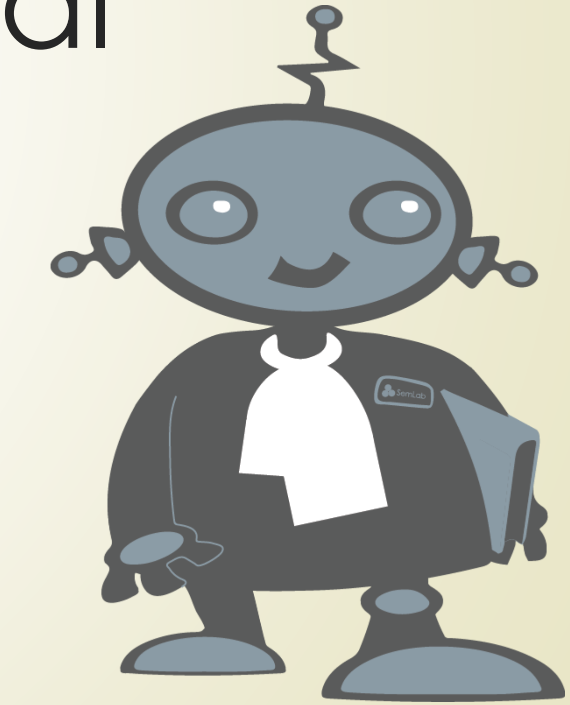
Roie Ben Simon

Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective

Nikolaos Aletras, Dimitrios Tsarapatsanis,

Daniel Preoțiuc-Pietro and Vasileios Lampos



Predicting judicial decisions

- Can a computer predict a judicial decision?
- The European Court of Human Rights (ECHR)
- The ECHR – Structure of decisions
- The dataset
- Dataset Preprocessing
- Training the Model
- Results
- Discussion

Can a computer predict a judicial decision?

- How would you predict a decision?

- Previous attempts mostly focused on disciplinaries in political science and economics:
  - the nature and gravity of the crime
  - Preferred policy position of each judge
  - Age of judge, time to retirement, …

The European Court of Human Rights

- European Court of Human Rights = ECHR
- Established in 1959 in Strasbourg, France.
- Rules mostly in conflicts between individuals and states.
- Handles cases regarding the violation of **the Human Rights convention** (an international treaty to protect human rights and political freedoms in Europe).
- Sort of a 'supreme court' (cases are first handled by other authorities).

- **Why cases from the ECHR?**

The European Court of Human Rights

- **Why cases from the ECHR?**
- Main assumption: there is enough similarity between published judgments and applications/briefs submitted by the parties to the case.
- We can predict future court rulings based on prior documents with similar structure.

The ECHR – Structure of Decisions

- All cases from the ECHR have a specific structure:

1. Procedure

2. The facts:
   - Circumstances of the case
   - Relevant law

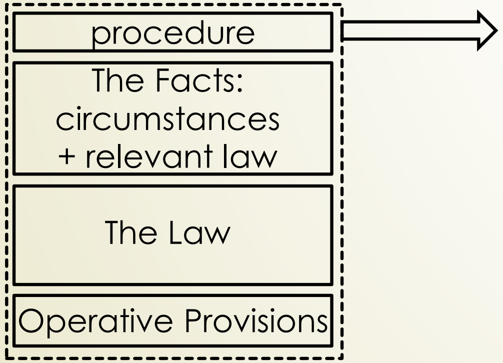
3. The law:
   - Alleged violation of article x
   - Main arguments of both parties

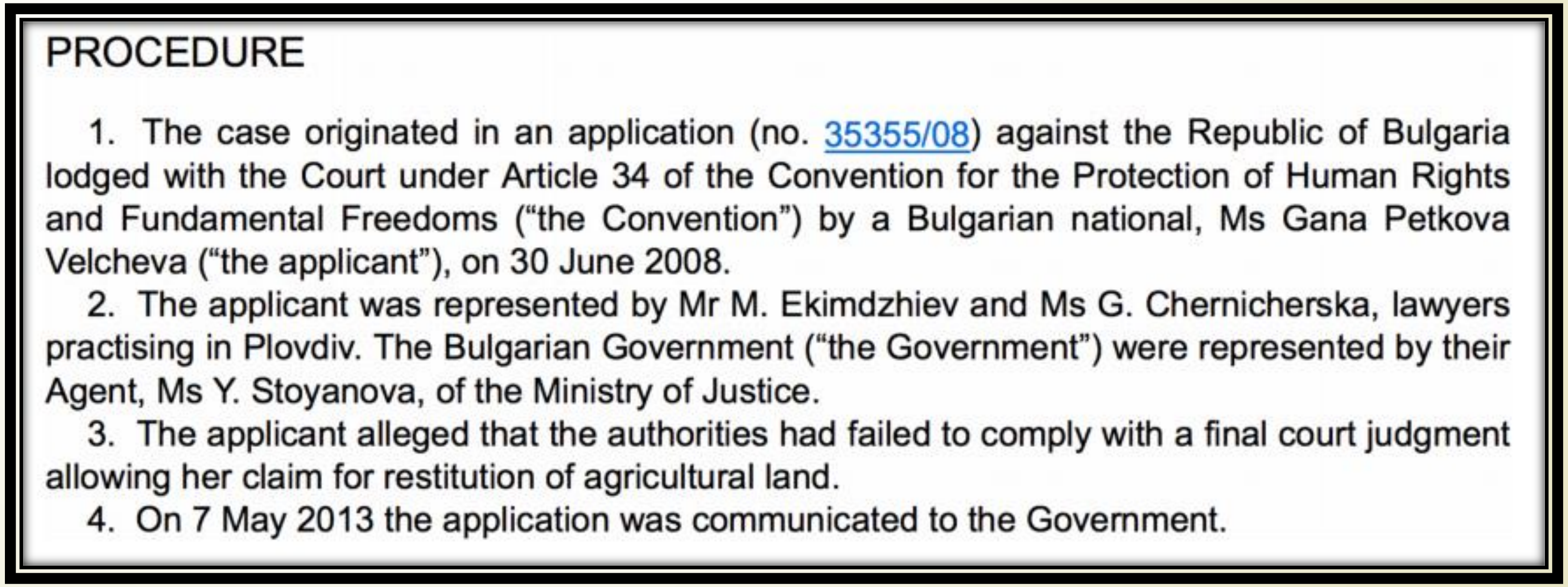
4. Operative provisions

The ECHR – Structure of Decisions

1. procedure

- The procedure followed before the Court, from the lodging of the individual application until the judgment was handed down.





