3 DAYS, 6 SESSIONS

UNIVERSAL AI COURSE

July 3 - 5, 2023 | NCSR "Demokritos" Congress Centre

Universal AI and its applications in science, engineering, humanities, law, medicine and art.

AI in practice - Part 2

Decision trees and interpretable machine learning

HELLENIC INSTITUTE of ADVANCED STUDIES

Machine learning engineering

Problem understanding

Data wrangling

Data analysis

Model training

Evaluation setting

Analysis of the results

Train yourself before training neural networks

Take your data seriously

ML systems development

Evaluate real-world performance

Model deployment

Choose the right model

Set baselines and targets

Understand the mechanisms behind model decisions

Problem: Supreme Court decision prediction

Context
The Supreme Court of the United States (SCOTUS) consists of nine justices, appointed by the President. Justices are distinguished judges, professors of law, and state and federal attorneys deciding on the most difficult and controversial cases that often involve the interpretation of the Constitution. The decisions made by the Supreme Court can have significant social, political, and economic consequences.

Motivation
Legal academics and political scientists try to predict SCOTUS decisions through in-depth studies of cases and individual justices.

Task
Use an AI model constructed from data to make predictions about decisions. It is highly desirable that the model allows for a more intuitive understanding of the decision-making process. The AI model should provides a clear and interpretable representation of the factors and criteria used in reaching a particular outcome.

Problem: Supreme Court decision prediction

Dataset
During the period from 1994 to 2001, the same nine justices served on the SCOTUS. These justices were Breyer, Ginsburg, Kennedy, O'Connor, Rehnquist (Chief Justice), Scalia, Souter, Stevens, and Thomas. This time frame represents a rare and valuable dataset as it encompasses the longest period with the same set of justices in over 180 years.
For the purpose of prediction, the focus will be on **Justice Stevens' decisions**.

Dataset characteristics
It includes over 566 records and 8 attributes (we consider 6 attributes representing properties of the case that affect the decision).
The output is whether Justice Stevens voted to *reverse the lower court decision*. A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

Data exploration

Preliminary and more complex tasks

Load the data

Use Pandas to load the CSV file
Perform basic operations on the data (access, visualise, filter etc)
Data handling (handle missing values, preprocessing etc)

View the data

Get a glimpse of the data
Preview simple data statistics
Visualise complex data statistics

Data exploration

Attributes

Circuit: ranging from 1st to 11th, DC, and FED
Issue: e.g., civil rights, federal taxation
Petitioner: e.g., US, an employer
Respondent: e.g. US, an employer
LowerCourt: conservative or liberal
Unconst: Whether the petitioner argued that a law or practice was unconstitutional

Target variable to predict

Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

Docket	Term	Circuit	Issue	Petitioner	Respondent	LowerCourt	Unconst	Reverse
93-1408	1994	2nd	EconomicActivity	BUSINESS	BUSINESS	liberal	0	1
93-1577	1994	9th	EconomicActivity	BUSINESS	BUSINESS	liberal	0	1
93-1612	1994	5th	EconomicActivity	BUSINESS	BUSINESS	liberal	0	1
94-623	1994	1st	EconomicActivity	BUSINESS	BUSINESS	conser	0	1
94-1175	1995	7th	JudicialPower	BUSINESS	BUSINESS	conser	0	1
95-129	1995	9th	EconomicActivity	BUSINESS	BUSINESS	conser	1	0
95-728	1996	FED	EconomicActivity	BUSINESS	BUSINESS	conser	0	1
96-1768	1997	9th	EconomicActivity	BUSINESS	BUSINESS	conser	1	1
96-843	1997	DC	EconomicActivity	BUSINESS	BUSINESS	conser	0	1
98-1480	1999	11th	CriminalProcedure	BUSINESS	BUSINESS	conser	0	1
99-150	1999	2nd	EconomicActivity	BUSINESS	BUSINESS	conser	0	1
99-1571	2000	6th	EconomicActivity	BUSINESS	BUSINESS	conser	1	1

Data exploration

Preliminary and more complex tasks

Understand the data
Understand the impact of the different attributes
Understand what are the attributes that are important
Understand possible biases

Data exploration

Circuit: ranging from 1st to 11th, DC, and FED

1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	DC	FED
15	35	32	46	53	38	47	44	122	33	49	30	22

Issue: e.g., civil rights, federal taxation

Attorneys	3
CivilRights	74
CriminalProcedure	132
DueProcess	43
EconomicActivity	98
FederalismAnd...rstateRelations	28
FederalTaxation	15
FirstAmendment	39
JudicialPower	102
Privacy	9
Unions	23

Respondent: e.g., US, an employer

AMERICAN.INDIAN	4
BUSINESS	79
CITY	13
CRIMINAL.DEFENDENT	89
EMPLOYEE	30
EMPLOYER	17
GOVERNMENT.OFFICIAL	38
INJURED.PERSON	9
OTHER	175
POLITICIAN	16
STATE	48
US	48

Data exploration

Respondent: e.g., US, an employer

AMERICAN.INDIAN	13
BUSINESS	80
CITY	9
CRIMINAL.DEFENDENT	58
EMPLOYEE	28
EMPLOYER	21
GOVERNMENT.OFFICIAL	24
INJURED.PERSON	14
OTHER	177
POLITICIAN	17
STATE	56
US	69

LowerCourt: conservative or liberal

conser	293
liberal	273

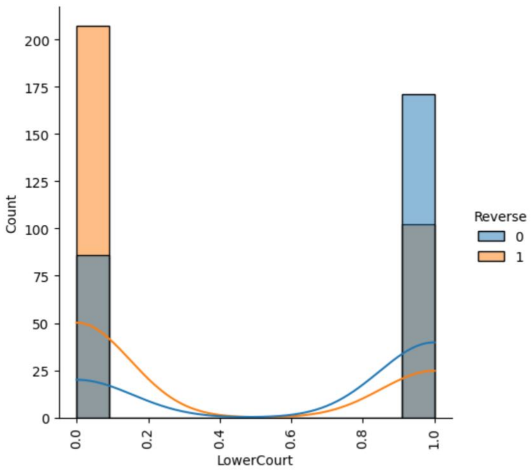
Unconst: Whether the petitioner argued that a law or practice was unconstitutional

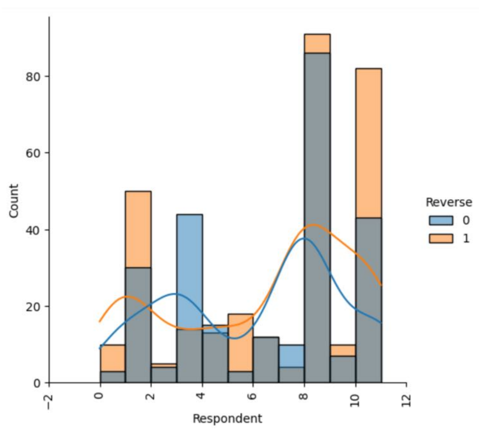
0	426
1	140

Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

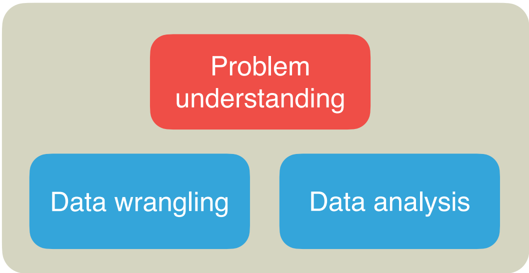
0	257
1	309

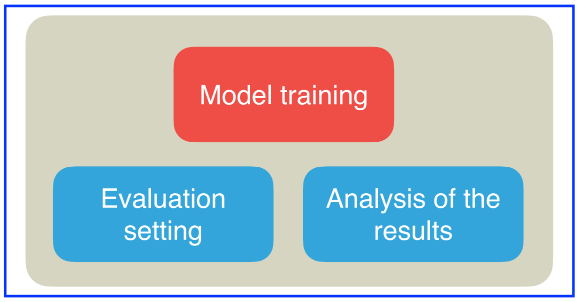
Data exploration





Machine learning engineering



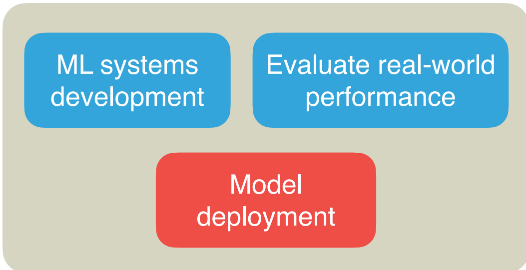


Train yourself before training neural networks

Take your data seriously

Choose the right model

Set baselines and targets



More data exploration and baseline

Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

0	257
1	309

A simple classifier that always predicts 1 has very low accuracy

$$\text{acc} = \frac{309}{257 + 309} \approx 55\,\%$$

Idea

First check the value of one of the attributes

$$\text{prob(LowerCourt=liberal)} = \frac{273}{273 + 293} \approx 0{,}48$$

$$\text{prob(LowerCourt=conser)} = \frac{293}{273 + 293} \approx 0{,}52$$

LowerCourt=liberal (273 items)

Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

0	171
1	102

$$\text{acc} = \frac{171}{171 + 102} \approx 63\,\%$$

LowerCourt=conser (293 items)

Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

0	86
1	207

$$\text{acc} = \frac{207}{207 + 86} \approx 71\,\%$$

More data exploration

After checking first an attribute, can we continue by checking another attribute?

LowerCourt=liberal (273 items)

Unconst: Whether the petitioner argued that a law or practice was unconstitutional

0	426
1	140

1 : 3,05

Unconst: Whether the petitioner argued that a law or practice was unconstitutional

0	235
1	38

1 : 6,19

Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

0	257
1	309

1 : 0,83

Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

0	171
1	102

1 : 1,68

More data exploration

After checking first an attribute, can we continue by checking another attribute?

LowerCourt=conser (293 items)

Unconst: Whether the petitioner argued that a law or practice was unconstitutional

0	426
1	140

1 : 3,05

Unconst: Whether the petitioner argued that a law or practice was unconstitutional

0	191
1	102

1 : 1,87

Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

0	257
1	309

1 : 0,83

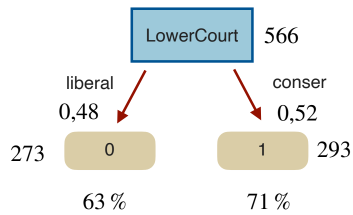
Reverse: A value of 1 indicates a vote to reverse, while a value of 0 indicates a vote to affirm.

0	86
1	207

1 : 0,42

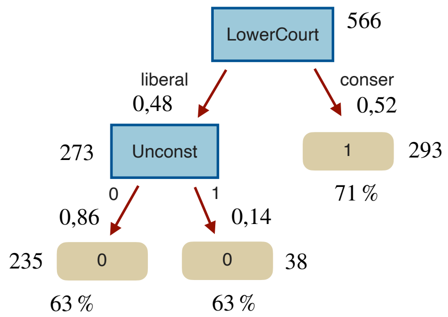
Model selection - Decision trees

Circuit	Issue	Petitioner	Respondent	LowerCourt	Unconst	Reverse
2nd	EconomicActivity	BUSINESS	BUSINESS	liberal	0	1
2nd	EconomicActivity	BUSINESS	BUSINESS	conser	0	1
2nd	CivilRights	EMPLOYEE	BUSINESS	conser	0	1
2nd	FirstAmendment	OTHER	BUSINESS	liberal	1	0
2nd	JudicialPower	STATE	CITY	liberal	0	1
2nd	JudicialPower	BUSINESS	CRIMINAL.DEFENDENT	liberal	0	0
2nd	DueProcess	OTHER	CRIMINAL.DEFENDENT	liberal	0	1
2nd	CriminalProcedure	STATE	CRIMINAL.DEFENDENT	liberal	0	1
2nd	CriminalProcedure	STATE	CRIMINAL.DEFENDENT	liberal	0	1



IF LowerCourt IS liberal THEN Reverse IS 0

IF LowerCourt IS conser THEN Reverse IS 1



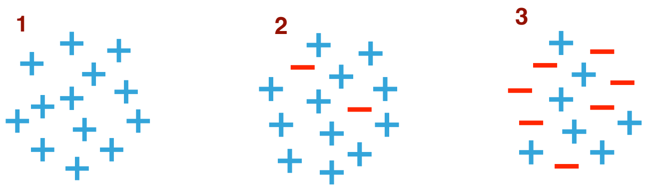
IF LowerCourt IS liberal AND Unconst IS 0 THEN Reverse IS 0

Model selection - Decision trees

DesicionTree . train($\mathbb{D}$)

1. Choose an attribute a with different values on $\mathbb{D}$
2. Insert a new node A with a as the node condition
3. For any value of a:
    - 3.1 Insert a new child a
    - 3.2 Define $\mathbb{D}'$ as the subset of $\mathbb{D}$ with the instances that have this value
    - 3.3 If all $y \in \mathbb{D}'$ have the same value, then make this node a leaf

      Else DecisionTree . train($\mathbb{D}'$)



$$\text{gini}(\mathbb{D}) = \sum_{i \in \text{labels}(\mathbb{D})} p_i(1 - p_i) = 1 - \sum_{i \in \text{labels}(\mathbb{D})} p_i^2$$

Model selection - Decision trees

Reverse

0	257
1	309

LowerCourt=liberal (273 items)

Reverse

0	171
1	102

LowerCourt=conser (293 items)

