Hybrid AI Analysis of the Drug Micro-trafficking in Italy

Salvatore Sapienza(✉)

Department of Legal Studies, University of Bologna, Bologna, Italy
salvatore.sapienza@unibo.it

Abstract. This paper analyses factual and legal aspects on the *quantum* of criminal sanctions in cases of drug micro-trafficking. This case study considers the Italian jurisdiction, which contemplates two legal qualifications of drug dealing, i.e., "minor" and "non-minor" offences. As a case-by-case analysis is required to courts of merits in deciding how the offence has to be legally qualified, the study aims to identify trends and to cast light and provide explanations on the judicial decision-making about the legal qualification of a set of facts. The study emphasizes the importance of combining criminal judgment with innovative tools to facilitate the work of judges and allow information-based studies about jurisprudential trends.

Keywords: Legal Data Analysis · Digital Justice · Micro-trafficking · Digital transformation · Data Science in justice

1 Introduction

When a criminal judge of merits is tasked with evaluating specific facts and applying syllogistic reasoning to relevant criminal facts, transparency, uniformity, and coherence must be ensured in the decision-making process. Although Italy is a Civil Law country, the principles of predictability of criminal sanctions and formal equality require that some degree of relevance is given to past cases. The similarity between the factual elements of the current case and previous ones, as well as the legal proximity of relevant institutions, can determine the level of closeness between analogous cases. To uphold equality in criminal justice, it would be necessary to provide judges with information that can quickly help them to navigate legal and factual elements of a case, enabling them to make well-informed decisions about the quantum of sanctions. Moreover, such information could be helpful to identify latent trends or emerging patterns in jurisprudence, thus being helpful for legal scholars, practitioners, students, and the whole judicial system.

The urgency of the issue arises particularly in cases of micro-trafficking that are often brought to the judge's attention. Micro-trafficking involves easily identifiable typifying elements such as criminal conduct, the object of the crime, the criminally relevant quantity of traded substance, and the subjective element.

A. Kö et al. (Eds.): EGOVIS 2023, LNCS 14149, pp. 103–117, 2023.
https://doi.org/10.1007/978-3-031-39841-4_8

Taken together, these factors make micro-trafficking a suitable scenario for the automated analysis of legal texts. The repetitive nature of these cases provides an ideal testing ground for combining criminal judgment with innovative tools aimed at the analysis of rulings.

The frequency of micro-trafficking cases requires the semi-automated identification of constituent elements of the criminological type directly from previous case law texts. Once identified, these elements can be further extracted, refined, and analysed using legal data analytics tools to facilitate the work of judges facing similar cases. This pipeline constitutes the object of this introductory study. By analysing a corpus of rulings, this study identifies trends and provide explanations on judicial decision-making. Moreover, the goals of this study do not entail "predictive justice" or "judicial profiling". Although some of the models presented hereafter have the ability to predict the outcome of the case, we opted to refrain from predicting the judgements as this would have led to ethically-questionable results.

The paper is structured as follows. Section 2 identifies some related works which constitute the background of this study. Section 3 discusses the research question and the methodology adopted in this paper. Section 4 presents the legal background of drug micro-trafficking in the Italian legislation and case law. Section 5 discusses how the dataset has been built, whereas Sect. 6 presents some findings. A discussion (Sect. 7) on such findings anticipates some final remarks (Sect. 8).

2 Related Works

In his presentation at the seventeenth International Conference on Artificial Intelligence and Law (ICAIL'19), Verheij [1] argued that AI systems in the legal domain should be seen as "hybrid critical discussion systems" (emphasis added). According to this approach, AI systems construct and evaluate hypothetical perspectives to find satisfactory solutions, with the aim of assisting legal professionals in making informed decisions and providing accurate legal advice, not necessarily relying solely on AI systems. Legal analytics through Artificial Intelligence consists of several approaches [2]. The field of computational law aims to address complex problems, such as legal interpretation [3], argument mining [4–6], rule extraction [7], and the management of temporal aspects of legal documents [8,9]. To achieve these goals, researchers have developed two main categories of computational approaches: legal expert systems and legal text analysis [2,10,11].

Automated legal expert systems use rule-based systems, case-based reasoning, and machine learning algorithms to provide legal advice or decision-making support. Rule-based systems mimic logical reasoning used by legal experts for interpretation and rule extraction. Legal text analysis uses natural language processing and machine learning to extract information from legal documents and provide insights into legal cases. Techniques such as argument mining and classification are used to identify arguments and categorize legal texts. The overviews provided in [2, p. 73, 260] clarify the scope of legal information retrieval from statutory text (e.g., prohibitions, obligations, and so forth and from rulings (e.g., arguments). The scenario presented in this study differs from this literature because it aims at extracting factual information from rulings. Therefore, some adaptations to the methods presented in the related works are necessary. Such changes are discussed in the section below.

