Professional Certificate in Machine Learning and Artificial Intelligence

Module 20: Transparency and Interpretability

Building an Interpretable Decision Tree

Imagine a loans organisation has created the decision tree described below based upon data on its borrowers. They now want to create an interpretable version of that decision tree so that any employee from the organisation can easily understand why the tree made a particular decision.

For this activity, recreate the following decision tree to make it interpretable to employees in the organisation.

* Root node
  + Field\_HUMANITIES <= 0.5
  + gini = 0.352
  + samples = 500
  + value = [386, 114]  
    *The root node branches into decision two nodes detailed below.*
    - Decision node 1
      * Field\_STEM <= 0.5
      * gini = 0.136
      * samples = 353
      * value = [327, 26]  
        *Decision node 1 branches into two terminal nodes detailed below.*
        + Terminal node 1

gini = 0.339

samples = 97

value = [76, 21]

class = No (22% default)

* + - * + Terminal node 2

gini = 0.338

samples = 256

value = [251, 5]

class = No (1.9% default)

* + - Decision node 2
      * graduationYear <= 2009.5
      * gini = 0.481
      * samples = 147
      * value = [59, 88]
      * class = Yes  
        *Decision node 2 branches into two terminal nodes detailed below.*
        + Terminal node 1

gini = 0.436

samples = 78

value = [53, 25]

class = No (32% chance of default)

* + - * + Terminal node 2

gini = 0.159

samples = 69

value = [6, 63]

class = Yes (91% chance of default)

In trying to make this decision tree interpretable, keep in mind the following considerations:

* Think about how the predictions at the leaves should be reported.
  + Show class or final decision using majority vote. If yes samples are more than no samples, report yes and vice versa. Can also be reported as percentages or probability of defaulting
* Think about how a person might interpret the nodes
  + Use feature names to depict the nodes. Samples can be depicted as proportions.
* Make sure you include information about the dataset.
  + It’s a loan dataset, features are: (['field', 'graduationYear', 'loanAmount', 'selectiveCollege', 'sex']
* Make sure you include information about the model and what it was created for.
  + Model is a decision tree classifier using a max depth of 2 and accuracy of 87.2 to classify if a person defaulted on their loan or not.
* What issues do you think could arise from how this model might be used?
  + This model uses very features to determine if a person defaulted on a loan or not and uses questionable features such as field and graduation year. It might need more research to see if it can be trusted.

In addition to these considerations, answer the following questions about your decision tree:

* Where is the dataset from, and which features does it include?

It’s a loan dataset, features are: (['field', 'graduationYear', 'loanAmount', 'selectiveCollege', 'sex']

* field: the field in which each student is taking their studies in
* graduationYear: the year in which each student graduated
* loanAmount: the amount each student owns
* selectiveCollege: binary valued column: 1 for students who attend a selective college, 0 for students that do not
* sex: sex of the student
* Is it fair to assess loans based on this decision tree? For example, is it fair to discriminate based on subject?

It might not be fair but this is what the decision tree is doing based on the data it is trained on.

Example of an interpretable decision tree:

