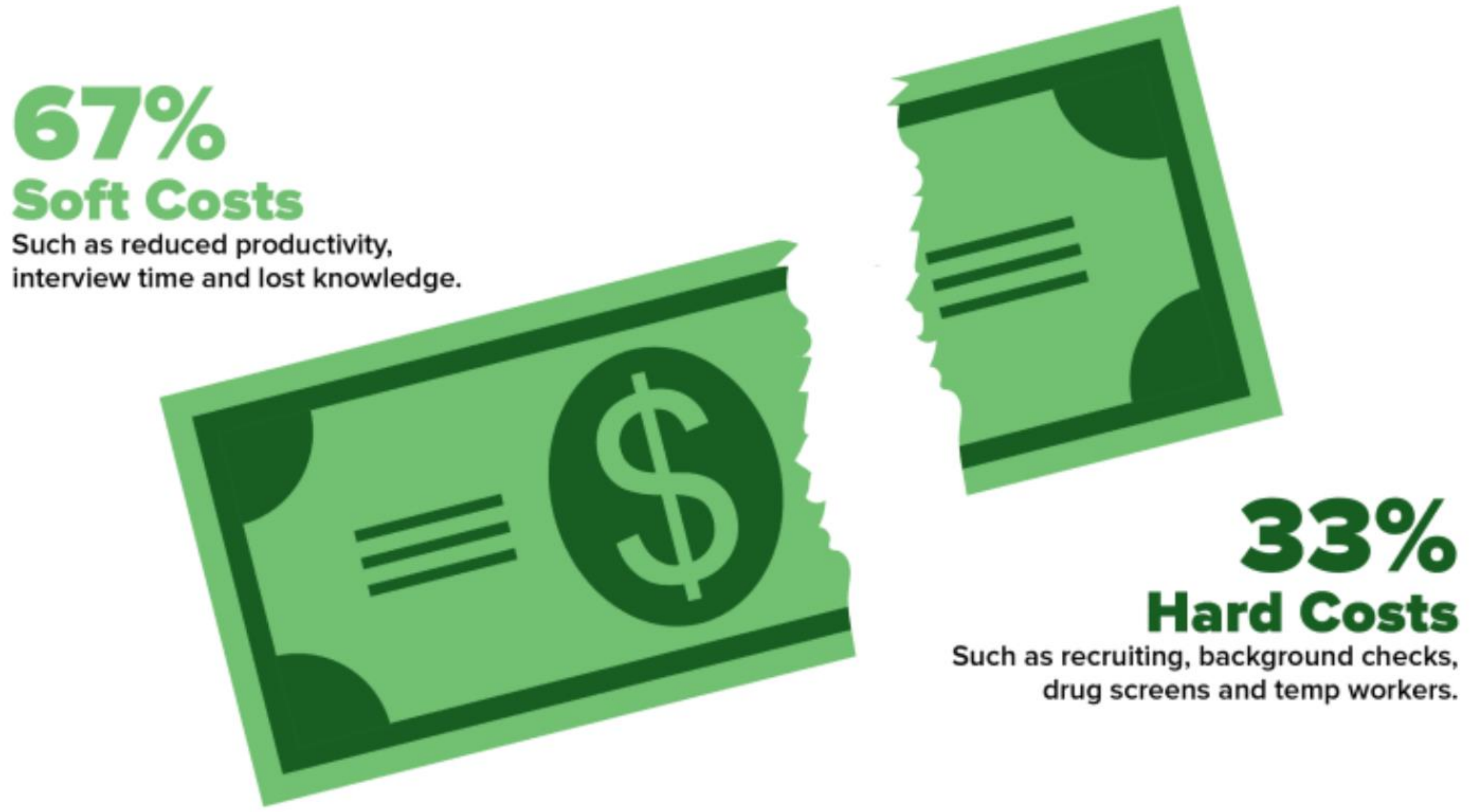
Use Case and Research Question

Employee departures cost a company time, money, and other resources.

SHRM, 2008

Employee turnover can cost up to 150% of his/her annual income.

W. G. Bliss, 2011



Is it possible to predict whether an employee is going to leave the company?

How does the model make the prediction ?

Picture: https://www.shrm.org/hr-today/news/all-things-work/pages/to-have-and-to-hold.aspx, Allen, D. G. (2008). *Retaining talent: A guide to analyzing and managing employee turnover*. SHRM Foundations. Bliss, W. G. (2011). The Advisor-Cost of Employee Turnover

Dataset and Source of the Data

	MarriedID	MaritalStatusID	GenderID	DeptID	PerfScoreID	FromDiversityJobFairID	PayRate	PositionID
1	1	1	0	1	3	1	28	1
2	0	2	1	1	3	0	23	1
3	0	0	1	1	3	0	29	1
4	1	1	0	1	3	0	22	2
5	0	0	0	1	3	0	17	2
6	1	1	0	1	3	1	20	2
7	1	1	0	6	3	0	55	3
8	0	0	0	6	3	0	55	3
9	0	0	0	6	1	0	55	3
10	1	1	1	6	3	0	56	3

Sex	HispanicLatino	ManagerID	EngagementSurvey	EmpSatisfaction	SpecialProjectsCount	DaysLateLast30
F	No	1	2	2	6	0
M	No	1	5	4	4	0
M	No	1	4	5	5	0
F	No	1	3	3	4	-99
F	No	1	5	3	5	0
F	No	1	4	4	4	-99
F	No	17	3	5	0	-99
F	No	17	5	5	0	0
F	Yes	17	2	1	0	0
M	No	17	4	5	0	0

kaggle

Usability 9.4/10

Open License

https://creativecommons.org/licenses/by-sa/4.0/

Code examples available

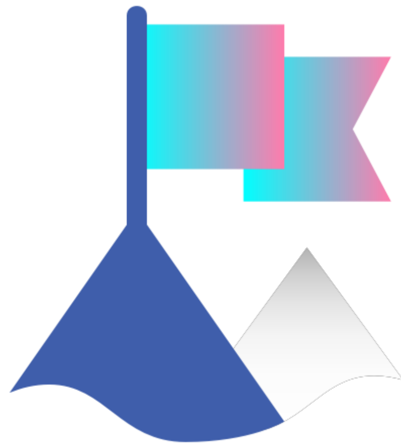
Source of Data: https://www.kaggle.com/rhuebner/human-resources-data-set

Dataset and Data Preparation

	MarriedID	MaritalStatusID	GenderID	DeptID	PerfScoreID	FromDiversityJobFairID	PayRate	PositionID
1	1	1	0	1	3	1	28	1
2	0	2	1	1	3	0	23	1
3	0	0	1	1	3	0	29	1
4	1	1	0	1	3	0	22	2
5	0	0	0	1	3	0	17	2
6	1	1	0	1	3	1	20	2
7	1	1	0	6	3	0	55	3
8	0	0	0	6	3	0	55	3
9	0	0	0	6	1	0	55	3
10	1	1	1	6	3	0	56	3

Sex	HispanicLatino	ManagerID	EngagementSurvey	EmpSatisfaction	SpecialProjectsCount	DaysLateLast30
F	No	1	2	2	6	0
M	No	1	5	4	4	0
M	No	1	4	5	5	0
F	No	1	3	3	4	-99
F	No	1	5	3	5	0
F	No	1	4	4	4	-99
F	No	17	3	5	0	-99
F	No	17	5	5	0	0
F	Yes	17	2	1	0	0
M	No	17	4	5	0	0

- replace missing values with -99
  - DaysLateLast30 and ManagerID
- round float variables and convert them to integer
  - PayRate and EnagagementSurvey
- dummy coding of categorical variables
  - Sex and HispanicLatino
- Split 80% Training Data and 20 % Test Data

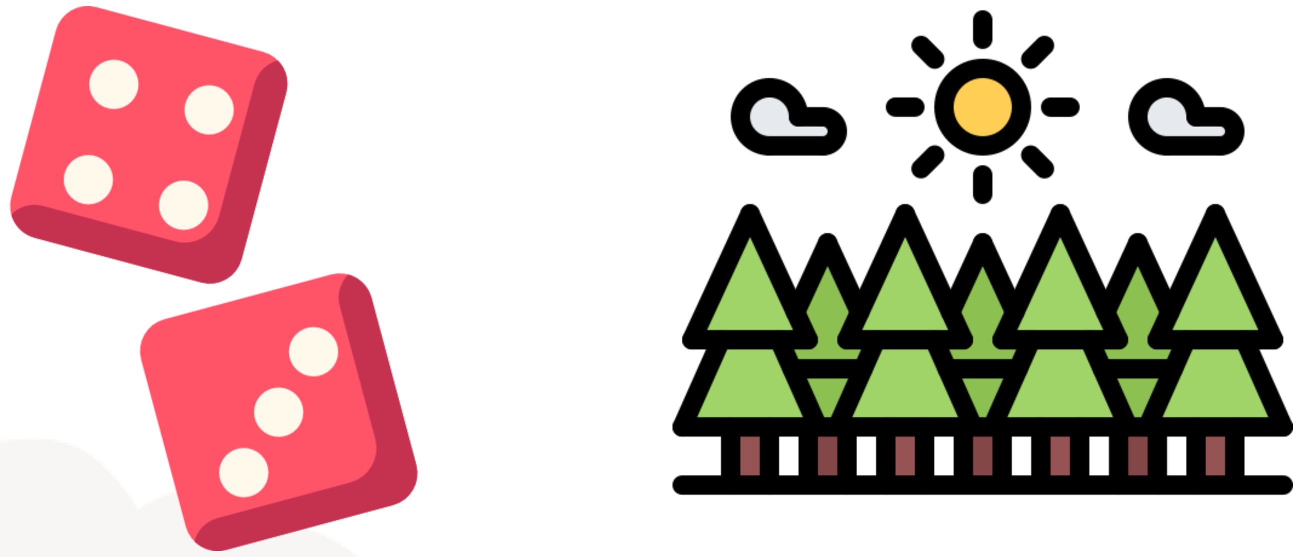


Target variable: Termd

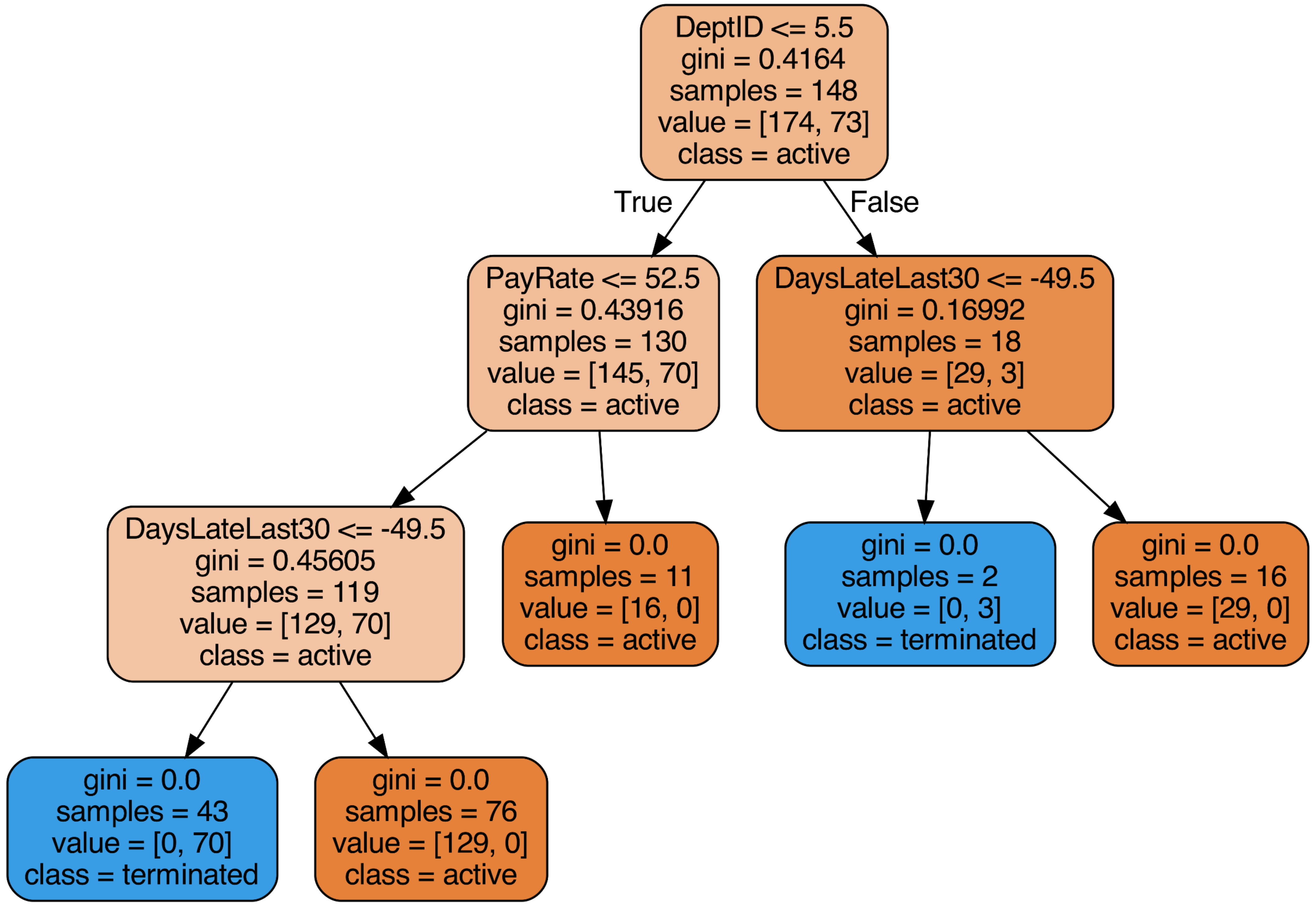
- for 0 = still working for the company
- for 1 = terminated