3 Research Question and Methodology

The research methodology used in this study is a hybrid approach that combines unsupervised and supervised learning experiments on a legal corpus. This methodology involves several steps, which are illustrated in Fig. 1. The first step is a legal analysis that identifies an interesting or controversial research question for legal scholars. In this study, the research question is: "What factual legal factors, both quantitative and qualitative, contribute to the classification of drug-dealing actions as "minor" or "non-minor" crimes?" As we will see in the next section, this question is significant for legal scholars seeking to provide answers to the demand for legal certainty in cases where the classification of the offence is uncertain. While the motivation behind this question is primarily legal, a legal-informatics approach can be useful in formulating or validating legal hypotheses. Then, following the definition of the question, legal research aims to identify a) legal "signals" in the language that are relevant to address the research question and b) what legally-qualified factors can be of interests. For instance, a) the occurrence of the string <number + unit of measure + "of" + illegal substance> is frequent, and b) it can be used to investigate whether correlations with legal factors such as the applicability of drug trafficking legislation.

Legal signals are formalised in a light taxonomy, which helps navigating the signals and their relationship. Such taxonomy does not consists of all the elements of legal ontologies (e.g., restrictions and limitations), yet it allows a broad understanding of how the elements that are necessary to perform the analysis are related. Thus, where applicable, MeLOn methodology for legal ontologies is used to design such light taxonomy [12]. Then, unsupervised experiments are carried out to understand the extent to which the dataset contains features that allow the automated retrieval of information. Extracted factors are analysed and, if necessary, the light taxonomy is refined to include the newly-extracted factors. Legally-relevant factors are extracted from the text. Such factors can be extracted manually (e.g., in [13]) by means of annotating legal documents, or automatically (e.g., in [14]). This study adopted both the approaches by means of regular expressions and manual extraction/validation of the legal factors. Finally, following the measurement of the performance by metrics, these results are interpreted in a legal sense. Limitations of the method are discussed and mitigating solutions are proposed for further refinements. An answer to the research question is provided after the twofold (technical and legal) validation.

The methodological pillars of the proposed method include exaplainability and knowability [15] of the results. This implies that attention is placed on

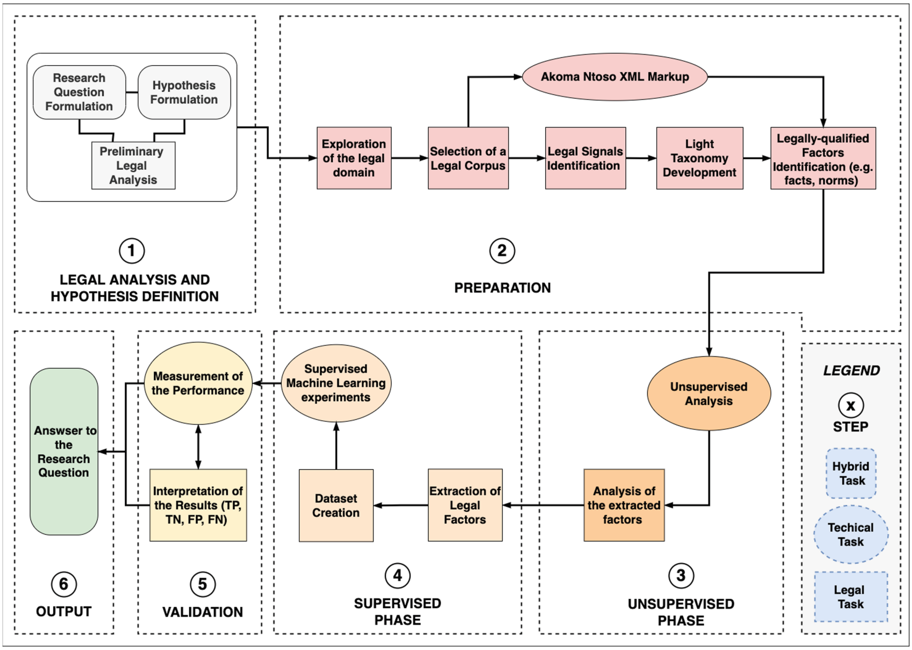

Fig. 1. Hybrid Research Methdology

allowing the correct interpretation of the results from a legal perspective. This is in line with recent studied in legal informatics, which have correctly pointed out the necessity of explaining conclusions in legal analysis software [16,17].

4 The Legal Background in the Italian Micro-trafficking

Article 73 of the Presidential Decree 309/199n outlines the penalties for drug sale and trafficking. If someone grows, makes, sells, transports, or trades any narcotic or psychotropic substance without the necessary authorization, they will be sentenced to imprisonment for a period of 6 to 20 years and a fine ranging from €26,000 to €260,000. In 2019, the Italian Constitutional Court has lowered the minimum sentence from 8 years to 6 years by considering the 8-year minimum disproportionate to the offence[1], also in the light of the offence range of more serious crimes like personal injuries or murder attempts [18]. Therefore, great attention is placed on the offensiveness of the drug-dealing action when the case is decided and the defendant is sentenced.