PROCEDURE

1. The case originated in an application (no. 35355/08) against the Republic of Bulgaria lodged with the Court under Article 34 of the Convention for the Protection of Human Rights and Fundamental Freedoms ("the Convention") by a Bulgarian national, Ms Gana Petkova Velcheva ("the applicant"), on 30 June 2008.

2. The applicant was represented by Mr M. Ekimdzhiev and Ms G. Chernicherska, lawyers practising in Plovdiv. The Bulgarian Government ("the Government") were represented by their Agent, Ms Y. Stoyanova, of the Ministry of Justice.

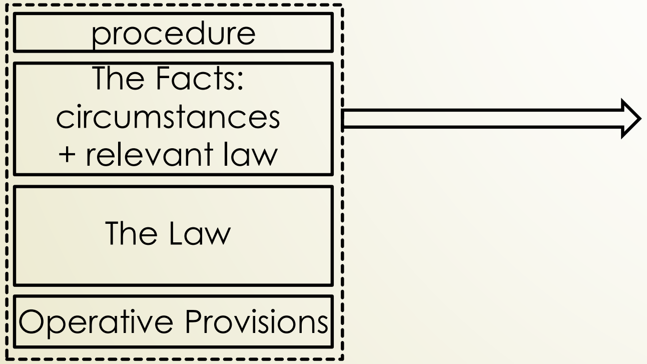
3. The applicant alleged that the authorities had failed to comply with a final court judgment allowing her claim for restitution of agricultural land.

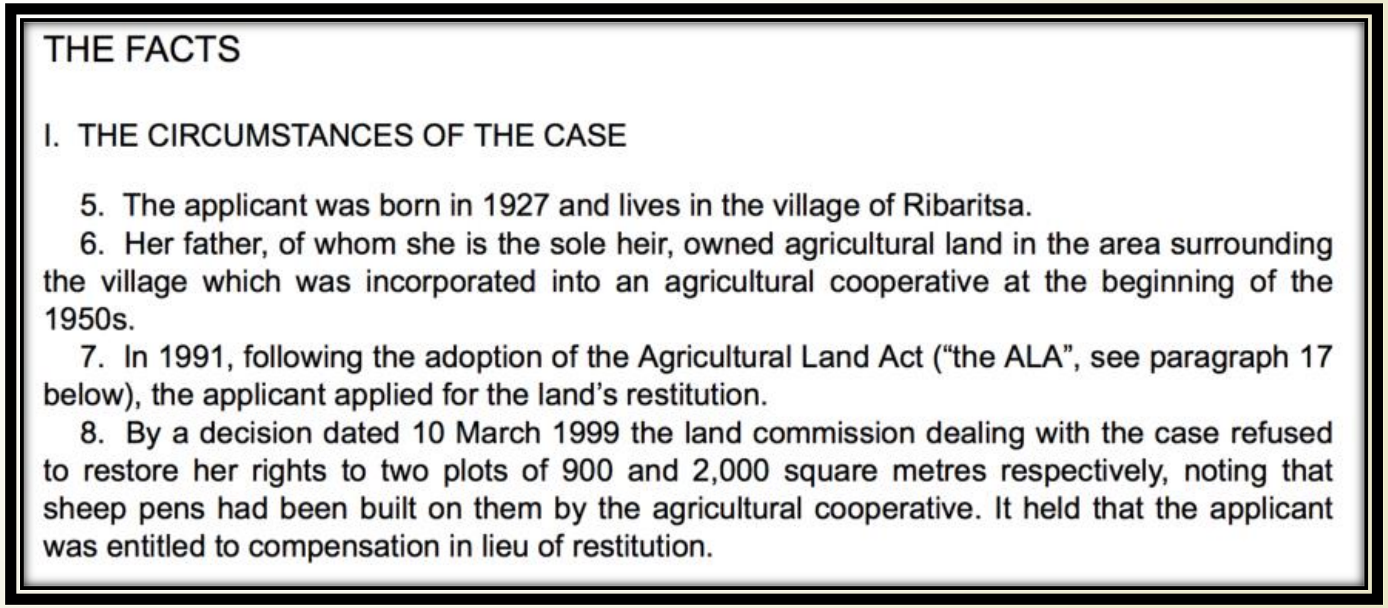
4. On 7 May 2013 the application was communicated to the Government.

The ECHR – Structure of Decisions

2. The Facts

- Circumstances - the factual background of the case and the procedure
- All actions and events that have allegedly given rise to a violation of the ECHR
- Relevant law





THE FACTS

I. THE CIRCUMSTANCES OF THE CASE

5. The applicant was born in 1927 and lives in the village of Ribaritsa.

6. Her father, of whom she is the sole heir, owned agricultural land in the area surrounding the village which was incorporated into an agricultural cooperative at the beginning of the 1950s.

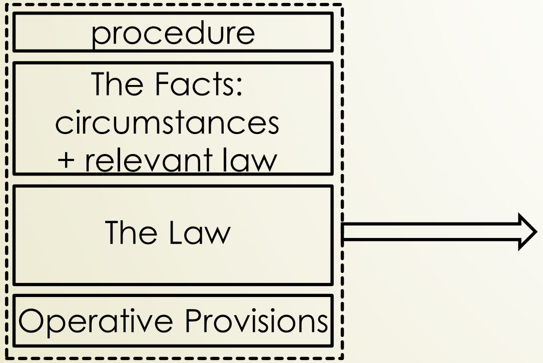
7. In 1991, following the adoption of the Agricultural Land Act ("the ALA", see paragraph 17 below), the applicant applied for the land's restitution.

