



HR ANALYTICS IN PYTHON: PREDICTING EMPLOYEE CHURN

Tuning employee turnover classifier

Hrant Davtyan

Assistant Professor of Data Science
American University of Armenia



Overfitting

Existence of overfitting:

- Training accuracy: 100%
- Testing accuracy: 97.23%

Methods to fight it:

- Limiting tree maximum depth
- Limiting minimum sample size in leafs



Pruning the tree

Limiting Depth

```
model_depth_5 = DecisionTreeClassifier(  
    max_depth=5, random_state=42)  
  
# Train set Accuracy: 97.71%  
# Test set Accuracy: 97.06%
```

Limiting Samples

```
model_sample_100 = DecisionTreeClassifier(  
    min_samples_leaf=100, random_state=42)  
  
# Train set Accuracy: 96.58%  
# Test set Accuracy: 96.13%
```



HR ANALYTICS IN PYTHON: PREDICTING EMPLOYEE CHURN

Let's practice!



HR ANALYTICS IN PYTHON: PREDICTING EMPLOYEE CHURN

Evaluating the model

Hrant Davtyan

Assistant Professor of Data Science
American University of Armenia



Prediction errors

Confusion Matrix		Reality	
		0	1
		0	1
Predicted	0	TN	FN
	1	FP	TP



Evaluation metrics 1

- If target is leavers, focus on FN
 - Recall score = $TP/(TP+FN)$
 - Lower FN, higher Recall score
 - Recall score - % of correct predictions among 1s (leavers)
- If target is stayers, focus on FP
 - Specificity = $TN/(TN+FP)$
 - Lower FP, higher Specificity,
 - Specificity - % of correct predictions among 0s (stayers)



Evaluation metrics 2

- Even if target is leavers, you may still focus on FP:
 - Precision score = $TP / (TP + FP)$
 - Lower FP, higher Recall score
 - Precision score - % of leavers in reality, among those predicted to leave



HR ANALYTICS IN PYTHON: PREDICTING EMPLOYEE CHURN

Let's practice!



HR ANALYTICS IN PYTHON: PREDICTING EMPLOYEE CHURN

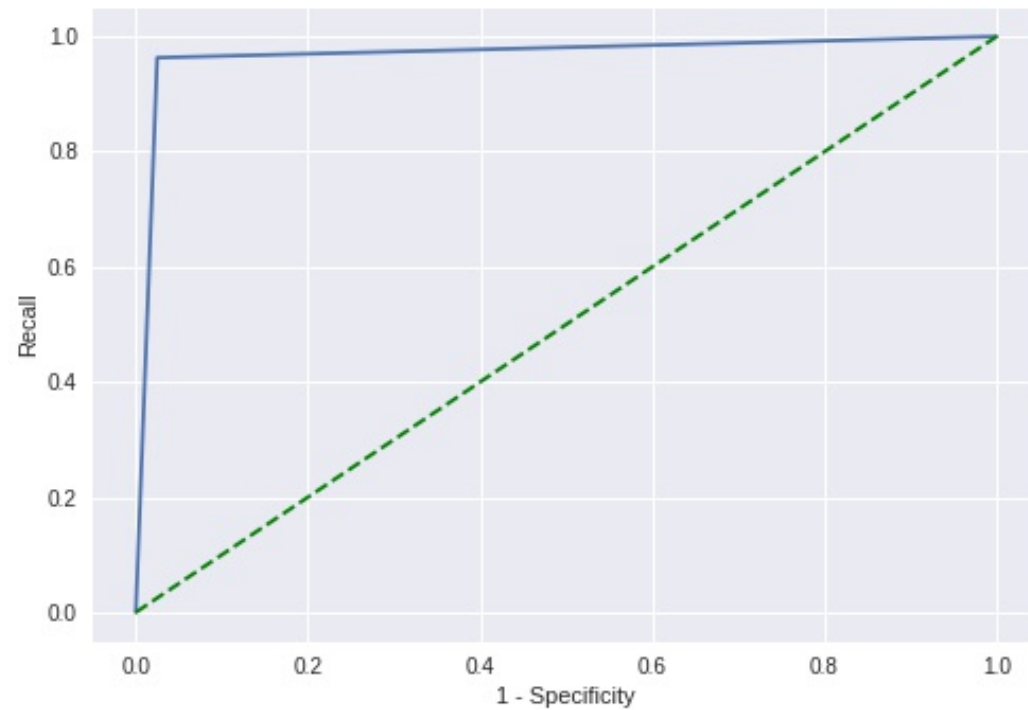
Targeting both leavers and stayers

Hrant Davtyan

Assistant Professor of Data Science
American University of Armenia



AUC score



- Vertical axis: Recall
- Horizontal axis: 1 - Specificity
- Blue line: ROC
- Green line: baseline
- Area between blue and green:
AUC



HR ANALYTICS IN PYTHON: PREDICTING EMPLOYEE CHURN

Let's practice!



HR ANALYTICS IN PYTHON: PREDICTING EMPLOYEE CHURN

Class Imbalance

Hrant Davtyan

Assistant Professor of Data Science
American University of Armenia



Prior probabilities

Without balance

- $P_0 = 0.76$
- $P_1 = 0.24$
- $Gini = 0.36$

With balance

- $P_0 = 0.5$
- $P_1 = 0.5$
- $Gini = 0.5$



HR ANALYTICS IN PYTHON: PREDICTING EMPLOYEE CHURN

Let's practice!