Alongside the provision against serious drug trafficking, a specific mitigating circumstance is the one provided in Article 73(5), which states that if the actions committed, the way they were carried out, or the substances involved are considered "minor", the punishment for violating the law will be imprisonment for

[1] Italian Constitutional Court Ruling 40 of 2019.

a period of 1 to 6 years and a fine ranging from €3,000 to €26,000. Given the differences in the legal regime, legal reasoning is necessary to correctly qualify the facts as a "minor" or "non-minor" offence[2]. The significance of "borderline cases", i.e., those "placed in a grey zone between the two types of offence"[3] is relevant considering the *nullum crimen sine lege* principle, according to which individuals should now the possible no conduct shall be held criminal unless it is specifically described in the behaviour circumstance element of a penal statute [19].

Crucially, the legislation does not specify any threshold separating "minor" and "non-minor" facts. The evaluation shall be carried out in courts of merits. The Italian Court of Cassation has stated that a) a case by case analysis has to be carried out by each judge and, b) the attribution of the "minor" or "non-minor" charge has to be grounded on quantitative and qualitative elements of the facts[4].

The risk is that "[t]he extent of the penalty gap inevitably conditions the overall assessment that the trial judge must make in order to ascertain the minor extent of the fact (deemed necessary by the Court of Cassation, joint criminal sections, ruling n. 51063/2018), with the risk of giving rise to punitive inequalities, in excess or in default, as well as unreasonable application discrepancies in a significant number of conducts"[5]. The problem lies both on the extent of the penalty range and on the twofold regime of "minor" and "non-minor", which requires a necessary qualitative and quantitative assessment. However, such evaluation is not straightforward as it seems. Qualitative and quantitative elements may be defined in an (almost) infinite number of possibilities, which include information related to the illegal substance as well as factual elements linked to *mens rea* and *actus rei* such as the habituality of the charged person.

Moreover, other elements may have an impact over the final decision ("minor" or "non-minor") despite not being directly mentioned by the law. For instance, a condition of recidivism or the procedure (e.g., shortened proceedings, pleas) might influence the decision. Therefore, it is necessary to make a selection of the features that might be relevant. To preserve the legal integrity of the study, only certain elements which are undoubtedly relevant to the illegal substance have been considered, such as the quantity/quality and purity indicators of the

---

[2] As one of the reviewers has correctly pointed out, this differentiation might lead to ambiguity in the definition of "non-minor" offences, which may encompass every legal qualification other than "minor". However, this lexical choice was done on purpose to mimic the Italian jurisdiction, which contemplates two different charges ("minor" and "non-minor" offences), and an aggravating circumstance for "serious" offences without a specific charge. Accordingly, the problem has been approached as a binary classification problem.

[3] Italian Constitutional Court Ruling 40 of 2019 §5.2.

[4] Court of Cassation, ruling n. 51063/2018 issued by the joint criminal sections, §5; Court of Cassation, 7th criminal section, ruling n. 6621/2019; Court of Cassation; 7th criminal section, ruling n. 3350/2019; Court of Cassation, 4th criminal section, ruling n. 2312/2019.

[5] Italian Constitutional Court Ruling 40 of 2019 §5.2.

substance. Alongside them, procedural elements are used to investigate potential correlations with the final legal qualification of the facts ("minor"/"non-minor").

In 2022, a study carried by the Italian Court of Cassation found out that the judges of its Sixth Section consider the "minor" qualification to be correctly applied by looking also (and not exclusively) to the quantitative elements related to the absolute weight of the traded substance[6] [20]. Crucially, judges from the Court of Cassation do not decide as courts of merits, but ultimately provide the correct interpretation of the law in case of jurisprudential debate or uncertainty on points of law. Its interpretation becomes highly influential, yet an in-depth analysis is required to judges on the merits in cases such as micro-trafficking. Moreover, the adaptation from the courts of merits to such "minor"/"non-minor" thresholds identified by the Court of Cassation might take some time. In the meanwhile, it is worth reconstructing previous decisions rather than forecasting future outcomes. In fact, the soundness of any forecast on previous cases would be seriously hindered by the new trend emerging in the rulings issued by the Court of Cassation.

5 Dataset Composition

The composition of the dataset is the following. From a corpus of rulings (N = 340), a subset (N = 88) of rulings in which the suspect has been found guilty has been extracted. In 49 rulings (.558), the offender has been charged with "minor" drug dealing, whereas in the remaining 39 (.442), the offence has been considered "non-minor". ll the rulings are anonymised prior to further processing in compliance with the applicable laws and codes of ethics. Table 1 describes the data available in the dataset, a short description of the information available, the data type, the unit of measure (when available) and how the information has been extracted.