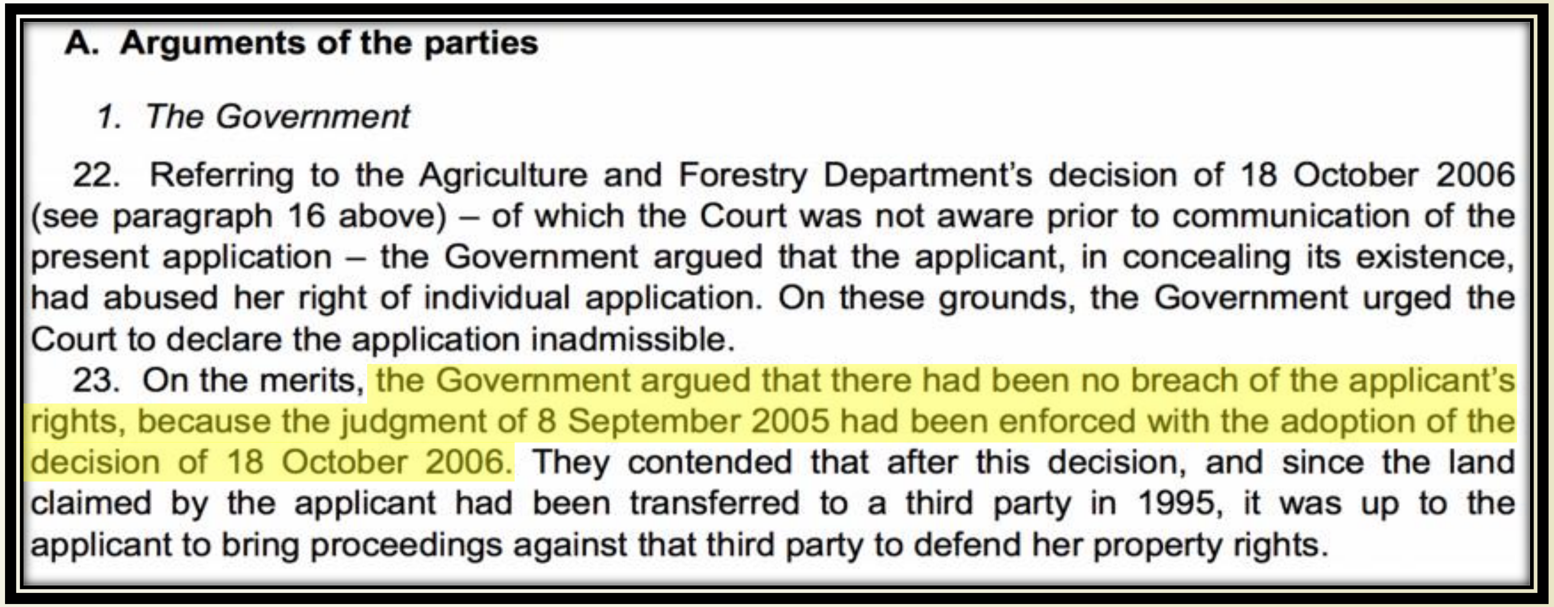
8. By a decision dated 10 March 1999 the land commission dealing with the case refused to restore her rights to two plots of 900 and 2,000 square metres respectively, noting that sheep pens had been built on them by the agricultural cooperative. It held that the applicant was entitled to compensation in lieu of restitution.

The ECHR – Structure of Decisions

3. The Law

- The merits of the case, through the use of legal argument
- The legal arguments used by the parties or the legal reasons provided by the Court





A. Arguments of the parties

1. The Government

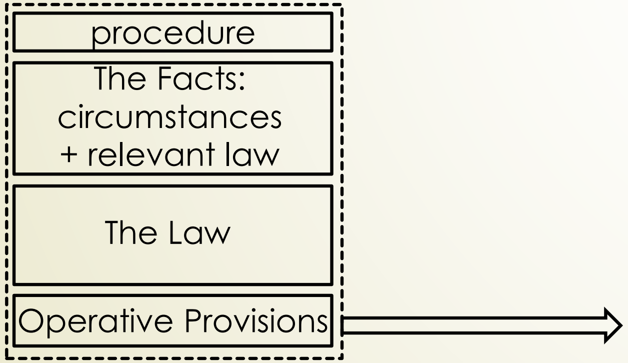
22. Referring to the Agriculture and Forestry Department's decision of 18 October 2006 (see paragraph 16 above) – of which the Court was not aware prior to communication of the present application – the Government argued that the applicant, in concealing its existence, had abused her right of individual application. On these grounds, the Government urged the Court to declare the application inadmissible.

23. On the merits, the Government argued that there had been no breach of the applicant's rights, because the judgment of 8 September 2005 had been enforced with the adoption of the decision of 18 October 2006. They contended that after this decision, and since the land claimed by the applicant had been transferred to a third party in 1995, it was up to the applicant to bring proceedings against that third party to defend her property rights.

The ECHR – Structure of Decisions

4. Operative Provisions

- the Court announces the outcome of the case
- Was there a violation of the convention or not