Reverse

0	86
1	207

Information Gain: $\text{ig}(\text{LowerCourt}) = \text{gini}(\text{Root}) - \text{gini}(\text{LowerCourt}) =$

$$= \text{gini}(\text{Root}) - \sum_{v \in \text{values}(\text{LowerCourt})} \frac{|\text{LowerCourt} = v|}{|\text{Root}|} \text{gini}(\text{LowerCourt} = v)$$

$$\text{gini}(\text{Root}) = 1 - \left(\frac{257}{566}\right)^2 - \left(\frac{309}{566}\right)^2$$

$$\text{gini}(\text{LowerCourt}) = \frac{273}{566}\text{gini}(\text{LowerCourt} = \text{liberal}) + \frac{293}{566}\text{gini}(\text{LowerCourt} = \text{conser})$$

$$\text{gini}(\text{LowerCourt} = \text{liberal}) = 1 - \left(\frac{171}{273}\right)^2 - \left(\frac{102}{273}\right)^2$$

$$\text{gini}(\text{LowerCourt} = \text{conser}) = 1 - \left(\frac{86}{293}\right)^2 - \left(\frac{207}{293}\right)^2$$

Model training

Prepare the training and testing dataset

Split the dataset into training and testing datasets
Training Dataset has 424 rows
Testing Dataset has 142 rows

Experimentation

Set parameters: **min_samples_leaf**, **max_depth**
Explore the intuitions behind the results
Check accuracy measures together with tree parameters (tree depth etc)
Understand the interpretability of the model (check the rules extracted from the tree and their intuitive meaning)

Model training

Evaluation measures

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

Confusion matrix

True Positive (TP): a sample belonging to the positive class being classified correctly
True Negative (TN): a sample belonging to the negative class being classified correctly
False Positive (FP): a sample belonging to the negative class but being classified wrongly as belonging to the positive class
False Negative (FN): a sample belonging to the positive class but being classified wrongly as belonging to the negative class

Confusion matrix

TN	FP
FN	TP

Testing the model

The model prediction

Compute the prediction of the model for the testing data
Evaluate the results using the basic measures

Confusion matrix (testing data) (min_samples_leaf=25)

