

# A practical application of Artificial Intelligence techniques for legal context analysis



Ilaria Angela **Amantea**, Chiara **Bonfanti**, Michele **Colombino**, Giorgia **Iacobellis**, Susanna **Marta**, Rachele **Mignone**, Marianna **Molinari**, Ivan **Spada**, Laurentiu Jr Marius **Zaharia**

**Supervisor:** Emilio **Sulis**, Luigi **Di Caro**, Guido **Boella**



**UNIONE EUROPEA**  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



*Agenzia per la  
Coesione Territoriale*



# Legal context Analysis

- ❖ Augmented reading for understanding juridical context
- ❖ Taxonomy transfer and classification
- ❖ Principles of law extraction



# Data and sources

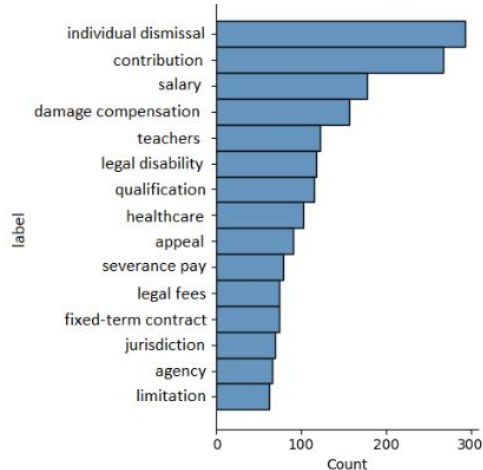
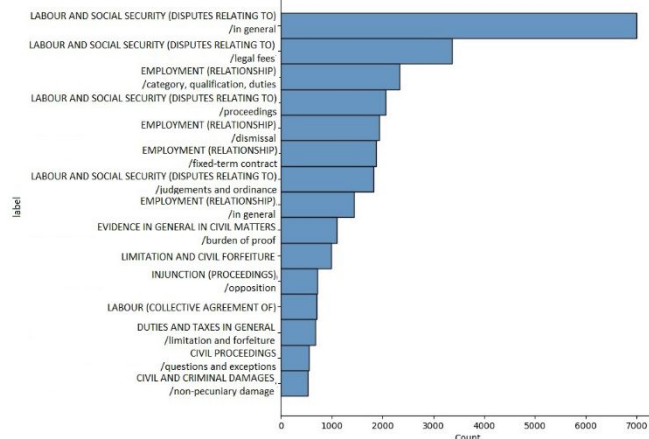
- **Turin corpus** (309 labels)

- 27,477 judgments from the **Court of Turin** (Labor)
- only **4804** are labelled
- Linear structure

- **Leggi di Italia corpus** (300 labels)

- 21,562 judgements from **Leggi d'Italia** web archive
- Taxonomic relationship

- ❖ Corpora are **highly unbalanced!**
- ❖ Labels from the two sources are structured **differently!**
- ❖ All the judgments were been converted into a **machine-readable format**



# Understanding the juridical context

- ❖ Reading, analysis, and comparison follow a partially standard procedure
- ❖ Relevant informations can be south in specific location
- ❖ A semantic context can be extracted automatically



# Locate the relevant info

The judgments have a specific **structure**, specific **language**, and different importance of information depending on **location**

- ❖ Relevant information can be sought in **recurring positions**
- ❖ The document is divided into **core parts** (header, facts, reasoning, and decisions)
- ❖ Each section describes a **different core point** of the legal case
- ❖ End users may be interested in a **core point analysis**

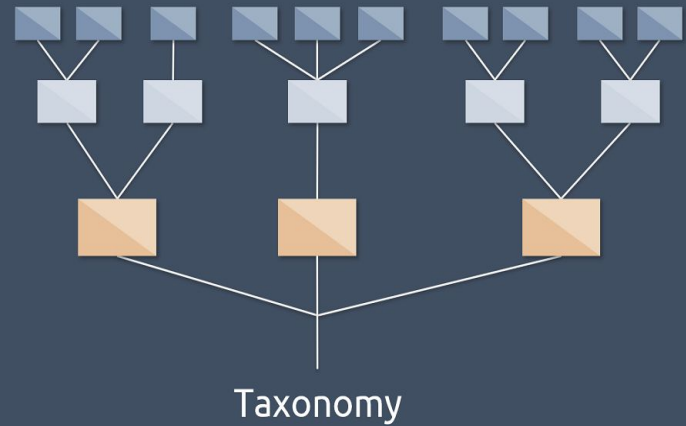
# Building a semantic context

Scraping and segmented the document into core parts, led us to perform an “**augmented reading**”