The issue of managing legal references (legislation and precedents) is addressed by modeling legal knowledge through an international XML standard, Akoma Ntoso [21], adopted by OASIS and used by many institutions for marking up and sometimes even drafting their own legal documents[7] The benefits of using Akoma Ntoso lie in enriching the digitized text with precise legal references (articles, commas, statutes, etc.) and point-in-time information, allowing for a secure extraction of provisions and precedents. Unlike other models, this enrichment preserves the legal meaning of the original text and serves as useful metadata for automatic processing while respecting the semantics attributed by humans. Akoma Ntoso supports judicial documents by identifying `<introduction>`, `<background>` (facts), `<motivation>` and `<decision>` elements within the `<judgementBody>` of the cases alongside metadata [22].

---

[6] These thresholds are the following 1) 23.66 g for cocaine; 2) 28.4 g for heroin; 3) 108.3 g for marijuana; 4) 101.5 g for hashish.

[7] See http://docs.oasis-open.org/legaldocml/ akn-core/v1.0/os/part1-vocabulary/akn- core-v1.0-os-part1-vocabulary.html, last accessed on 19 April 2023.

Table 1. Information types extracted from the rulings, alongside a short description by, the data type, the unit of measure (when available) and the extraction method

Information	Description	Data Type	Unit of Measure	Extraction
substance	The illegal substance object of trade, possession, etc	String		Regular Expression
absolute_weight	Absolute weight of the illegal substance	Float	Grams	Regular Expression
doses	Number of doses of illegal substance	Float		Regular Expression
weight_active_principle	Weight of the illegal substance relative to the absolute number	Float	Grams	Regular Expression
percentage	Percentage of purity of the substance	Float	Percentage	Regular Expression
punishment_days	Days of confinement in jail	Integer	Days	Regular Expression
pecuniary_punishment_amount	Amount of the pecuniary sanction	Integer	Euro	Regular Expression
minor_offence	Whether the offence has been considered non-minor (0) or minor (1)	Boolean		Manual
recidivism	Whether the offender is not a re-offender (0) or he/she is a re-offender (1)	Boolean		Manual
plea_bargain	Whether the offender has not negotiated a plea bargain (0) or he/she has (1)	Boolean		Manual
shortened_proceeding	Whether the offender has opted for full proceeding (0) a shortened proceeding (1)	Boolean		Manual

Some clarifications on certain information types might be necessary. Regarding `substance`, it is worth noticing that judges usually prefer the common name (e.g., cocaine, hashish) instead of the scientific name reported provided by the law. Therefore, the regular expressions used to extract the information did not include scientific formulae in favour of the common name. In the case multiple substances have been found in the same case, the one with the highest number of doses have been kept. Regarding `absolute_weight` and `weight_active_principle`, the regular expressions contemplate several units of measures ("kg", "mg", "mgr", etc.), also in full-text word in Italian (e.g., "chilogrammo").

The extracted information have been standardised in grams, which was the most common unit of measure. Regarding `percentage` and `weight_active_principle`, missing data have been calculated by dividing or multiplying the available information types according to the target data. Regarding `punishment_days`, the regular expression matches all the possible combinations of year, month, and days expressed in numeric or string format. Dictionary-based conversions were used to convert string to numbers, and then multiplications to standardise the duration of the punishments in days were done. Regarding `pecuniary_punishment_amount`, the extracted information consisted of numbers, string and combinations of the two. As with the previous case, dictionary-based conversions were used to make the amount unique. The other information types have been extracted manually from the text of the ruling. Regarding `substance`, a closed list of the most frequent illegal narcotics (marjuana, heroine, cocaine, hashish, crack, ketamine) have been encoded to be represented as a categorical attribute. Finally, regarding Regarding `doses`, the number has been either extracted by the ruling or calculated by applying the average weight : dose ratio available from cases in which the same substance was traded.

6 Findings from Data Analysis

Following the hybrid methodology described in Sect. 3, an unsupervised K-Means clustering (K = 3) experiment was run to understand whether groups of "minor" and "non-minor" rulings can be automatically identified taking into account all the available attributes, with potential outliers a third option for outliers. The results are displayed in Fig. 2 with days of punishment and amount of pecuniary sanction as axes.

As expected, a division can be observed between a first groups of rulings (highlighted in red), possibly associated to "minor" cases, and another group of rulings (highlighted in green). A third group was added to the clustering to identify potential outliers. This group likely corresponds to crimes in which the case has been qualified as a drug dealing action in which a "huge amount" of narcotics to art. 80(2) of the Presidential Decree 309/1990. This legal qualification is not the object of this study. However, this additional legal factor identified in the unsupervised experiment signals that there is potential to expand the study in this direction. For the purposes of this study, the division between "minor" and "non-minor" seems present in the features of the dataset. Therefore, the analysis can go on without the necessity of immediately refining the light taxonomy and proceeds with additional features. First, we run an experiment to determine what model is more promising keeping constant certain parameters. We use `scikit-learn` in Python[8] instantiating a Decision Tree and a Random

[8] Scikit-learn is a Python library for machine learning, providing tools for data preprocessing, classification, regression, clustering, and more.

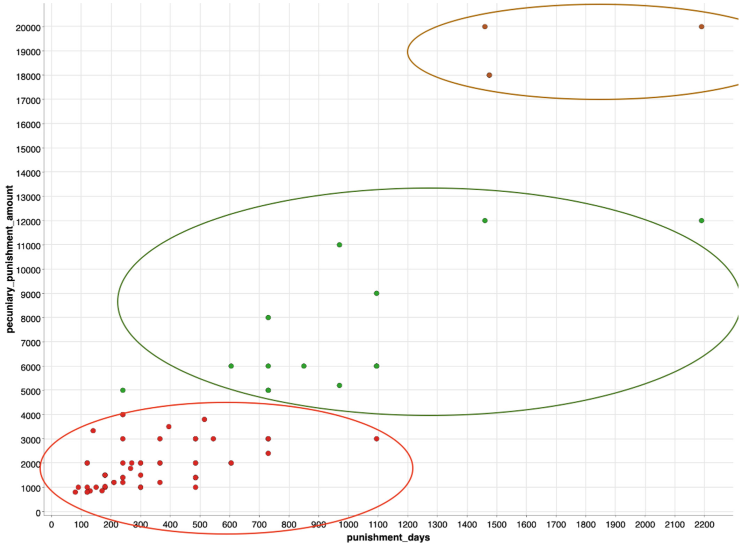

Fig. 2. Clustering of cases with K=3. The chart displays groups of court rulings (red and green) for "minor" and "non-minor" offences. A third group (brown), contains outliers which might include cases in which the aggravating circumstance for "serious" offences has been applied (Color figure online)

Forest models with `max_depth=8`. At the same time, similar experiments were run with KNIME[9].

The table presents the results of several classification experiments using different machine learning algorithms, namely Decision Tree, Support Vector Machine (SVM), Random Forest, and Gradient Boosted Trees. The classification tasks uses the predictive features to classify `minor_offense`. The experiments were performed with varying training and test ratios of 80:20, 75:25, and 70:30 to evaluate the impact of dataset compositions in a binary classification with few data available. The evaluation metrics used to compare the algorithms were precision, accuracy, recall, and F1-measure[10].

Observing the results, it can be seen that the performance of the algorithms varies significantly depending on the training and test ratios. In general, the Decision Tree algorithm performed well, especially when using the KNIME

[9] KNIME is a data analytics platform that allows users to visually design workflows, integrating various data processing and machine learning algorithms for advanced analytics and predictive modelling.

[10] These metrics are widely utilized for assessing the performance of classification models. Precision and recall are especially valuable for datasets that exhibit significant class imbalance, wherein the number of instances in various classes differs substantially. In contrast, accuracy is well-suited for balanced datasets. The F1-measure is a composite metric that integrates both precision and recall and is considered a more equitable evaluation criterion for classification models.

Table 2. This table presents classification metrics for different models. The metrics include precision, accuracy, recall, and F1-measure, calculated for different ratios of training to test data.

Model	Training: Test	Precision	Accuracy	Recall	F1-measure
Decision Tree (scikit-learn)	**80:20**	**0.55**	**0.56**	**0.56**	**0.55**
Decision Tree (scikit-learn)	75:25	0.50	0.50	0.50	0.49
Decision Tree (scikit-learn)	70:30	0.45	0.44	0.45	0.44
SVM (scikit-learn)	80:20	0.38	0.39	0.39	0.37
SVM (scikit-learn)	75:25	0.34	0.36	0.36	0.34
SVM (scikit-learn)	70:30	0.42	0.44	0.44	0.42
Random Forest (scikit-learn)	80:20	0.34	0.39	0.39	0.34
Random Forest (scikit-learn)	75:25	0.39	0.41	0.41	0.38
Random Forest (scikit-learn)	70:30	0.48	0.48	0.48	0.47
Decision Tree (KNIME)	80:20	0.61	0.61	0.63	0.61
Decision Tree (KNIME)	75:25	0.53	0.55	0.53	0.51
Decision Tree (KNIME)	**70:30**	**0.65**	**0.65**	**0.65**	**0.65**
Tree Ensemble (KNIME)	80:20	0.54	0.53	0.56	0.53
Tree Ensemble (KNIME)	75:25	0.64	0.54	0.54	0.54
Tree Ensemble (KNIME)	70:30	0.54	0.62	0.53	0.49
Random Forest (KNIME)	80:20	0.63	0.61	0.58	0.54
Random Forest (KNIME)	75:25	0.64	0.54	0.54	0.54
Random Forest (KNIME)	70:30	0.60	0.58	0.58	0.54
Gradient Boosted Trees (KNIME)	80:20	0.61	0.61	0.61	0.61
Gradient Boosted Trees (KNIME)	75:25	0.59	0.55	0.56	0.55
Gradient Boosted Trees (KNIME)	70:30	0.54	0.54	0.54	0.54

implementation, achieving higher precision, accuracy, recall, and F1-measure across all training and test ratios. The SVM algorithm, on the other hand, performed poorly, with relatively low precision, accuracy, recall, and F1-measure across all ratios. The Random Forest algorithm achieved moderate to good results, with the KNIME implementation generally performing better than the scikit-learn implementation (Table 2).