TN	FP
FN	TP

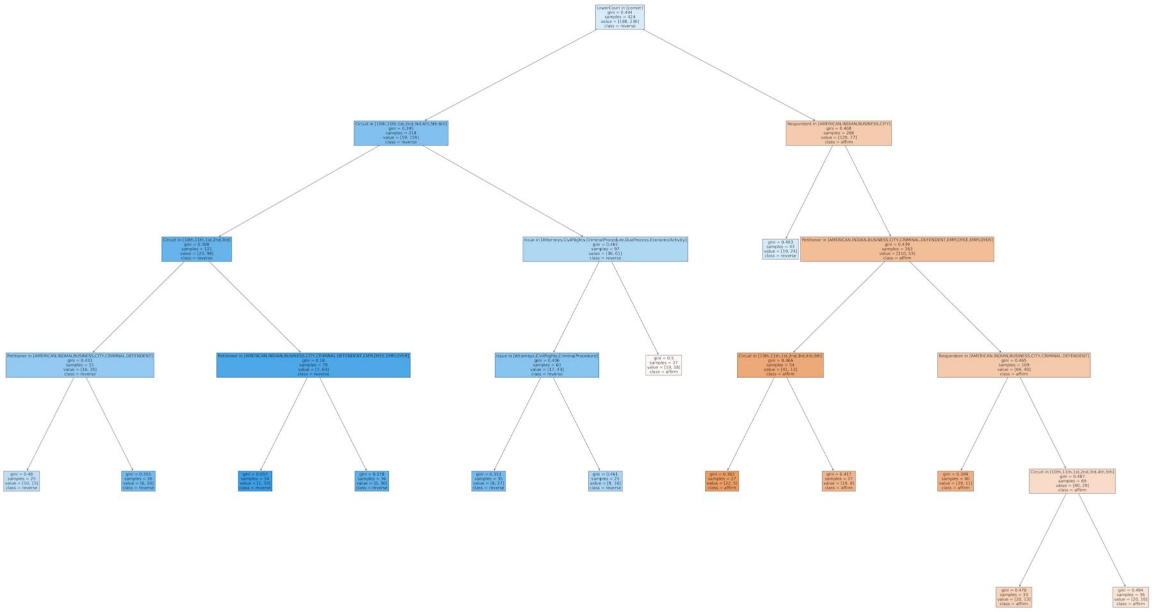
=

42	27
26	47

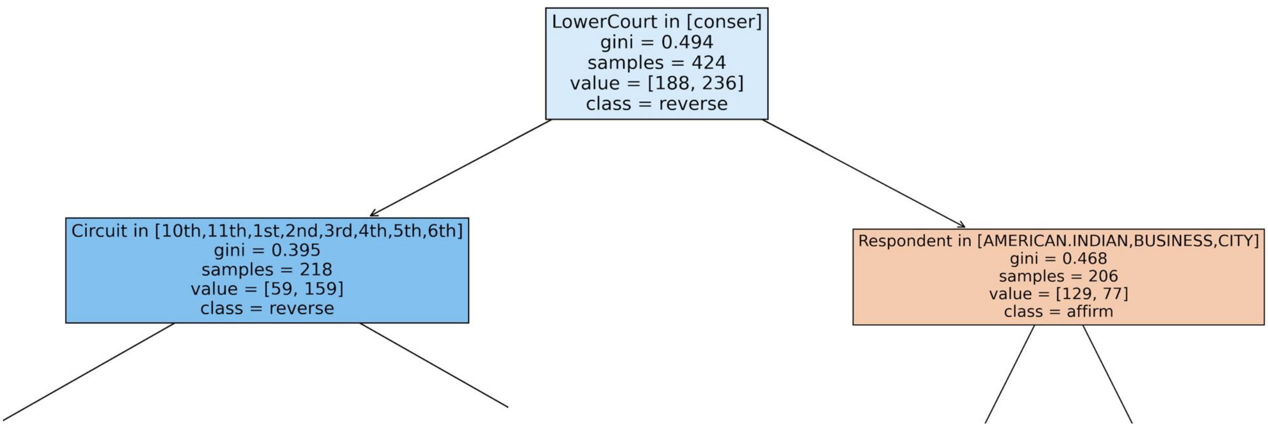
$$\text{Accuracy}(\text{model}) = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} = \frac{42+47}{42+27+26+47} = 0.6267605633802817$$

$$\text{Recall}(\text{model}) = \frac{\text{True Positive}}{\text{True Positive + False Negative}} = \frac{47}{26+47} = 0.64383562$$

Testing the model



Testing the model



Testing the model

```
|--- LowerCourt in ['conser']
|   |--- Circuit in ['10th', '11th', '1st', '2nd', '3rd', '4th', '5th', '6th']
|   |   |--- return 1
|   |--- Circuit in ['7th', '8th', '9th', 'DC', 'FED']
|   |   |--- Issue in ['Attorneys', 'CivilRights', 'CriminalProcedure', 'DueProcess', 'EconomicActivity']
|   |   |   |--- return 1
|   |   |--- Issue in ['FederalTaxation', 'FederalismAndInterstateRelations', 'FirstAmendment', 'JudicialPower', 'Privacy', 'Unions']
|   |   |   |--- class: 0
|--- LowerCourt in ['liberal']
|   |--- Respondent in ['AMERICAN.INDIAN', 'BUSINESS', 'CITY']
|   |   |--- class: 1
|   |--- Respondent in ['CRIMINAL.DEFENDENT', 'EMPLOYEE', 'EMPLOYER', 'GOVERNMENT.OFFICIAL', 'INJURED.PERSON', 'OTHER', 'POLITICIAN', 'STATE', 'US']
|   |   |--- return 0
```


Selecting the best model

Prepare the training and testing dataset

Split the dataset into training and testing datasets and validation dataset (unbiased evaluation of the different models)

Training Dataset has 319 rows

Testing Dataset has 142 rows

Validation Dataset has 105 rows

Experimentation parameter

min_samples_leaf

This parameter sets the minimum number of samples required to be at a leaf node. If a split would result in a leaf node with fewer samples than this threshold, the split will be canceled. Higher values of this parameter can lead to simpler trees.

Model testing and experimentation

Experimentation parameters

- Maximum Depth: This parameter restricts the maximum depth of the decision tree by specifying the maximum number of levels it can have. Increasing the maximum depth can lead to a more complex tree with more splits and branches.
- Minimum Samples Split: This parameter determines the minimum number of samples required to split an internal node further. If the number of samples at a node is below this threshold, the node will not be split, resulting in a simpler tree. Increasing this parameter can reduce the complexity of the tree.
- Minimum Samples Leaf: This parameter sets the minimum number of samples required to be at a leaf node. If a split would result in a leaf node with fewer samples than this threshold, the split will be canceled. Higher values of this parameter can lead to simpler trees.
- Maximum Features: This parameter determines the maximum number of features or attributes that are considered when looking for the best split at each node. By reducing the number of features, the complexity of the tree can be reduced.