Source of Data: https://www.kaggle.com/rhuebner/human-resources-data-set

Data Analysis with Random Forest



Visualisation of Tree Nr. 80 for our Input Data:



- Random sampling of training data points when building trees
- Random subsets of features considered when splitting nodes

Result of first Analysis: 100 % Accuracy for Test Data !?

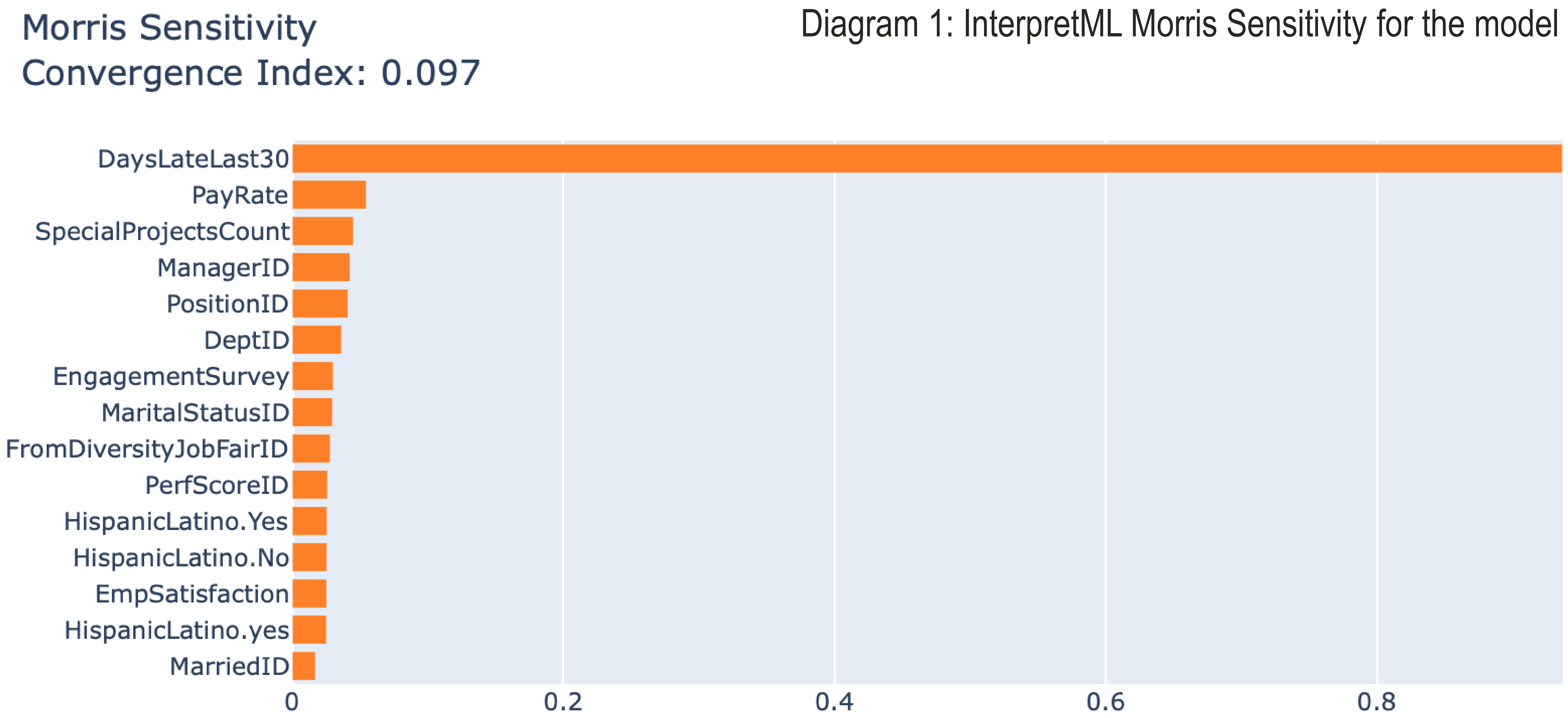
Source: Towards Data Science. (2020). An Implementation and Explanation of the Random Forest in Python. URL: https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76

Insights of first Analysis with InterpretML

The variable DayLateLast30 seems to have a very strong influence, but **why**?

Instead of replacing the missing values of that variable with -99 we tried droping the data with missing values.

Diagram 1: InterpretML Morris Sensitivity for the model



Result: only one class remained, thats why the variable was so important and the accuracy reached 100 %
InterpretML was helpful to detect that the variable should *not* be used for the model because it leads to bias.