After evaluating the performance of the different machine learning algorithms, it is worth exploring some of the factors that may explain the observed differences between the two sets of algorithms. With this regard, they may be explained by several factors. For instance, the decision tree algorithm used in KNIME may have a different stopping criterion than the one used in scikit-learn, which can result in different tree structures and ultimately, different classification accuracy. Also, Scikit-learn and KNIME may have different default hyperparameters for the same algorithm.

The choice of hyperparameters can have a significant impact on the performance of the algorithm, and if the hyperparameters are not tuned properly, it can lead to suboptimal performance. In these experiments, hyperparameters were left untouched. `GridSearchCV` method from Sckit-learn was used to perform an exhaustive search over specified hyperparameter values for a given estimator. This method performs cross-validation for each combination of hyperparameters and returns the best set of hyperparameters that yields the highest score on a specified evaluation metric. The improved model had a theoretical F1-measure of 0.67, which is compatible to the KNIME one. Since the decision tree was the best performing algorithm in both tests, it is convenient to visualise the criteria on which the classification has been performed. The visualisations purposively display the most influential factors from both algorithms to ease the global understanding and interpretability of the results (Figs. 3 and 4).

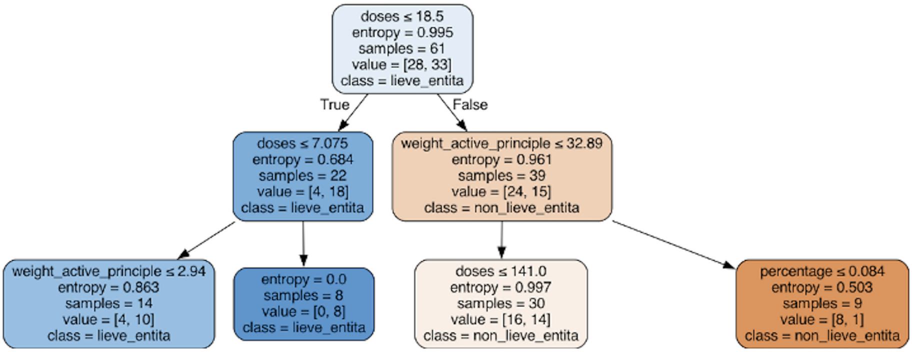

Fig. 3. The best-performing decision tree using Sckit-learn. Blue-highlighted cells represent the relevance for "minor" classification, orange colours stand for relevance of the factor in "non-minor" classifications. Darker colours represent higher relevance (Color figure online)

In both cases, the first criterion is `doses`, which varies from <18.5 and >18.5 (Sckit-learn) and <17.5 and >17.5 (KNIME) for "minor" and "non-minor" classification respectively. The difference may be due to the same factors (hyperparameters) identified when discussing the discrepancy in performance. However, they seem highly compatible given the small quantitative differences among them. In both cases, the quantity of `weight_active_principle` seems quite relevant as it is ranked among the most influential factors. As it will be discussed in the next section, this is expected with the legal analysis carried out in the preliminary stages of the study.

With doses being ranked as the most influential factor in the two best-performing experiments, further visualisations allow the interpretability of the results. Figure 5 below. The position of the rulings based some extracted factors such as `doses` can be visualized in 3 dimensions to be correlated to `pecuniary_punishment_amount`, and `punishment_days`, which should express the "minor" or "non-minor" punishment in tangible terms. Additionally, the volume of dots should to `weight_active_principle` which was considered a relevant factor alongside `doses`. Green dots represent "minor" cases, whereas red dots display "non-minor" cases.

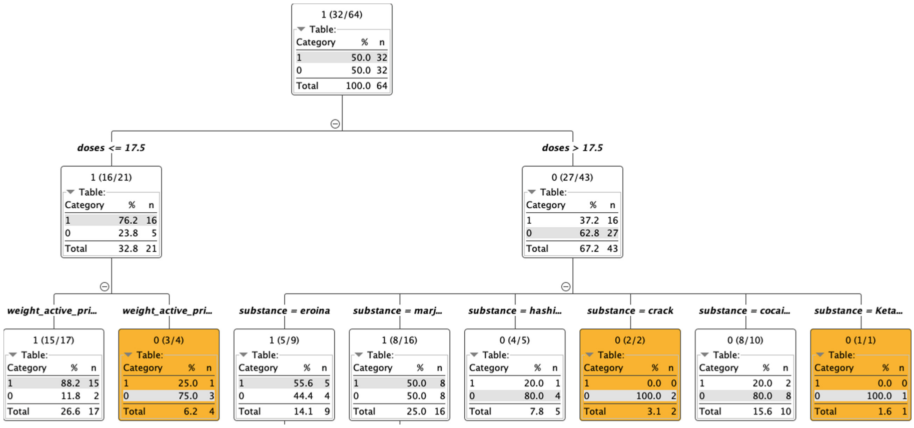

Fig. 4. The best-performing decision tree using KNIME. Each box contains and indication of the criterion and the number of instances from the training set (N = 64) classified according to the factor in bod (e.g. doses). Orange boxes are terminal leaves (Color figure online)

