H1B VISA APPROVAL STATUS SHORT VISUAL

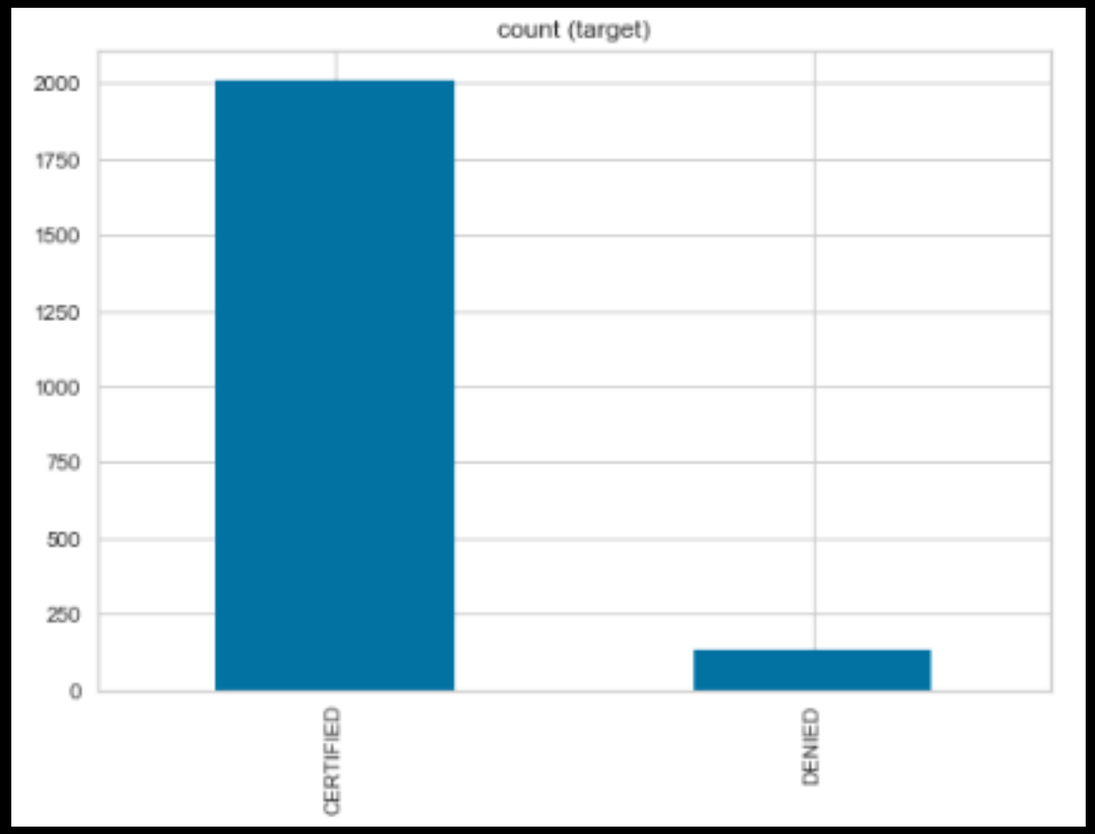
PATRICIA ATTAH

BACKGROUND

Background of this study comes from the need to have efficient staff, with the required skill set to complete projects and tasks on time and with high quality.

This survey is done on the company where Deloitte Consulting LLP had 103 denied out of 2009 this is approximately 7% of the applicants were denied. Every applicant that is denied causes project completion delay and substandard work.

In order to reduce the number of Denied cases we can look into the factors that affect the denied and certified outcomes individually. And the features that were most in influential in the model as a whole.