New Model after Bias was detected

Model information:

- number of trees in the forest = 100
- bootstrapping = True
- Random state = 1

Results for prediction:

- Accuracy value = 0.77
- Precision for active (class 0) = 0.75
- Precision for terminated (class 1) = 0.89

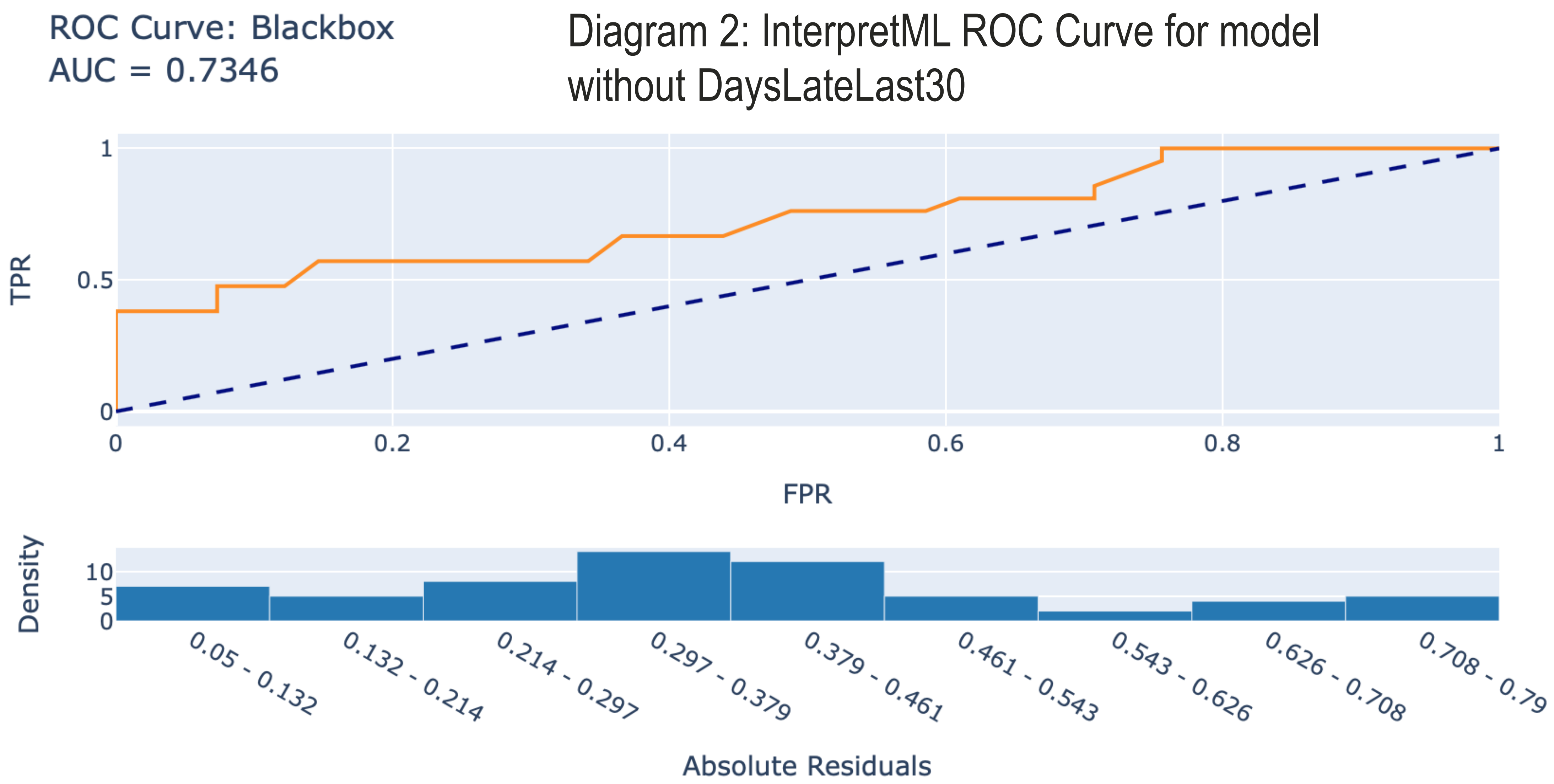