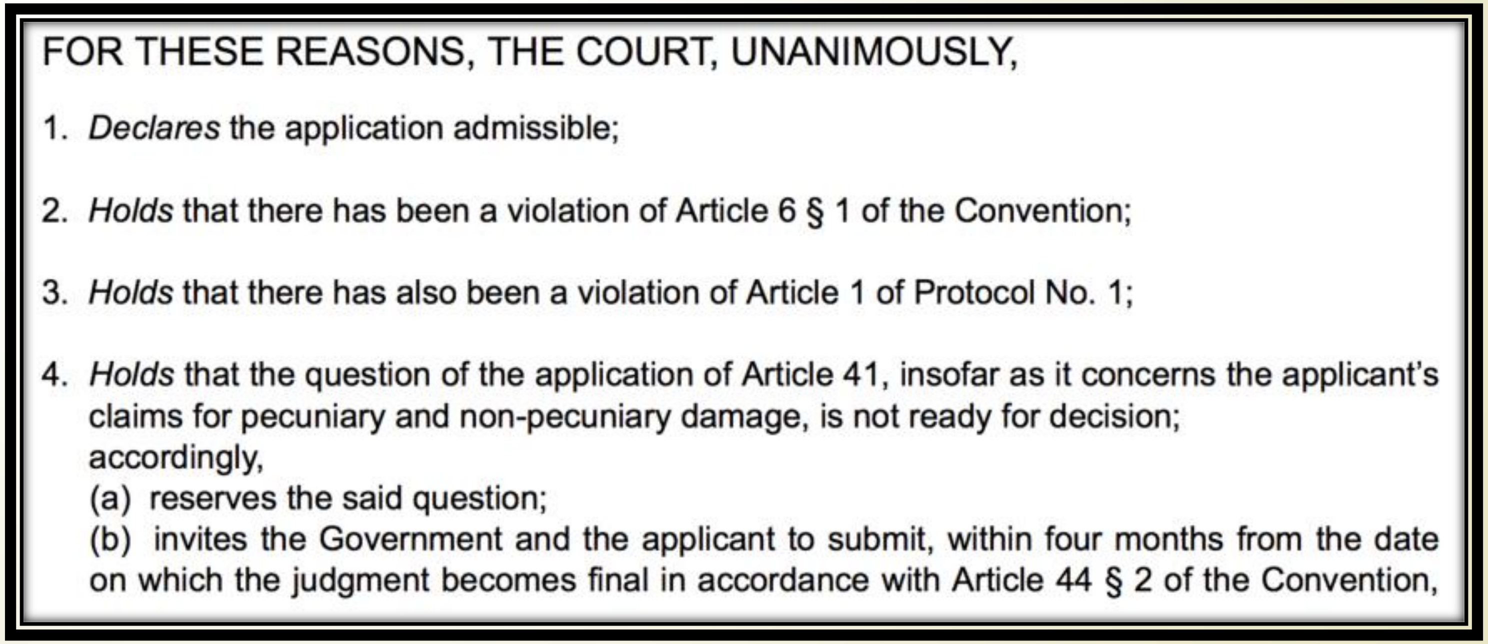


procedure

The Facts: circumstances + relevant law

The Law

Operative Provisions



FOR THESE REASONS, THE COURT, UNANIMOUSLY,

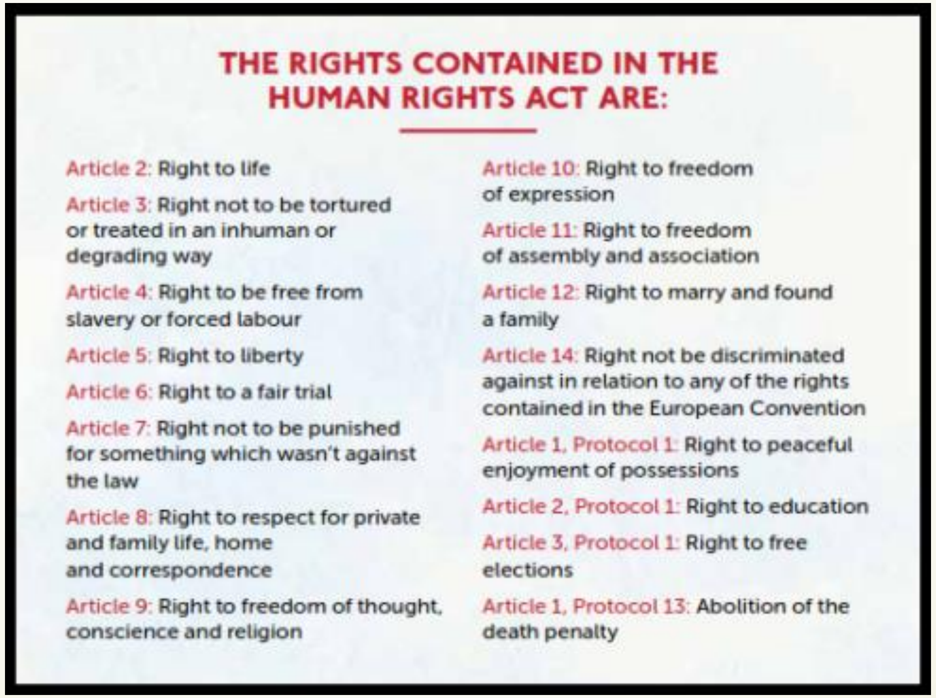
1. *Declares* the application admissible;
2. *Holds* that there has been a violation of Article 6 § 1 of the Convention;
3. *Holds* that there has also been a violation of Article 1 of Protocol No. 1;
4. *Holds* that the question of the application of Article 41, insofar as it concerns the applicant's claims for pecuniary and non-pecuniary damage, is not ready for decision; accordingly,
   (a) reserves the said question;
   (b) invites the Government and the applicant to submit, within four months from the date on which the judgment becomes final in accordance with Article 44 § 2 of the Convention,

The Dataset

- Each set of cases was tested separately
- 250, 80 and 254 cases for Articles 3, 6 and 8 from the convention, respectively.
- Cases are all in English.
- a balanced number of violation/non-violation cases for each article.

The Dataset

- 250, 80 and 254 cases for Articles 3, 6 and 8 from the convention, respectively.



THE RIGHTS CONTAINED IN THE HUMAN RIGHTS ACT ARE:

Article 2: Right to life

Article 3: Right not to be tortured or treated in an inhuman or degrading way

Article 4: Right to be free from slavery or forced labour

Article 5: Right to liberty

Article 6: Right to a fair trial

Article 7: Right not to be punished for something which wasn't against the law

Article 8: Right to respect for private and family life, home and correspondence

Article 9: Right to freedom of thought, conscience and religion

Article 10: Right to freedom of expression

Article 11: Right to freedom of assembly and association

Article 12: Right to marry and found a family

Article 14: Right not be discriminated against in relation to any of the rights contained in the European Convention

Article 1, Protocol 1: Right to peaceful enjoyment of possessions

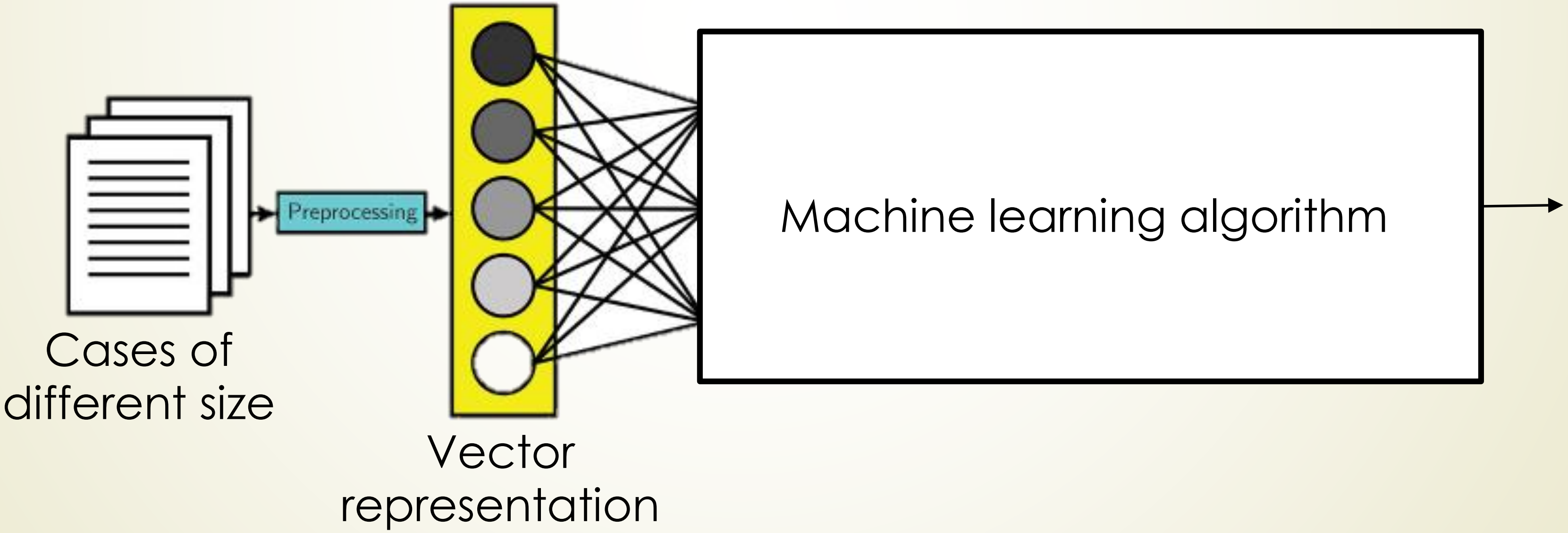
Article 2, Protocol 1: Right to education