	EMPLOYER_NAME	counts
0	DELOITTE CONSULTING LLP	2140

	CASE_STATUS	counts
0	CERTIFIED	2009
1	DENIED	131

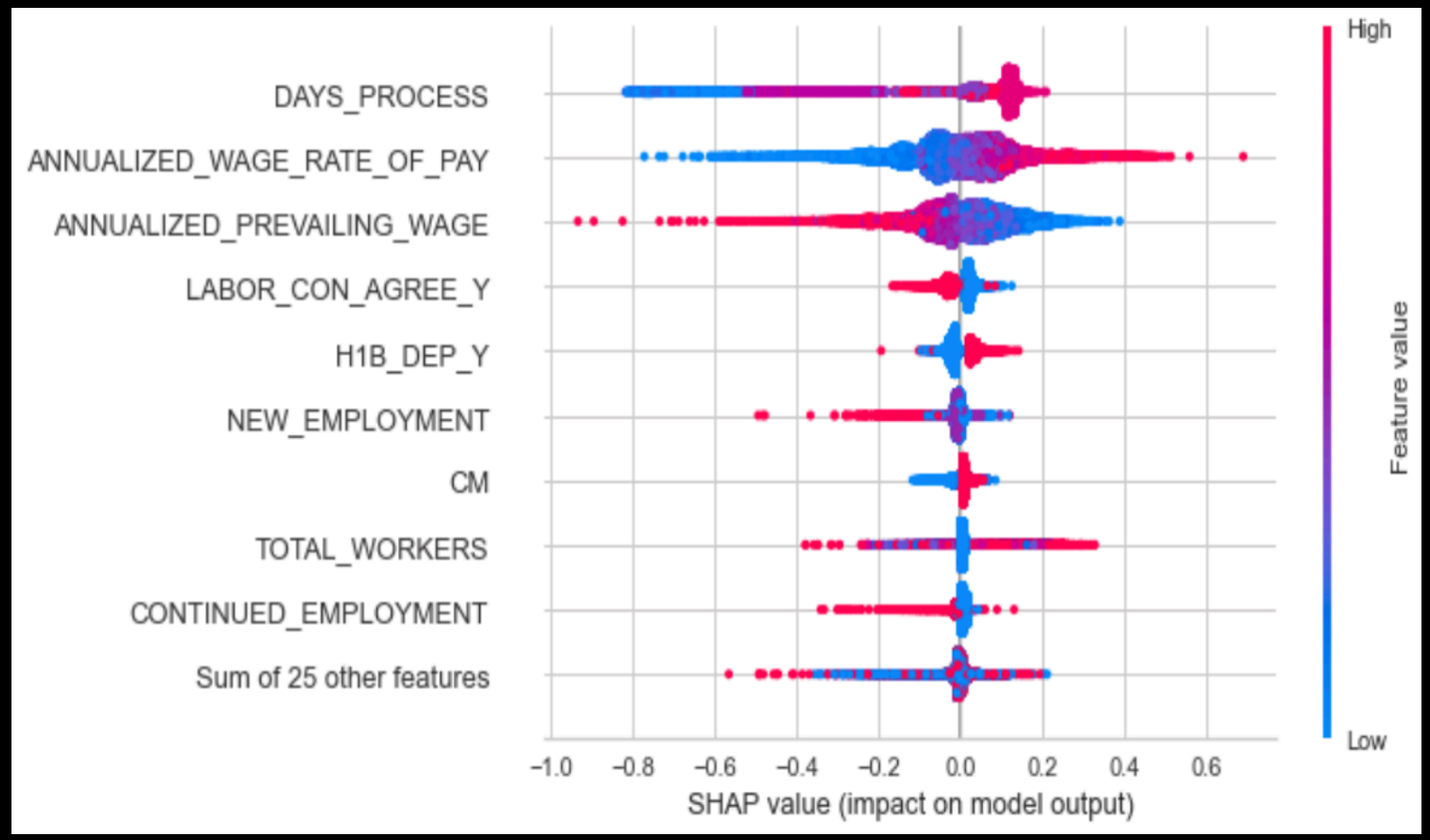
Note: Stratified k-fold and limited over sampling techniques were employed because of unbalanced targets values

Models implemented

1. Random forest with SHAP Visualization
2. K means with LIME Visualization

Random forest algorithm with SHAP visualization

The feature importance ranks in descending order so we see that DAY_PROCESS, ANNUALIZED_WAGE_RATE_PAY and the ANNUALIZED_PREVAILING_WAGE are the 3 most influential variables with. The red shows higher values while the blues shows the lower values. The more days it takes to process the visa means its more likely, while the lower the values of days process it strongly that the causes the visa to be denied. Also we can see the annual wage of pay specifies that the higher the pay of the job the more likely the visa will be approved while if the pay is lower it may not be approved. For the LABOR_CON_AGREEMENT was signed there was a lower chance of the visa to be certified.