Diagram 2: InterpretML ROC Curve for model without DaysLateLast30

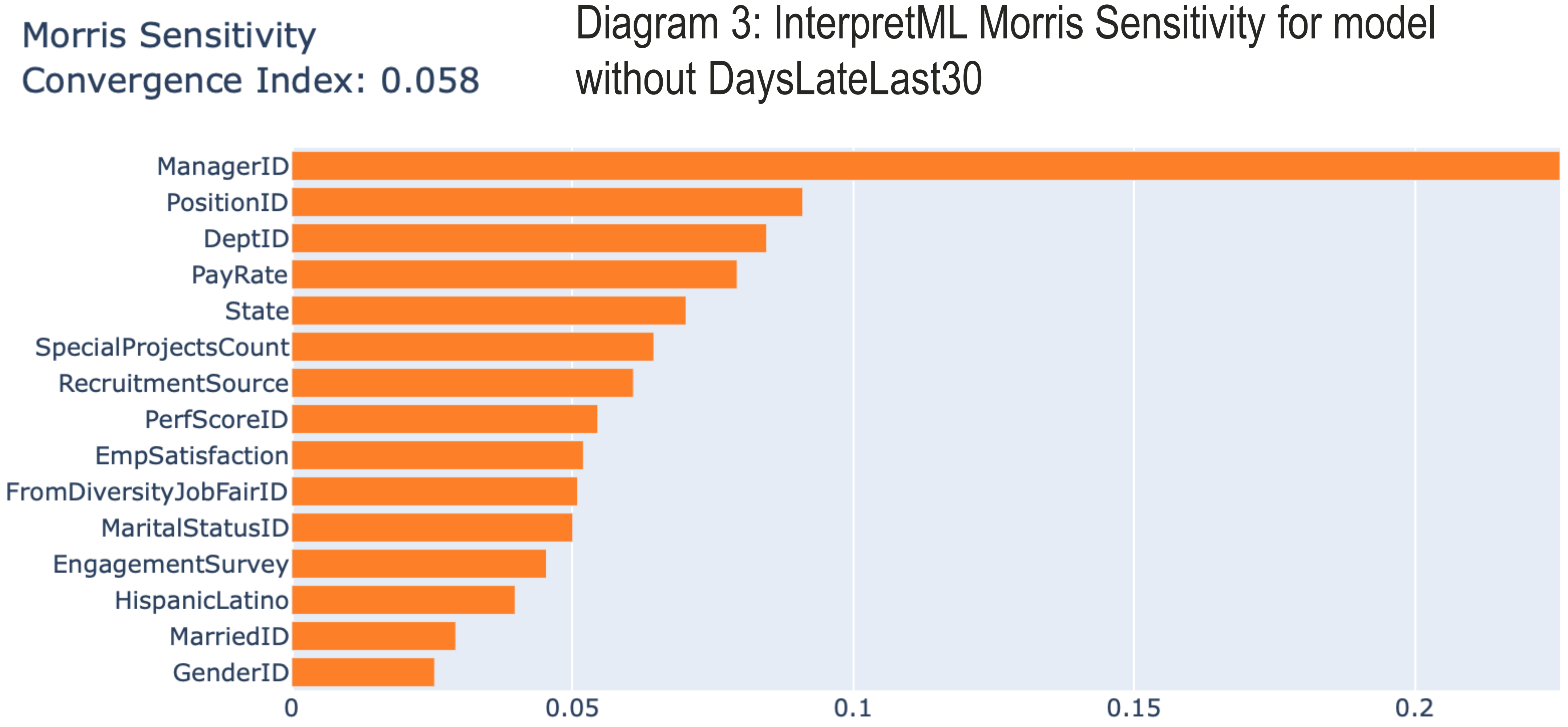


Diagram 3: InterpretML Morris Sensitivity for model without DaysLateLast30

How does the model make **correct** predictions?

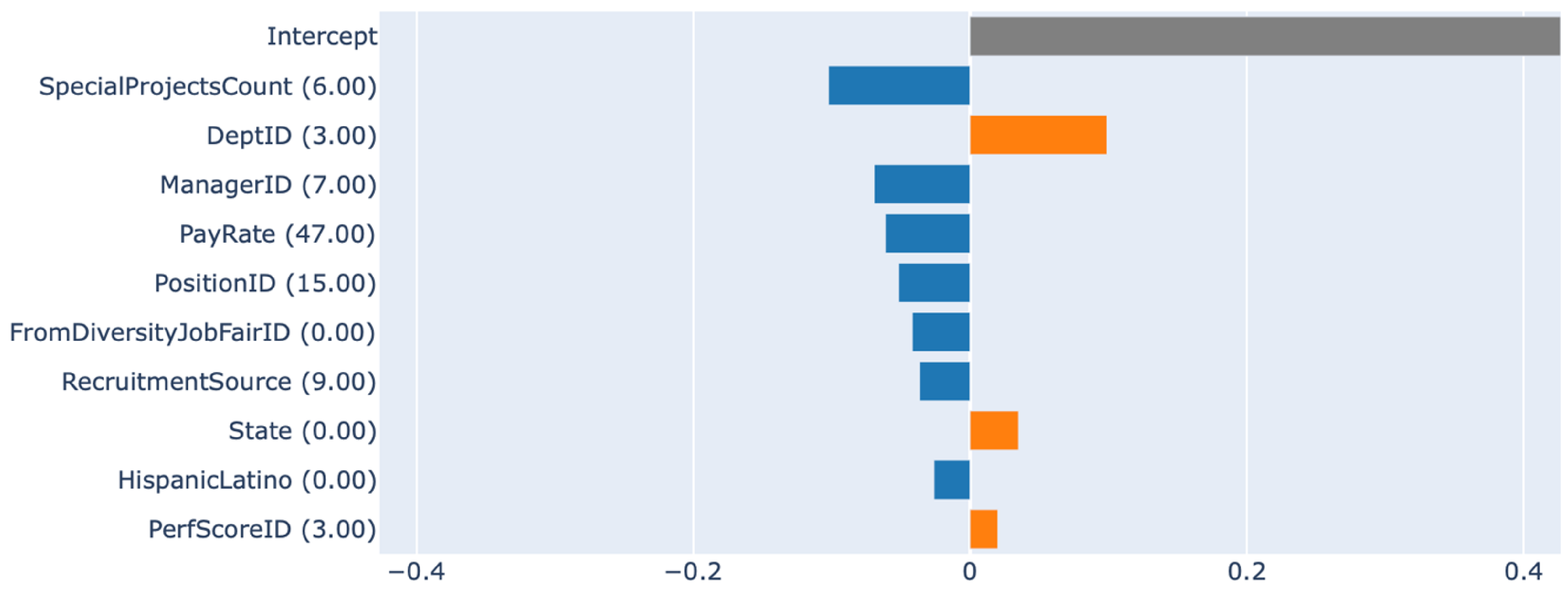


Diagram 4: InterpretML tabular Lime Explainer for object [3]