Article 3, Protocol 1: Right to free elections

Article 1, Protocol 13: Abolition of the death penalty

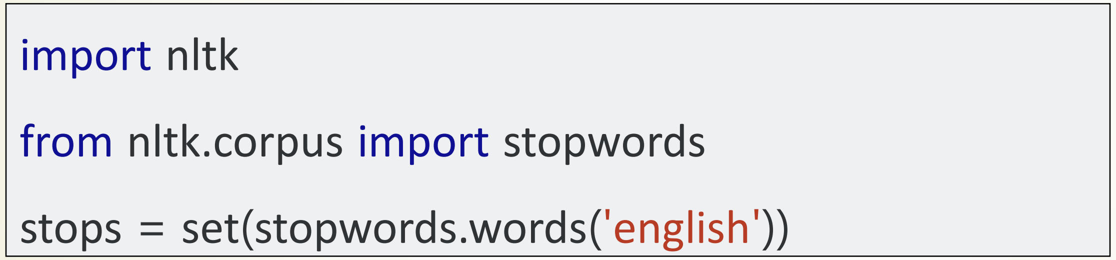
Dataset preprocessing

- Use the text “as is” in the machine learning algorithm?
- Problem: input should be of same length.
- We need to convert the input to a numeric vector



Dataset preprocessing

- All the sections on operative provisions were excluded.
- All cases were lower cased (small letters).
- Stop words removed (“I”, “the”, “and”, “its”, “this”, “that”, …).

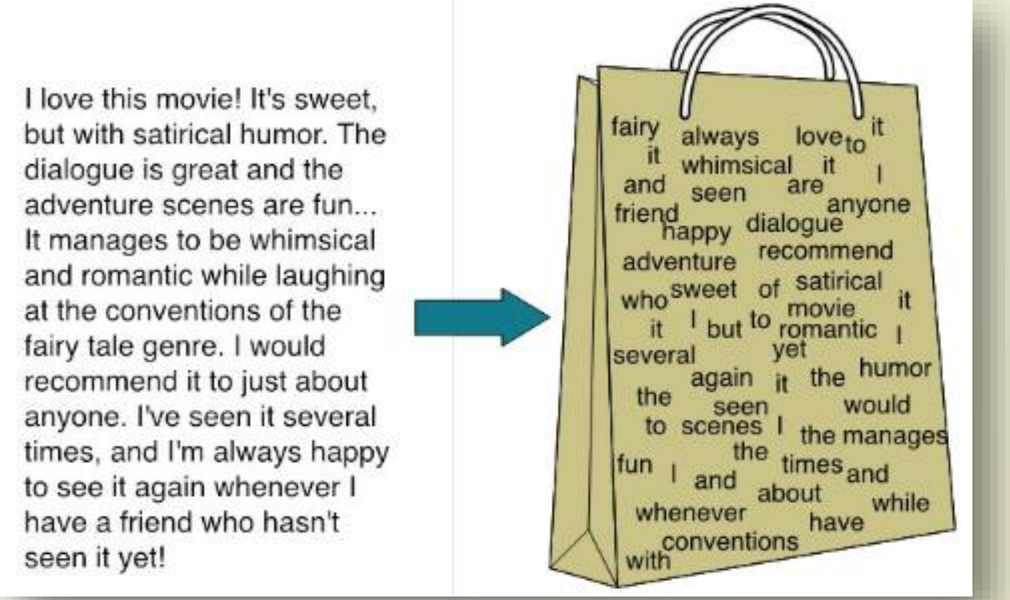


```
import nltk

from nltk.corpus import stopwords

stops = set(stopwords.words('english'))
```

- Can we now use the text as input to our algorithm? Not yet.

Dataset preprocessing

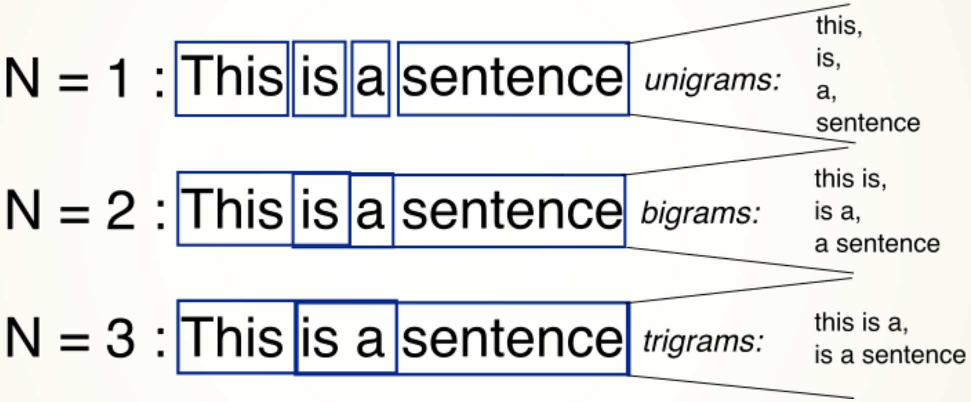
- **Bag of Words model**
- Good representation of text when using NLP
- Treat the text as a “bag of words”
- any information about the order, structure or grammar of words in the document is discarded



I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!

Dataset preprocessing

- **N-gram features**
  - Instead of words, use N-grams.