The figure seems to confirm the expected distribution of cases. "Minor" cases are close to the lower number of doses and receive a lower amount of punishment, both in pecuniary and sentencing terms. Outliers are represented by cases that, despite a lower number of doses, display higher punishments. Such cases seem related to the higher number of `weight_active_principle`, which was ranked as a relevant factor in the previous steps. This indicates a higher level of danger despite the lower number of doses, which might be linked to the purity of the traded narcotics.

7 Legal Analysis and Limitations of the Results

The analysis performed over a legal corpus allows for a legal interpretation of the findings. First, the research question posed in Sect. 3 might be answered by proposing the number of doses as the most important factor in the legal qualification of micro-trafficking facts, with the weight of the active principle playing a relevant role. This is confirmed by the double validation performed with different sets of algorithms. Despite some relatively small differences, the two approaches came to the same conclusion under different testing scenarios, in particular as regards the ratio between training and test set.

This finding seems to be confirmed by the recent indications of the Italian Court of Cassation, which has identified some thresholds in the amount (expressed in weight) that might help courts of merits in the legal qualification of the facts. While this jurisprudential trend is on its way of consolidation (in the absence of clear legislative indications), this analysis also contributes to this debate by providing a different perspective on a purely legal matter, which identifies the number of doses as the most influential criterion in courts of merits.

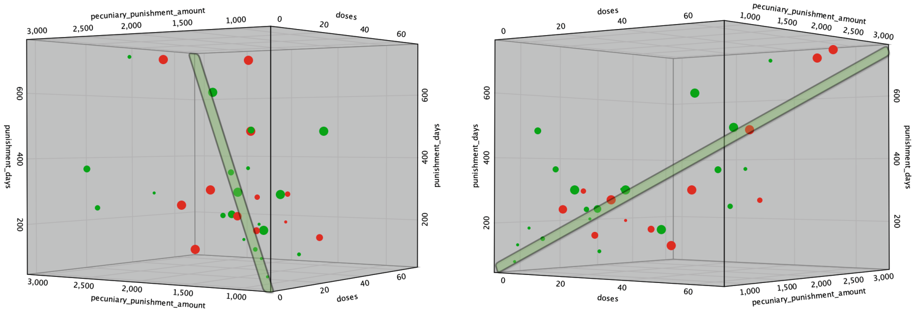

Fig. 5. 3-dimensional visualisations from two different angles of the position the rulings according to `doses`, `pecuniary_punishment_amount`, and `punishment_days`. Dots' volume represents `weight_active_principle`. The green plane represents an ideal linear trend in which the variables are directly proportionate

This paper does not explore further this divergence, as an in-depth discussion would require a different methodology aimed at investigating, as shown in other studies (e.g., in [20]) pros and cons of one quantitative criterion over the other.

Overall, the approach used in this study favours the explainability of the machine learning systems developed to support the legal analysis. By understanding, also by means of visualisation, what the impact of each feature in the final decision, a reasonable margin of appreciation is left to the legal expert in interpreting the results from a legal standpoint.

Let us also address some important limitations of this study. First and most importantly, the dataset used to perform the analysis is relatively small and it has to be increased in further refinements. There are important factors to keep into account. For instance, compliance with data protection law and privacy is a key aspect which has to be investigated before introducing new elements to the dataset. Another limitation regards the quantity and the quality of the feature extracted from the rulings. One key element to underline is that the light taxonomy developed in the early stages of the methodology constitutes an "open world" prone to be enriched by additional elements. Criminal rulings in micro-trafficking present several factual factors that should be kept into account in further refinements (e.g., precautionary measures). At the same time, increasing the complexity of the knowledge on which the model finds correlations and patterns may hinder the overall performances.

This paper has found a trade-off by relying on a method capable of grasping the *legal* significance of pre-identified factors. There are potential avenues for exploring other methods. One could be the automated extraction of linguistic formulas and patterns, already tested in other studies in the legal domain [13]. In this perspective study, it would be necessary to reconcile the linguistic patterns to their relevance in the legal language. Ontologies seem a viable way to perform such bottom-up approach in combination with the top-down pre-identification of the legal elements performed in this study.

8 Conclusive Remarks

Given the relative small size of this legal corpus, it is worth underlining that this study constitutes a preliminary attempt of analysing criminal rulings rather than a complete assessment. Overall, this study shows that the number of doses and the weight of the active principle are relevant factors in the legal qualification of micro-trafficking criminal cases in the Italian jurisdiction. This conclusion can support the research of scholars active in criminal law by providing novel perspectives on a current debate. Although significant under a legal point of view, much has to be done to ensure the reliability of the findings of this paper, in particular in the area of data availability. This might affect the generalizability of the results.