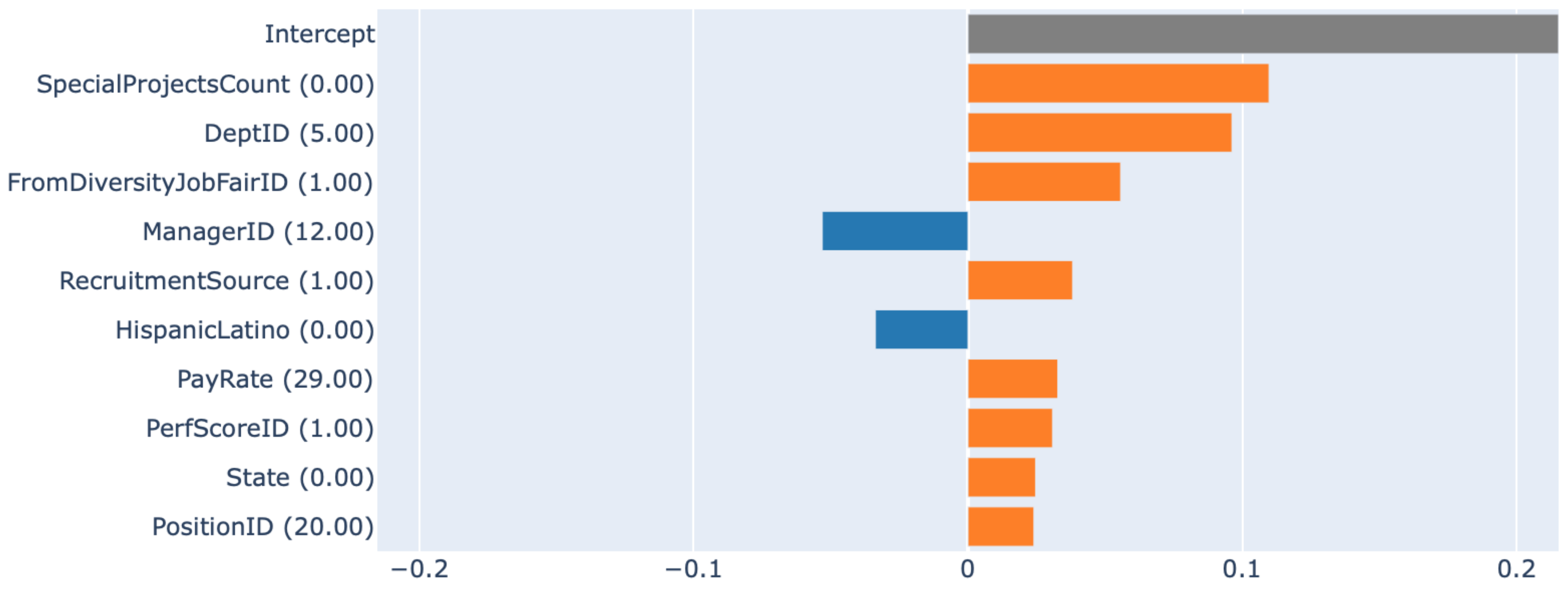


Diagram 5: InterpretML tabular Lime Explainer for object [18]

How to read Lime tabular Diagrams of InterpretML:
Predicted → probability for model to choose class 1 (terminated)
Actual → value of the acutal class

The bars show the influence of the variables with the linear model fit to that specific case.

Interpretation of Diagram 4 (person is still **active** class 0):

- a high number of special projects has a pos. influence (staying active)
- the manager, the pay rate and being in the position of a Network Engineer also lead to staying active

Interpretation of Diagram 5 (person **terminated** class 1):

- no special projects, working in production and being sourced from diversity job fair in this case had a strong impact on termination

→ In both predictions the two variables SpecialProjectCount and DeptID appear to have high influences

When does the Model make **mistakes**?

The variable HispanicLationo seems to have a influence on predicting that someone will stay active, when in reality that person terminated.

Since that specific variable is based on race the model could be biased.

Result of excluding that variable from the model:

- Precision from 0.89 to 1.00 for class 1 (terminated) no change for the other class
- leaving the variable out had no impact on overall accuracy
- therefore the variable can be excluded from the model without a negative influence

Diagram 6: InterpretML tabular Lime Explainer for object [1]

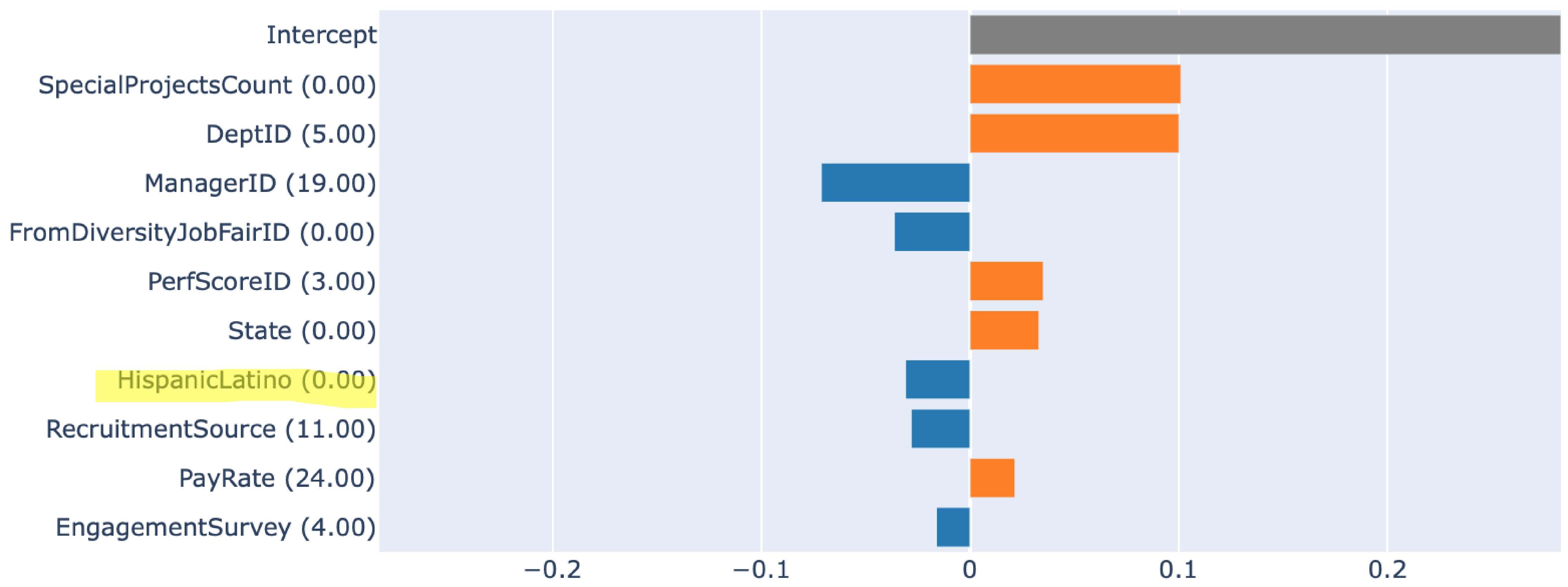
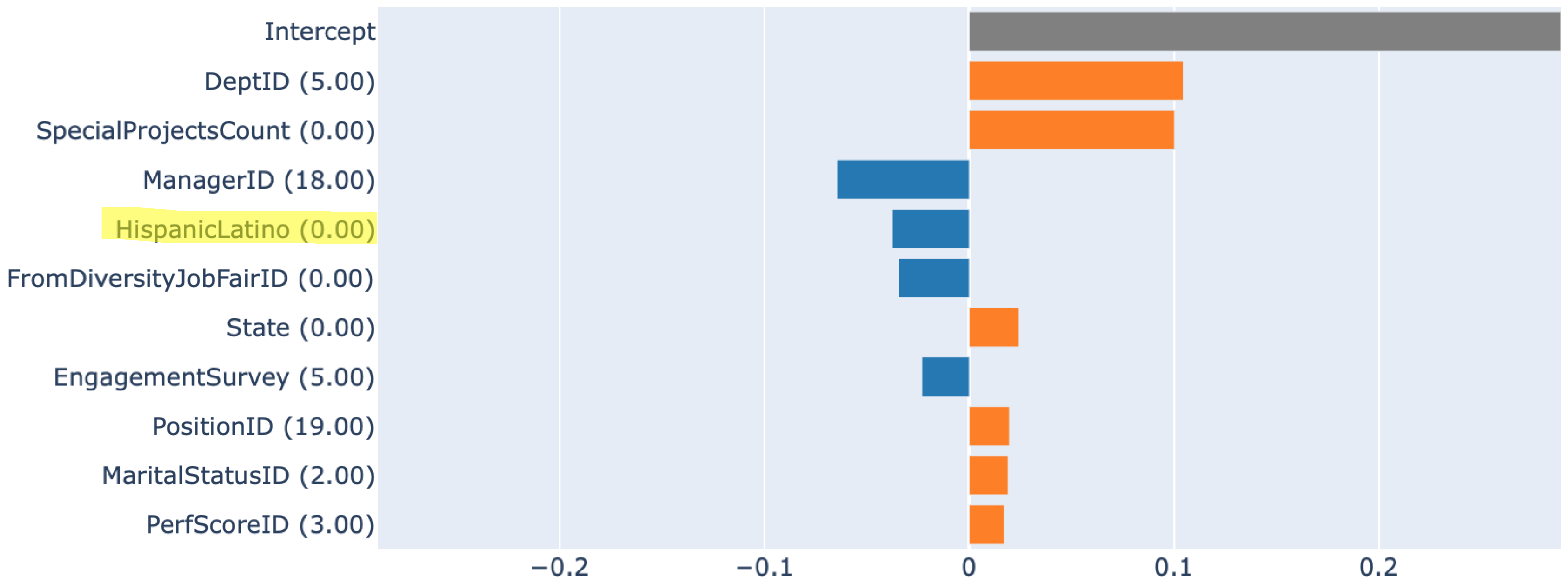