  - Compute the top 2000 N-grams for each set of cases, for N=1,2,3,4.
  - Each case is represented as a 1*2000 vector.
  - Each entry of the vector counts the number of times a specific N-gram appeared in the case.
  - The vector represents the features of the case

Dataset preprocessing

- Example:

“It was the best of times, → **[1, 1, 1, 1, 1, 1, 0, 0, 0, 0]**

it was the worst of times, → **[1, 1, 1, 0, 1, 1, 1, 0, 0, 0]**

it was the age of wisdom, → **[1, 1, 1, 0, 1, 0, 0, 1, 1, 0]**

it was the age of foolishness” → **[1, 1, 1, 0, 1, 0, 0, 1, 0, 1]**

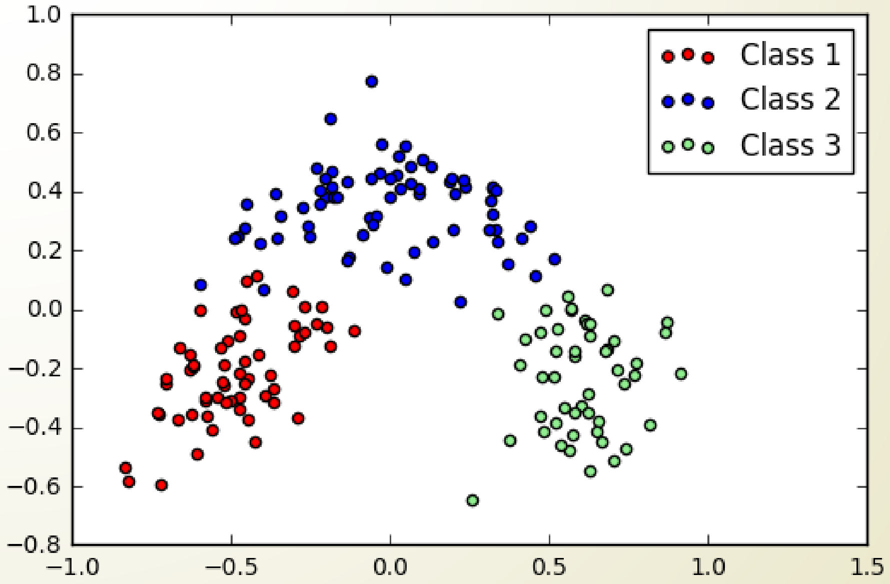
(A Tale of Two Cities, Charles Dickens)

{“it”, “was”, “the”, “best”, “of”, “times”, “worst”, “age”, “wisdom”, “foolishness”}

- In our case: the input is a 1*2000 vector of most common N-grams

Dataset preprocessing

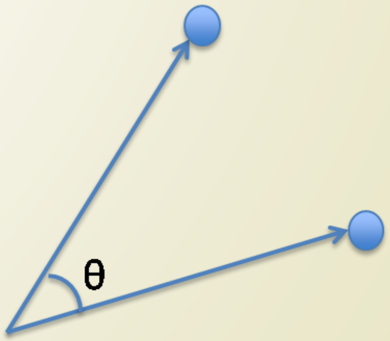
- **Topics/Word Clusters**
- Cluster all N-grams into sets.
- Instead of a 1*2000 vector, use a much smaller vector for each case.
- Main advantage: reduces the dimensionality of the feature space to a 1*30 vector:
  - Less overfitting
  - Less computation time.



Dataset preprocessing

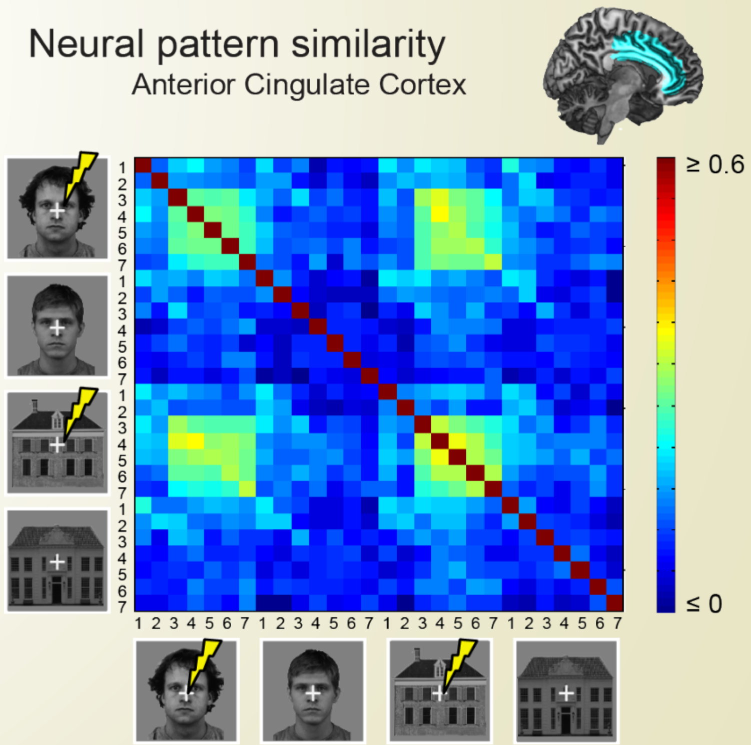
- **Topics/Word Clusters**
- Create a matrix C of all vectors representing one article of the ECHR
- For example, for Article 3: $C_{250*2000}$
- Each column vector in C represents an N-gram.
- Compute N-gram similarity between all vectors using the cosine metric (a measure of similarity between two non-zero vectors). and create an #N-grams*#N-grams similarity matrix.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$



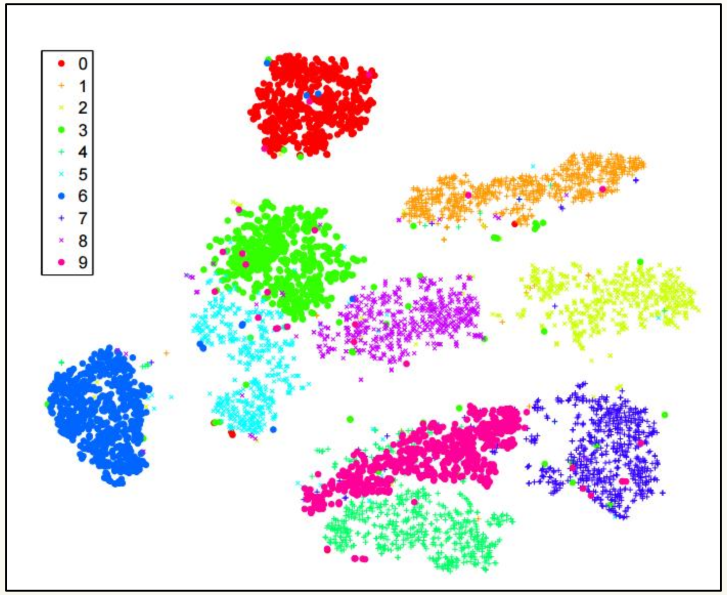
Dataset preprocessing

- **Topics/Word Clusters**
- Compute N-gram similarity between all vectors using the cosine metric (a measure of similarity between two non-zero vectors), and create an #N-grams*#N-grams similarity matrix:
  - Entry (n,k) in the matrix represents similarity between N-gram n and N-gram k.