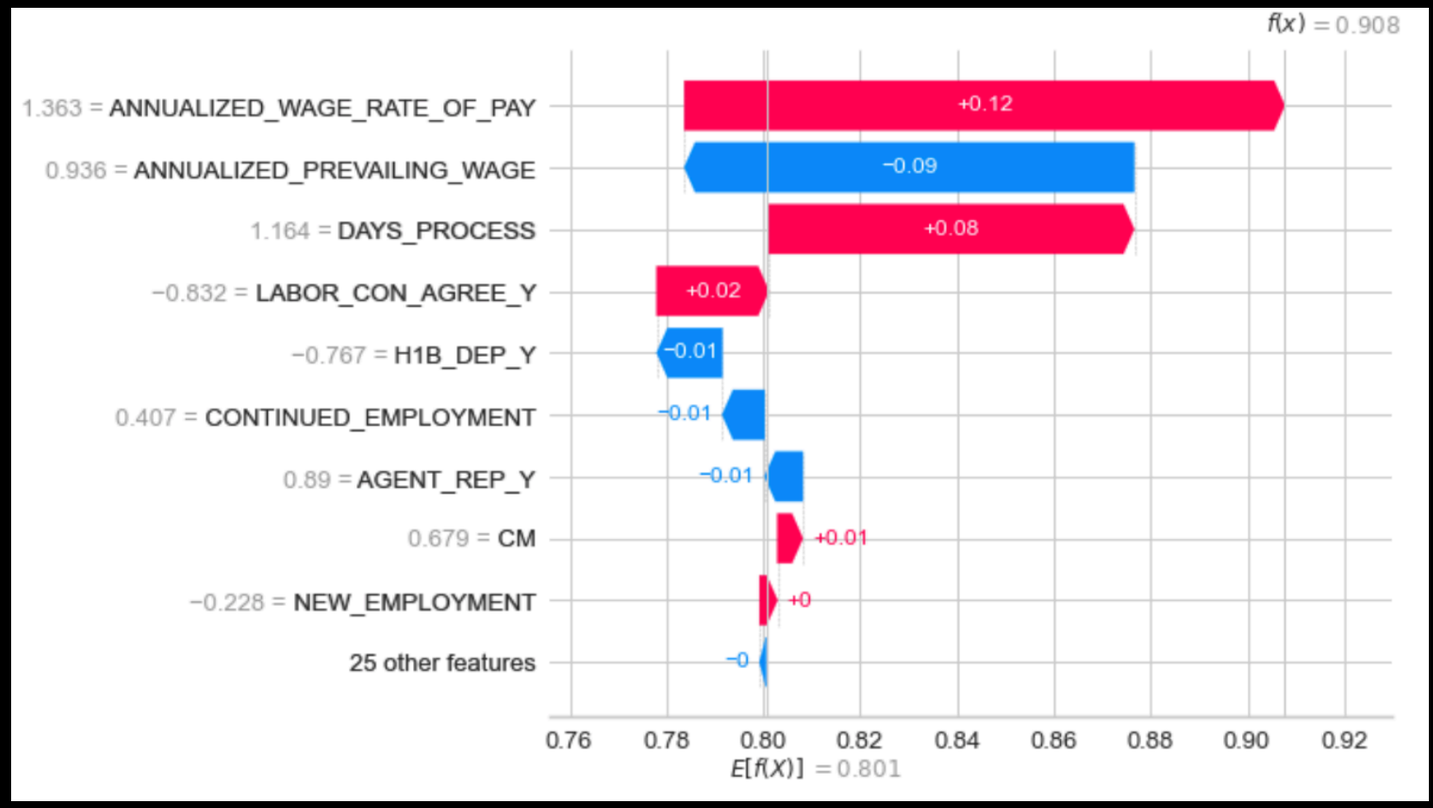
Variable Importance

Certified prediction (2nd row) - Shap plot

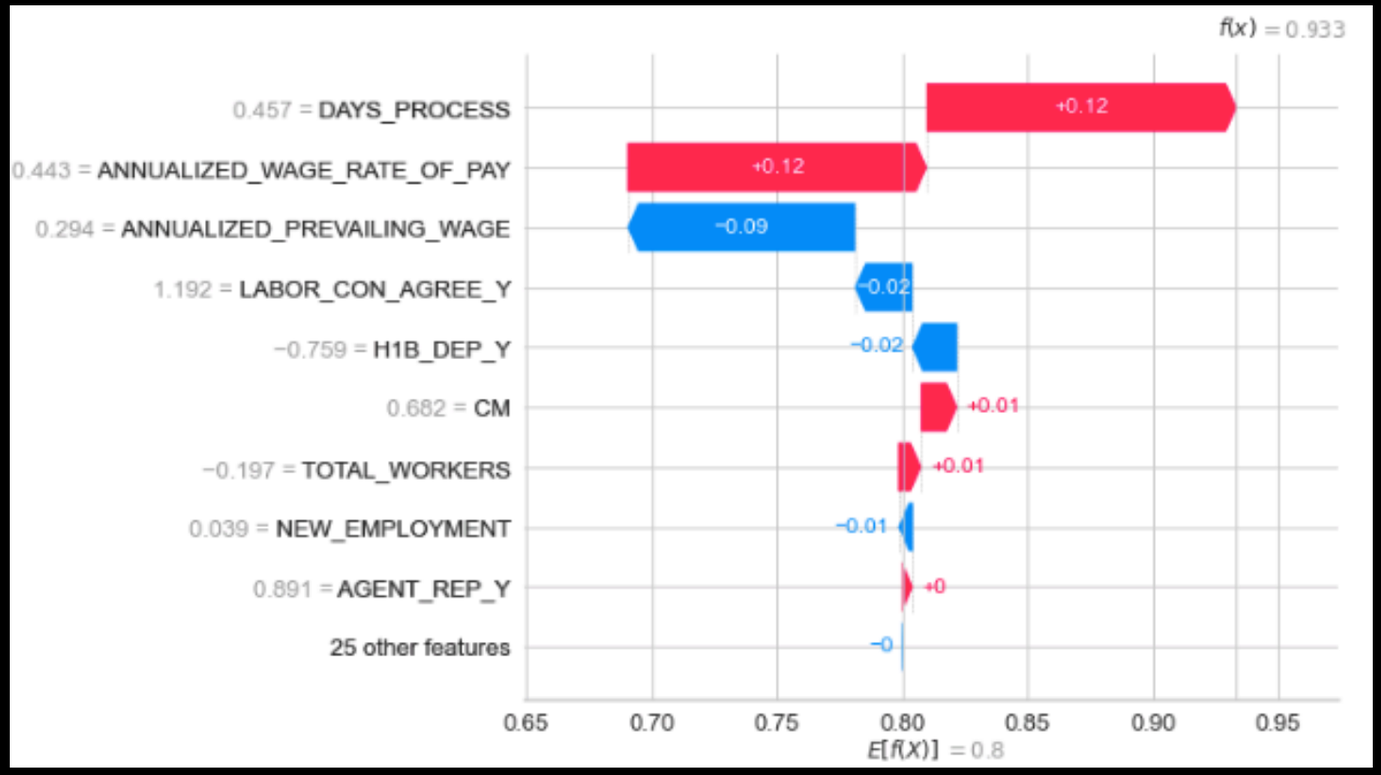
Denied prediction (7th row) - Shap plot

(Prediction of second row) in this instance the candidate was denied. The output was 0.91. The red values push the prediction higher while the blue values push them lower. The ANNUALIZED_WAGE_RATE_PAY pushes the prediction to certified because it has a positive effect in that instance and the value is higher than the mean (-.003<1.36) so it is pushes the prediction to right (which is certified). While the ANNUALIZED_PREVAILING_WAGE has a negative impact on the prediction while the value is higher than the mean(-.002<.936). So this gives an overall negative impact meaning it causes it to predict denied (and pushes it to the left).

This is another depiction of the visualization in slide 4 (2nd row certified prediction) and the second plot is the 6th row denied prediction where we can see the positive effect of ANNUALIZED_WAGE_RATE_PAY and the negative effect of ANNUALIZED_PREVAILING_WAGE. That was portrayed in the last slide. We see the determining features are both the same however their order of strength is not. Days processed is more important in the denied prediction than the certified and labor contract agreement has different impact directions on the denied and certified does not have the same

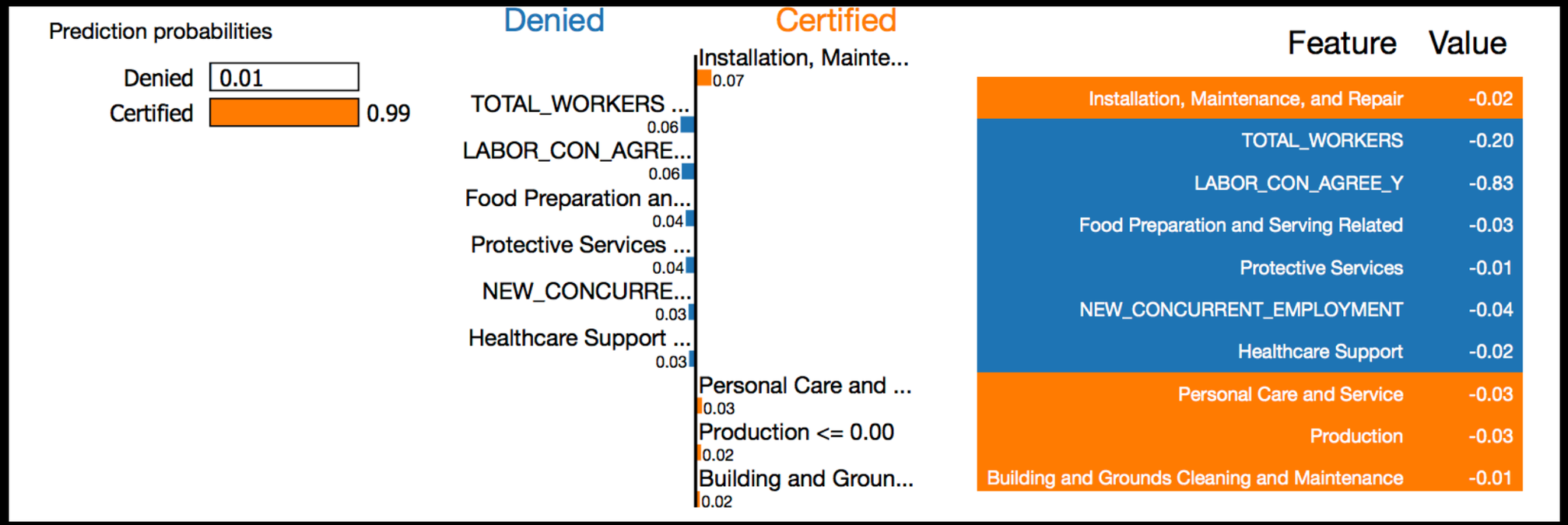


Certified prediction (2nd row) - Waterfall plot

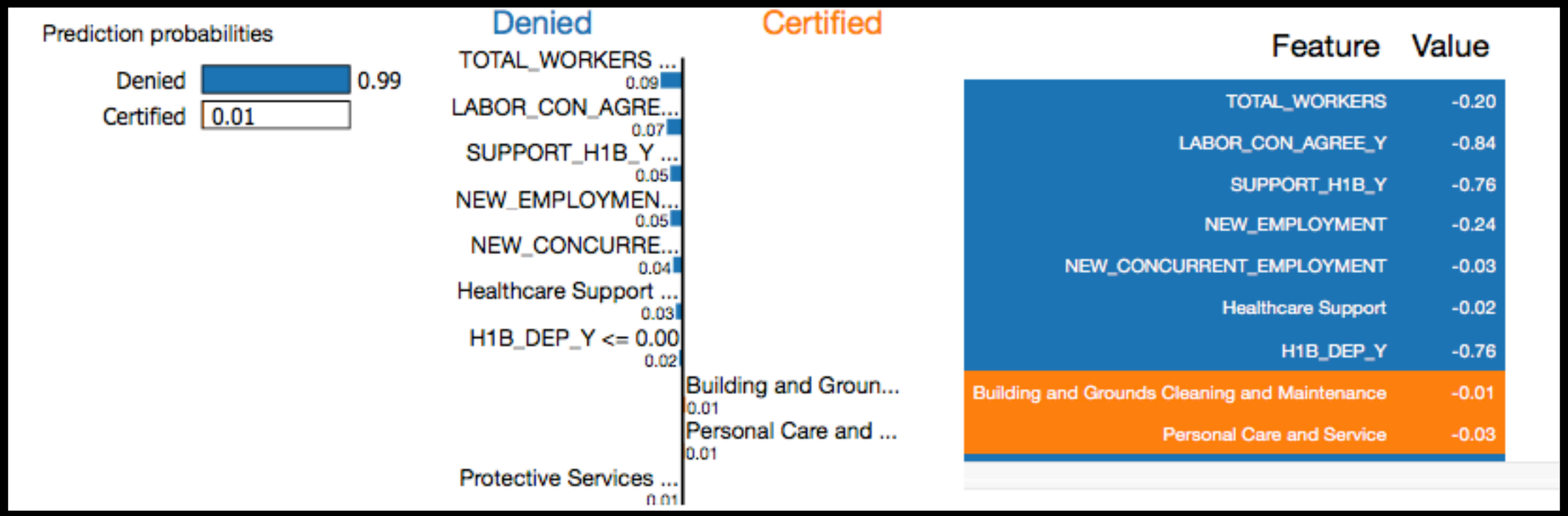


Denied prediction (7th row) - Waterfall plot

K means with LIME visualization



Certified prediction (2nd row) - LIME



Denied prediction (7th row) - LIME

Again we are looking at the 2nd row certified prediction and the 7th row denied prediction using a different software called LIME and different algorithm kmeans we see that the predictions were correct. The results show in both predictions that increase the influence to the certified prediction depends on Installation_maintenance_repair, personal care variables while the that total worker, labor con agreement and new concurrent employment increase the influence to the denied prediction