- ❖ Each one of the core parts Court, Section, Matter, Outcome, Citations may be used to define a **semantic context**
- ❖ Semantic context analysis can be customized by assigning a weight to each core part
- ❖ The weight depends on the criterion chosen for the definition of the legal context
- ❖ **Pros:** customized analysis, assign meaning to specific paragraph, reduction of the computational load

# Taxonomy transfer and classification

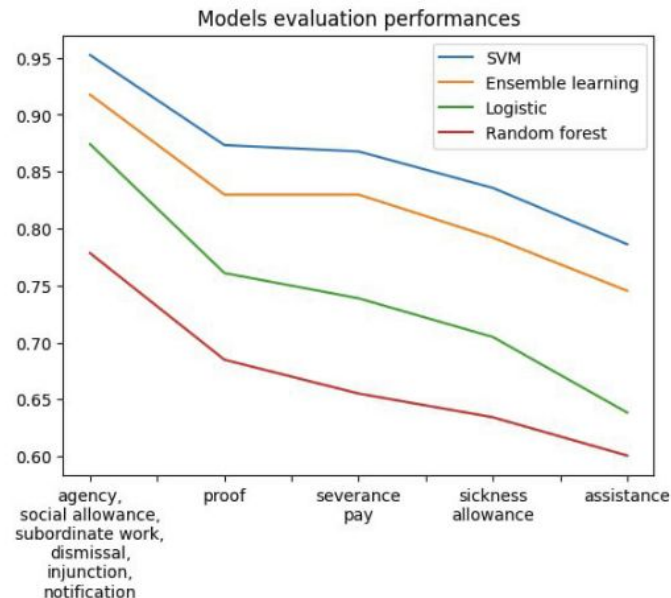
- ❖ Why alignment?
- ❖ Alignment or transfer?
- ❖ Transferring taxonomy pipeline
- ❖ Automatic classification



# Transferring taxonomy Pipeline

The entire process of **transferring taxonomies** takes place in three different phases:

- ❖ Label comparison
- ❖ Semantic similarity
- ❖ Validation with classification





# Automatic classification

## Datasets:

- ◆ Turin: 1872 judgments on 15 labels
- ◆ Leggi d'Italia: 7308 judgments on 11 labels
- ◆ TF, TF-IDF
- ◆ Italian-Legal-BERT
- ◆ Doc2Vec

## Machine Learning models:

- ◆ Support Vector Machine
- ◆ Logistic regression
- ◆ Random forest classifier

## Results

- ◆ **97% accuracy**, on the Turin dataset
- ◆ **94% accuracy**, on the Leggi d'italia dataset

Test 2 - corpus 15 classes				
Dataset	Random forest	SVM	Logistic Regression	Ensemble Voting
Average accuracy				
TF	0.784	0.776	0.802	0.816
TF-IDF	0.784	0.805	0.794	0.808
Ita-legal BERT	0.722	0.714	0.786	0.741
Doc2Vec	0.914	0.954	0.962	0.957
Average precision				
TF	0.859	0.829	0.791	0.765
TF-IDF	0.865	0.859	0.837	0.853
Ita-legal BERT	0.773	0.835	0.766	0.853
Doc2Vec	0.943	0.966	0.972	0.965
Average recall				
TF	0.730	0.723	0.785	0.788
TF-IDF	0.726	0.744	0.737	0.750
Ita-legal BERT	0.640	0.595	0.748	0.750
Doc2Vec	0.878	0.945	0.955	0.955

# Principles of law extraction

- ❖ What are principles of law?
- ❖ How they can help for the definition of a legal context
- ❖ A first approach to the extraction



# Principles of law

Principles of law can be defined as an **interpretation of rules**, in Italy issued by the **Court of Cassation**

- ❖ Indicate the situation to which a certain **rule may be applied**
- ❖ Principles of law do not constitute a source of law per se
- ❖ They may contribute to the construction and the analysis of a legal context

# Automatic extraction

Supreme Court citations			<i>Metrics and evaluations</i>				
<b>data</b>	<b>retrieved relevant</b>		<b>intersection threshold</b>		<b>precision</b>	<b>recall</b>	<b>f1score</b>
unstructured_data	5	10	4	70%	0.8	0.4	0.533
structured_data	8	13	8	100%	1.0	0.615	0.761

- ❖ A **supervised** approach to the automatic extraction using a **regex** methodology
- ❖ **Datasets** annotated by domain experts
- ❖ Two experiments: **structured data** and **unstructured data**
- ❖ Best results on structured data with **76%** of **f1-score**

# Thank You for Your Attention



UNIVERSITÀ  
DI TORINO

UniTO  
Department of Computer Science