Dataset preprocessing

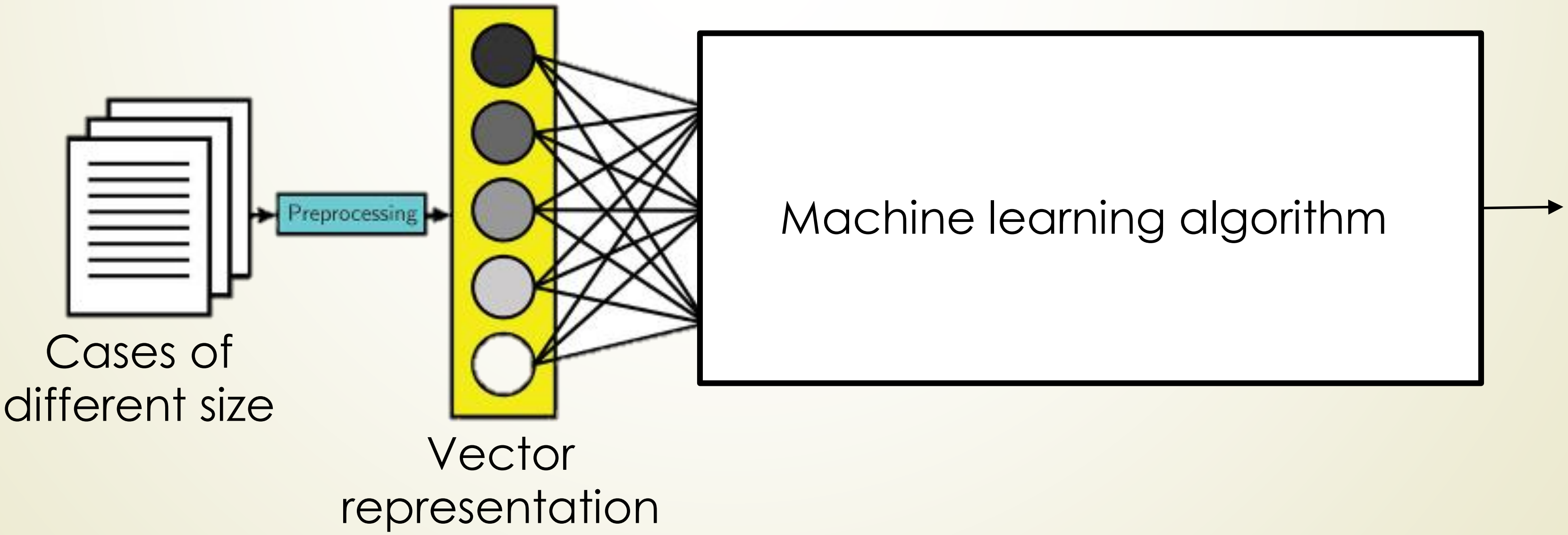
- **Topics/Word Clusters**
- Apply spectral clustering to obtain 30 clusters of N-grams.



- use a 30*1 vector as to represent each case.

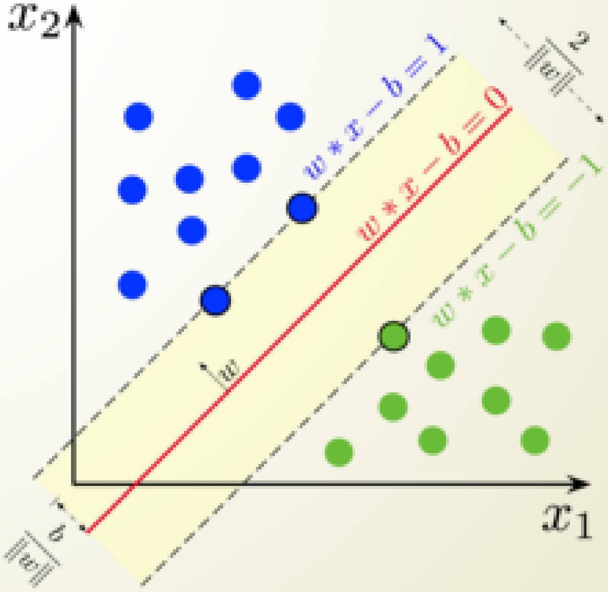
Training the model

- **Binary classification:** predict if, in the context of a particular case, there is a violation or non-violation in relation to a specific Article.
- Non-violation cases were labeled -1, violation cases labeled +1.
- Train a Support Vector Machine (SVM), well suited for text classification.



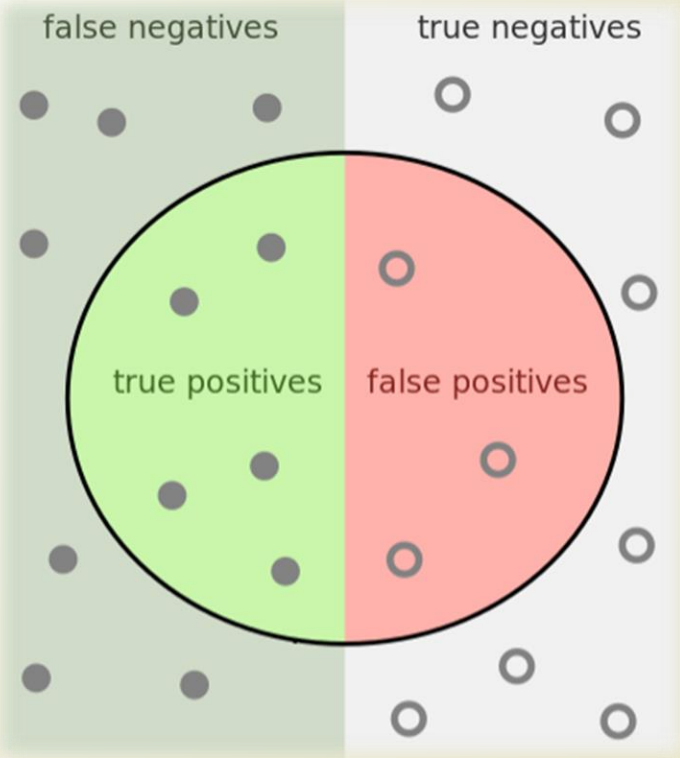
Training model

- **SVM** - Support Vector Machine
- Given a set of training samples, each marked as belonging to one of two categories, the SVM algorithm builds a model that assigns new samples to one of the categories.
- The samples of the two categories are divided by a clear gap that’s as wide as possible.
- The line doesn’t have to be linear.



Results

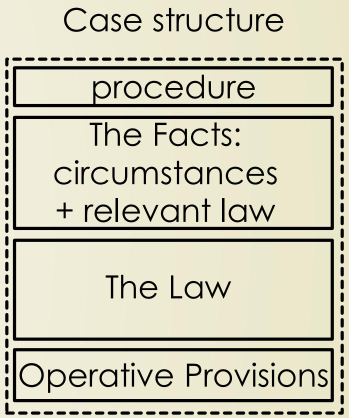
- **Accuracy** = $\frac{\mathrm{TV+TNV}}{\mathrm{V+NV}}$
- TV, TNV – number of true case classifications for violation/ non-violation.
- V, NV – total number of cases w/o violation (true and false classifications).



