HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN PYTHON

Splitting the data

Hrant Davtyan
Assistant Professor of Data Science
American University of Armenia

Target and features

- target = churn
- features = everything else

Train/test split

- train - the component used to develop the model
- test - the component used to validate the model

```
from sklearn.model_selection import train_test_split

target_train, target_test, features_train, features_test =
                           train_test_split(target,features,test_size=0.25)
```

Overfitting

an error that occurs when model works well enough for the dataset it was developed on (train) but is not useful outside of it (test)

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN PYTHON

Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN PYTHON

Introduction to Decision Tree classification

Hrant Davtyan
Assistant Professor of Data Science
American University of Armenia

Classification in Python

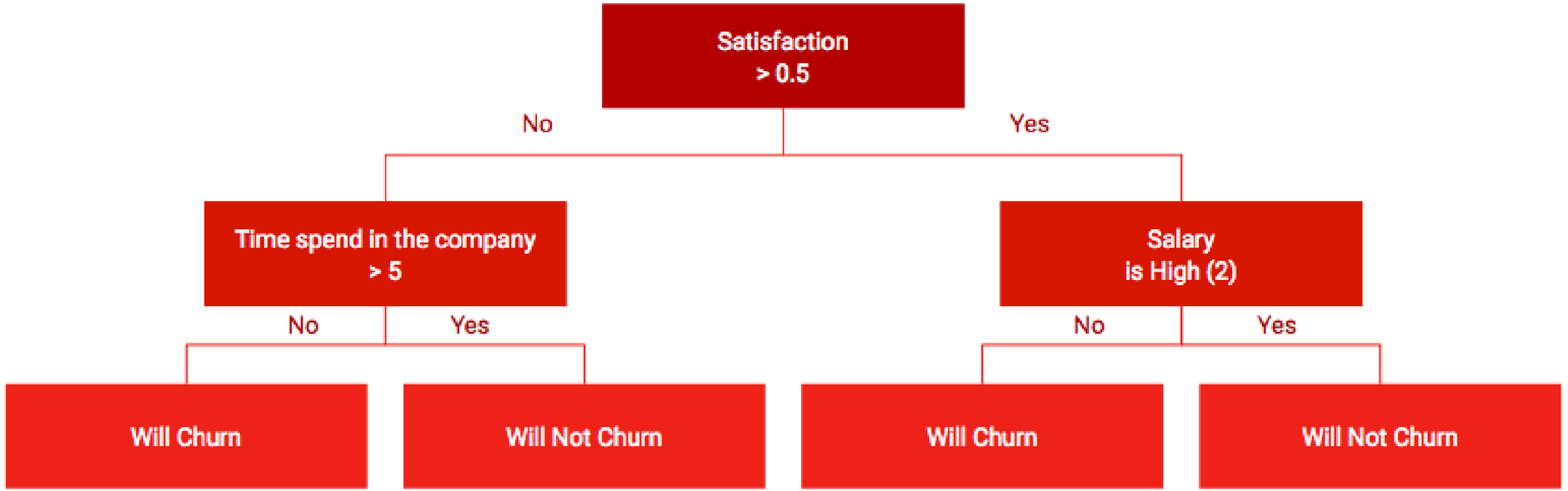
Classification algorithms

- Logistic regression
- Support Vector Machines
- Neural Networks
- Other algorithms

Algorithm we will use

- Decision Tree

Decision Tree Classification



Splitting rule

Splitting rules:

- Gini: `2*p*(1-p)`
- Entropy: `-p*log(p) - (1-p)*log(1-p)`

Decision Tree splitting: hypothetical example

Total set: 100 observations, 40 left, 60 stayed

- Gini: `2*0.4*0.6 = 0.48`

Splitting rule: satisfaction > 0.8

- Left branch (YES) - 50 people: all stayed
- Gini: `2*1*0 = 0`
- Right branch (NO) - 50 people: 40 left, 10 stayed
- Gini: `2*0.4*0.1 = 0.08`

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN PYTHON

Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN PYTHON

Predicting employee churn using decision trees

Hrant Davtyan
Assistant Professor of Data Science
American University of Armenia

Decision Tree in Python

```
from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier(random_state=42)

model.fit(features_train,target_train)

model.score(features_test,target_test)*100
```

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN PYTHON

Let's practice!

HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN PYTHON

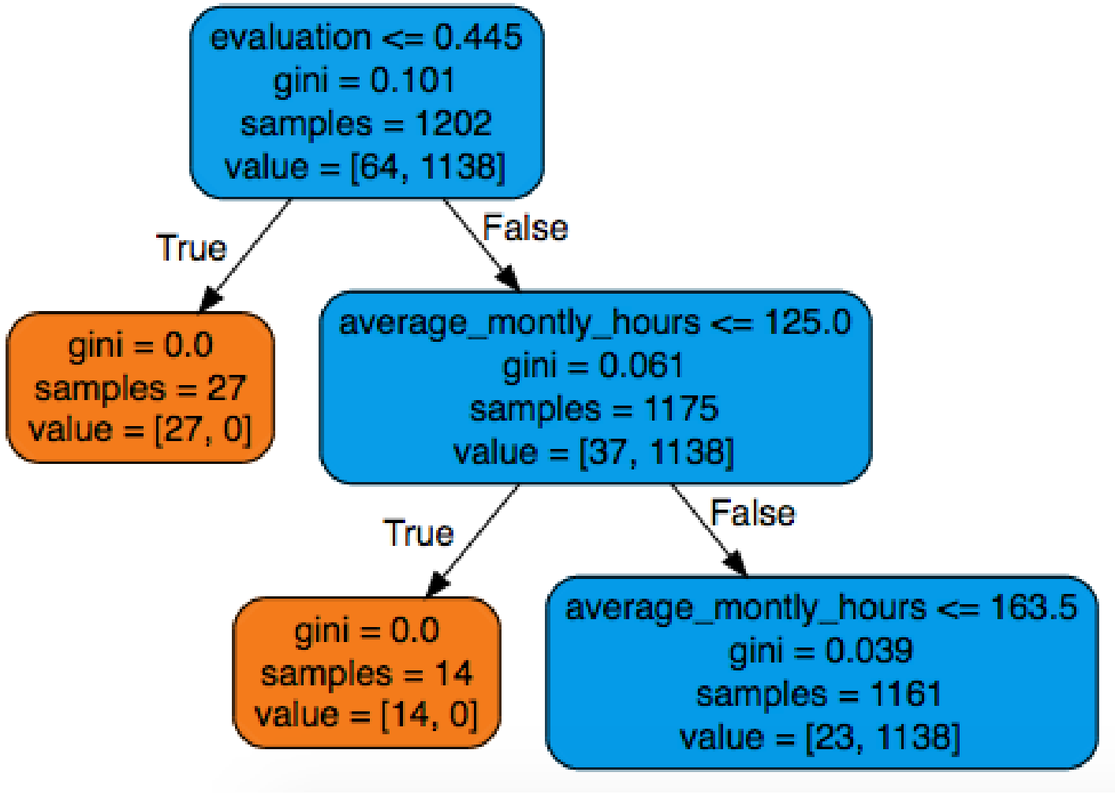
Interpretation of the decision tree

Hrant Davtyan
Assistant Professor of Data Science
American University of Armenia

Visualization

1. Export
2. Copy content
3. Paste it in www.webgraphviz.com

Interpretation



HUMAN RESOURCES ANALYTICS: PREDICTING EMPLOYEE CHURN IN PYTHON

Let's practice!