However, there are some important takeaways that deserve attention. First, drug micro-trafficking is a legal sector that can be explored with the lens of legal informatics for further experiments. Future studies should increase the "datafication" of rulings by expanding the factors to be taken into account, in particular in the direction of qualitative factual elements such as indications and evidence. Second, the methodology used in this paper seems appropriate for similar tasks and can be further explored in the domain of legal informatics. Finally, explainability and interpretability of results are necessary to validate the outputs of machine learning algorithms.

Acknowledgements. This presentation has been funded by Programma Operativo Nazionale (PON) "Governance e Capacità Istituzionale" 2014–2020, "Universitas per la Giustizia. Programma per la qualità del sistema giustizia e per l'effettività del giusto processo (UNI 4 JUSTICE) - Macro Area 2 - CUP J19J21026980006"

References

1. Verheij, B.: Artificial intelligence as law. Artif. Intell. Law **28**(2), 181–206 (2020)
2. Ashley, K.D.: Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press, Cambridge (2017)
3. Athan, T., Governatori, G., Palmirani, M., Paschke, A., Wyner, A.: LegalRuleML: design principles and foundations. In: Faber, W., Paschke, A. (eds.) Reasoning Web 2015. LNCS, vol. 9203, pp. 151–188. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21768-0_6
4. Prakken, H., Sartor, G.: Modelling reasoning with precedents in a formal dialogue game. In: Judicial Applications of Artificial Intelligence, pp. 127–183. Springer, Heidelberg (1998). https://doi.org/10.1007/978-94-015-9010-5_5
5. Prakken, H., Sartor, G.: Law and logic: a review from an argumentation perspective. Artif. Intell. **227**, 214–245 (2015)
6. Liga, D., Palmirani, M.: Classifying argumentative stances of opposition using tree kernels. In: Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence, pp. 17–22 (2019)
7. Wyner, A., Peters, W.: On rule extraction from regulations. In: Legal Knowledge and Information Systems, pp. 113–122. IOS Press (2011)

8. Palmirani, M.: Legislative change management with Akoma-Ntoso. In: Legislative XML for the Semantic Web, pp. 101–130. Springer, Heidelberg (2011). https://doi.org/10.1007/978-94-007-1887-6_7
9. Sartor, G., et al.: Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management, vol. 4. Springer, Heidelberg (2011). https://doi.org/10.1007/978-94-007-1887-6
10. Ashley, K.D.: Automatically extracting meaning from legal texts: opportunities and challenges. Ga. St. UL Rev. **35**, 1117 (2018)
11. Robaldo, L., et al.: Introduction for artificial intelligence and law: special issue "natural language processing for legal texts" (2019)
12. Palmirani, M., et al.: Legal ontology for modelling GDPR concepts and norms. In: Legal Knowledge and Information Systems, pp. 91–100. IOS Press (2018)
13. Palmirani, M., et al.: PrOnto ontology refinement through open knowledge extraction. In: Legal Knowledge and Information Systems, pp. 205–210. IOS Press (2019)
14. Palmirani, M., et al.: Hybrid AI framework for legal analysis of the EU legislation corrigenda. In: Legal Knowledge and Information Systems, pp. 68–75. IOS Press (2021)
15. Palmirani, M., Sapienza, S.: Big data, explanations and knowability. Ragion pratica **2**, 349–364 (2021)
16. Atkinson, K., Bench-Capon, T., Bollegala, D.: Explanation in AI and law: past, present and future. Artif. Intell. **289**, 103387 (2020)
17. Sapienza, S., Bomprezzi, C., et al.: Algorithmic Justice e classificazione di rischio nella Proposta AI ACT. In: La Trasformazione Digitale della Giustizia nel dialogo tra discipline: Diritto e Intelligenza Artificiale, pp. 65–114. Giuffré Francis Lefebvre (2022)
18. Gambardella, M.: Illeciti in materia di stupefacenti e riforma dei reati contro la persona: un antidoto contro le sostanze velenose. Cassazione penale **60**(2), 449–486 (2020)
19. Hall, J.: Nulla poena sine lege. Yale Law J. **47**(2), 165–193 (1937)
20. Lombardi, F.: La Cassazione fissa le soglie quantitative per la lieve entità ex art. 73 co. 5 DPR 309/1990. Osservazioni a prima lettura. In: Giurisprudenza Penale Web, vol. 12 (2022)
21. Palmirani, M., Vitali, F.: Akoma-Ntoso for legal documents. In: Legislative XML for the semantic Web, pp. 75–100. Springer, Heidelberg (2011). https://doi.org/10.1007/978-94-007-1887-6_6
22. Ceci, M., Palmirani, M.: Ontology framework for judgment modelling. In: Palmirani, M., Pagallo, U., Casanovas, P., Sartor, G. (eds.) AICOL 2011. LNCS (LNAI), vol. 7639, pp. 116–130. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35731-2_8