Diagram 7: InterpretML tabular Lime Explainer for object [10]



Interpret ML benefits and improvements

- gives a good overview over the specific decisions a model makes
- possibility to detect mistakes or bias in the model
- nice visualisation of results
- good interpretability for categorical variables
- easy usage also for general overview because you can use the dropdown option

- no support for float, therefore possible loss of information because we need to round variables
- usage not straight forward, notebook examples don't give explanations on how to interpret the results
- we could not include the definition of the target variables values instead the diagrams included only the integer of the class
- including a translation of categorical variables would make interpretation easier (DepID 5 = production)

Interpretability with LIME on AIX360

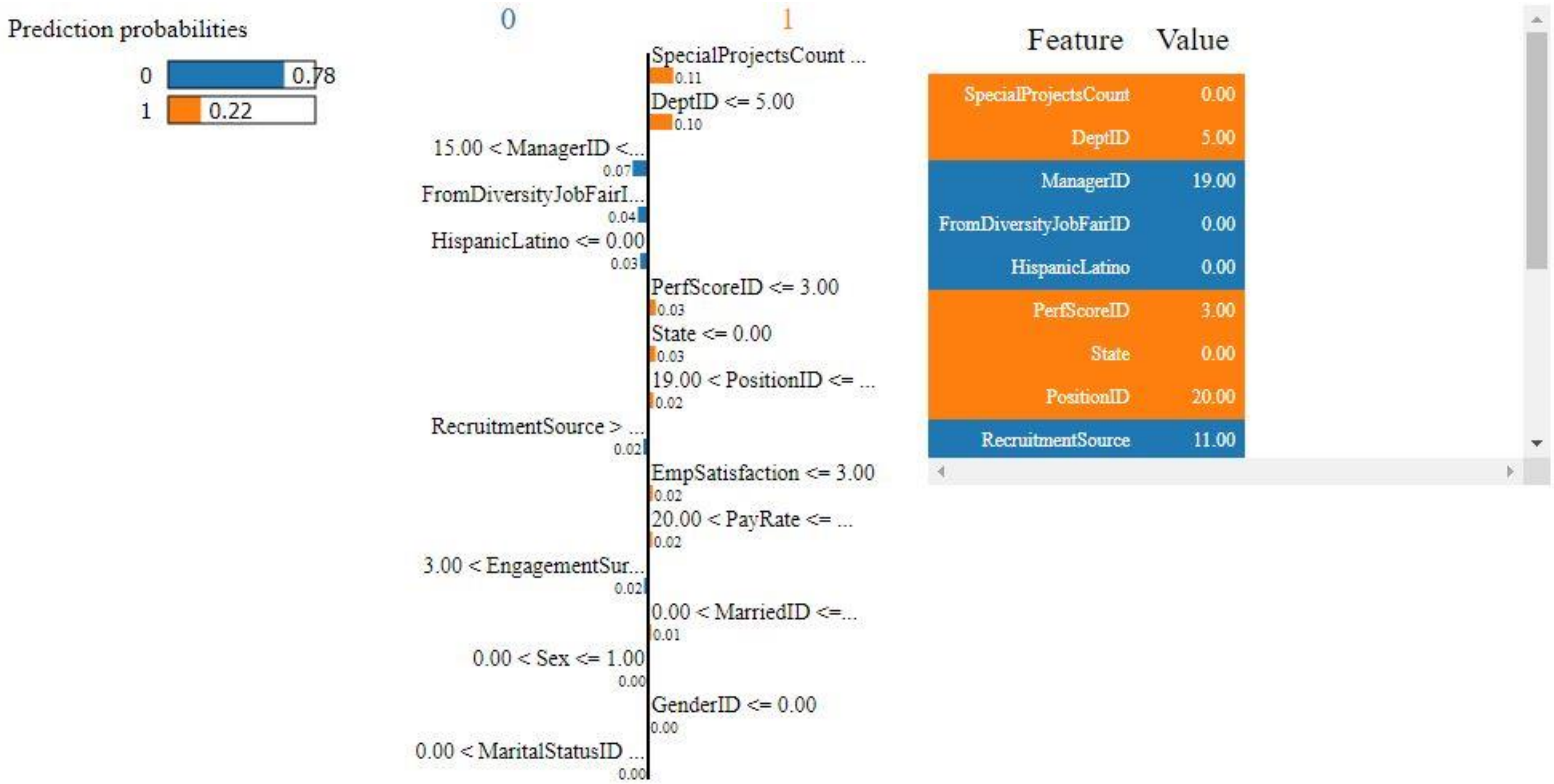
- Interpretation of AIX360 visualization for a specific tuple:

→ person still active

→ Manager and diversity have a significant impact

- Diagram has odd scaling

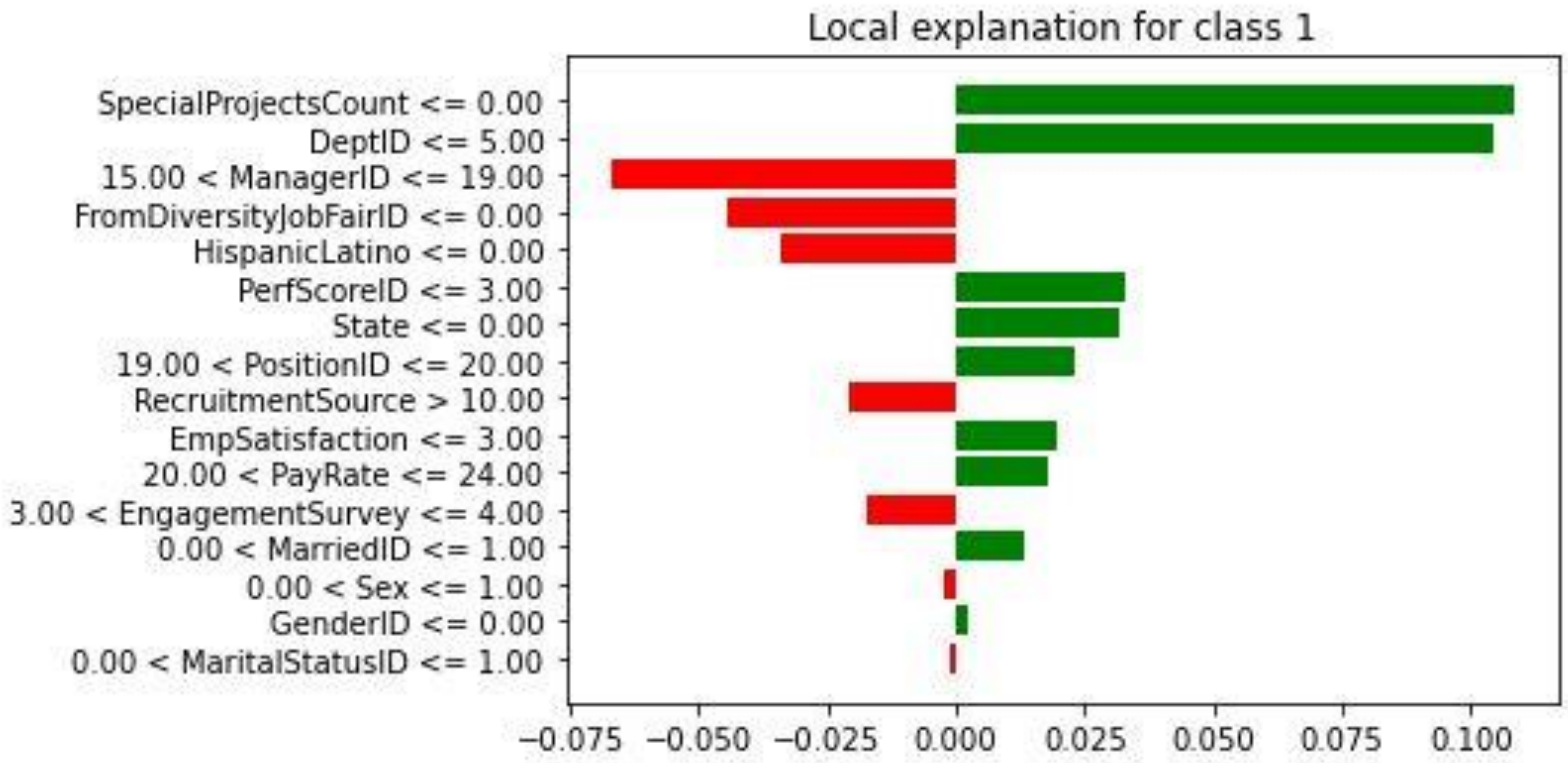
Diagram 8: AIX360 tabular Lime Explainer for object [1]



Interpretability with LIME on AIX360

- LIME includes a visualization using pyplot
- More intuitive chart and better scaling than AIX360
- Yet, not as good as the possibilities InterpretML offers

Diagram 9: AIX360 tabular Lime Explainer for object with pyplot [1]



Interpret ML vs. AIX360

- ✓ modern website and large GitHub
- ✓ easier coding for equivalent results
- ✓ many and nice visualization options
- ✓ can also be used to compare different models
- ✗ notebooks examples not well explained
- ✗ float variables need to be rounded

- ✓ powerful toolbox, but hard to navigate and understand
- ✓ supports images, text and tabular for explanation
- ✓ works with ranges of the variables for the local model
- ✗ few examples in official GitHub
- ✗ documentation is spare
- ✗ only one way for visualization

List of Sources

All icons used in the presentation are assesed via https://www.flaticon.com

AIX360. (2020). Docs. URL: https://aix360.readthedocs.io/en/latest/lbbe.html#lime-explainers (last access 10.06.2020)

Allen, D. G. (2008). *Retaining talent: A guide to analyzing and managing employee turnover*.

Github. (2020). AIX360. URL: https://github.com/IBM/AIX360 last access on 10.06.2020.

Github. (2020). InterpretML. URL: https://github.com/interpretml/interpret last access on 10.06.2020.

Kaggle. (2020). Human Resources Data Set. URL: https://www.kaggle.com/rhuebner/human-resources-data-set last access on 10.06.2020.

RPubs. (2020). HR Codebook-13. URL: https://rpubs.com/rhuebner/HRCodebook-13 last access on 10.06.2020.

SHRM Foundations. Bliss, W. G. (2011). The Advisor-Cost of Employee Turnover.

Towards Data Science. (2020). An Implementation and Explanation of the Random Forest in Python. URL: https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76 last access on 10.06.2020.

Wikipedia. Microsoft Logo. URL: https://de.wikipedia.org/wiki/Datei:Microsoft_logo.svg last access on 10.06.2020.

Wikipedia. Kaggle Logo. URL: https://de.wikipedia.org/wiki/Kaggle#/media/Datei:Kaggle_logo.png last access on 10.06.2020.