Results

- Accuracy of a random guess – 50%
- **Accuracy using only N-grams**

Section used when training	Article 3	Article 6	Article 8	Average
Procedure				
Circumstances				
Relevant law				
The Facts				
The Law				
Full (All sections combined)				


Case structure

procedure

The Facts:
circumstances
+ relevant law

The Law

Operative Provisions

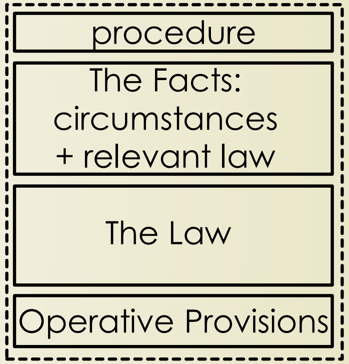
- **Accuracy using topics**

Results

- Accuracy of a random guess – 50%
- **Accuracy using only N-grams**

Section used when training	Article 3	Article 6	Article 8	Average
Procedure	67%	81%	71%	73%
Circumstances	68%	82%	77%	76%
Relevant law	68%	78%	72%	73%
The Facts	70%	80%	68%	73%
The Law	56%	68%	62%	62%
Full (All sections combined)	70%	82%	72%	75%

Case structure



procedure

The Facts:
circumstances
+ relevant law

The Law

Operative Provisions

- **Accuracy using topics**

	78%	81%	76%	78%

- **Accuracy when using both topics and circumstances**

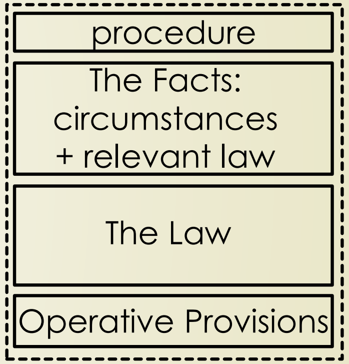
	75%	84%	78%	79%

Results

- Accuracy of a random guess – 50%
- **Accuracy using only N-grams**

Section used when training	Article 3	Article 6	Article 8	Average
Procedure	67%	81%	71%	73%
Circumstances	68%	**82%**	**77%**	**76%**
Relevant law	68%	78%	72%	73%
The Facts	70%	80%	68%	73%
The Law	56%	68%	62%	62%
Full (All sections combined)	**70%**	**82%**	72%	75%

Case structure



procedure

The Facts:
circumstances
+ relevant law

The Law

Operative Provisions

- **Accuracy using topics**

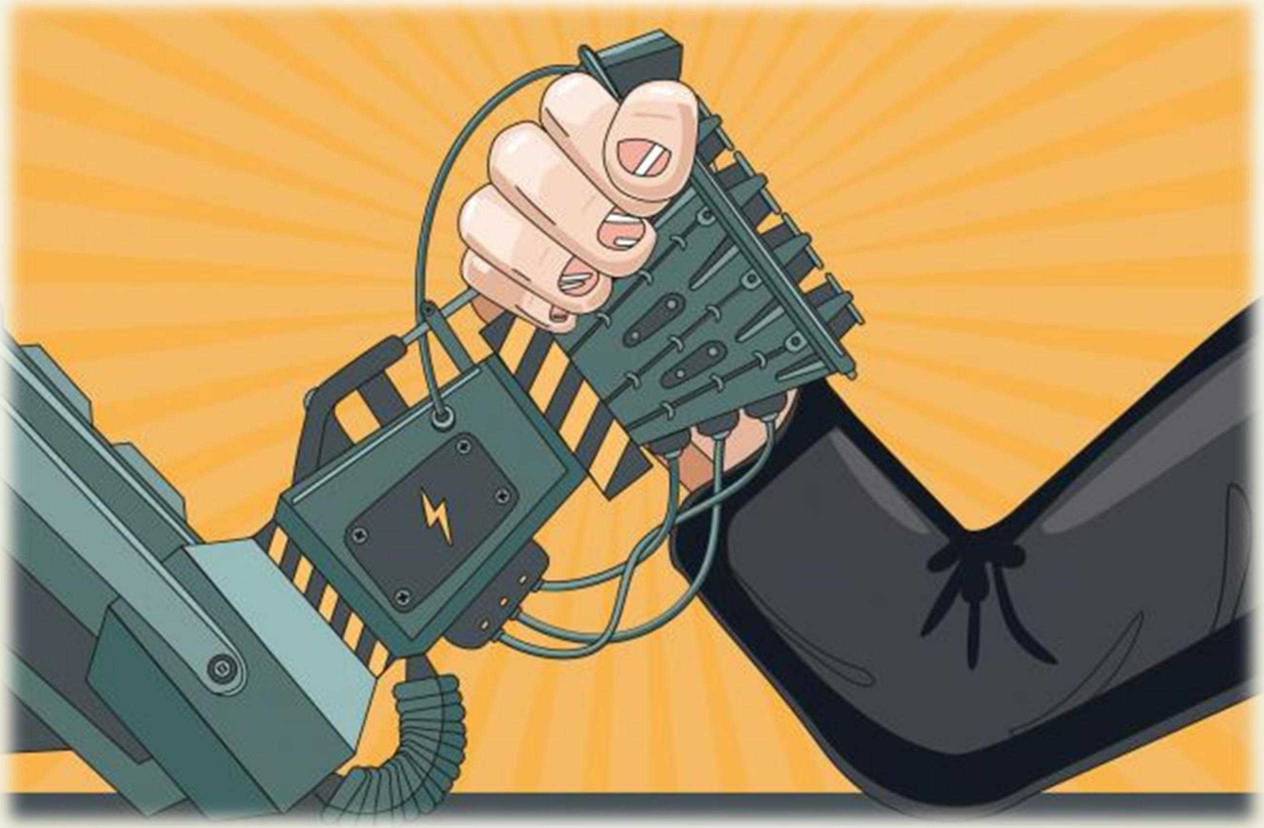
	78%	**81%**	**76%**	**78%**

- **Accuracy when using both topics and circumstances**

	75%	**84%**	**78%**	**79%**

Discussion

- Should a computer determine legal decisions?



Discussion

- Other, less controversial ways to use of machine learning in the legal field:
  - Find valid legal arguments
  - Summarize cases in order to find supportive arguments
  - Identify legal trends
  - divide cases into subsections
  - …

The end