3rd International Conference on Research in SCIENCE, ENGINEERING AND TECHNOLOGY

11-13 March, 2021

Oxford – United Kingdom

Auto Law: Web Solution for Registration and Support for Guidance On Rights of Victims of Acts of Violence

Fabricio Luiz de Souza Pereira[1], João Pedro Amaral de Oliveira[2], Luan Alves Godoy[3], Davi Pedro dos Santos[4], Victor Araújo Sguisato[5], Willian dos Santos Miranda[6], Luiz Melk de Carvalho[7], Flávio Henrique Batista de Souza[9]*

Centro Universitário de Belo Horizonte – UNIBH, Brazil
flabasouza@yahoo.com.br

Abstract

Unfortunately, society is marked by issues of intolerance and aggression. Many victims of aggression often do not know whom to appeal to, what are their rights and how to obtain support and help. Recent movements, such as the "Black Lives Matter", emphasize the need to pay attention to this sort of situation. Thus, the proposed web solution of this work, called Autolaw tool was developed, based on technologies such as: Cloud Computing, Natural Language Processing (NLP) and Machine Learning. This tool was developed according to the following methodology: a cloud structure was developed to gather and process the testimony of the user/victim; a set of algorithms was implemented in the back-end of the tool platform to process the demand via NLP and a machine learning process was used to classify the type of legal protection that is needed focusing according to the Brazilian laws applicable to violence against women and racism; and finally, an accuracy assessment via AUC (Area Under the Curve) ranging from 0 to 1 was performed. As a result, a functional web structure was created and accuracy of the Machine Learning algorithms was assessed. In summary, the web tool, represented by the interface via a website, gathers the testimony of the victim testimony; evaluates, via NLP, the present terms; it recognizes the similarity with the terms of protection laws for the elderly, women and racial ethnicity using machine learning algorithms; and returns the protection laws to the user. Experiments were carried out to gather consultations (700 samples) and the accuracy of the tests were evaluated considering the analysis of the resampling process. As a result, an AUC between 0.81 and 0.92 for the recommendations were obtained.

Keywords: women rights; machine learning; natural language processing; black lives matter; racism

1. Introduction

Unfortunately, humanity is marked by situations of lack of respect and consideration for others. Aggressions, insults, offenses and prejudices are inserted in the contexts of history in all nations of the planet. Discrimination and aggression are some of the characteristics of disrespect against human life and it is the focus of this research. The context of racial discrimination has recently gained a very sad chapter, which has resulted in a "Black Lives Matter" social movement (Dave et al., 2020). In relation to aggressions, a very growing context, mainly due to the COVID-19 pandemic, is violence against women (Peterman et al. 2020).

In this context, technologies have been developed to assist the dissemination and analysis of people's rights. In order to analyze the trend of the opinion of the courts, in Brazil, in 2018, supervised machine learning techniques were used in document classification, with an assertiveness above 90% of accuracy, resulting in solid information about the favored part, employee and employer (Barros et al., 2018).

Taking data from the European Court of Human Rights as an example, it was analyzed how natural language processing tools can be used to analyze texts from court proceedings in order to automatically predict (future) court decisions. With an average accuracy of 75% in predicting the violation of 9 articles of the European Convention on Human Rights, the use of machine language showed great potential in the legal scenarios addressed (Medvedeva et al., 2019).

In this research, two technological concepts were used to develop a solution to help the victim in the knowledge of their rights. In this case, NLP (Natural Language Processing) to recognize user demands and Machine Learning algorithms to classify the applicable laws. Thus, the objective of this work is to demonstrate an experimental process of a tool, called AutoLaw, which focuses on the use of NLP to understand the user's questioning about their rights and on the use of machine learning algorithms to classify the applicable laws. Accuracy assessments based on AUC (Area Under the Curve) were used during the experiment.

The laws used for the experiments are from the Brazilian scenario, in this case Law No. 11,340, of August 7, 2006 for the protection of women and Law No. 7,716 of January 5, 1989 for issues of Racism (Brazil, 1999; Brazil, 2006).

2. Material and Methods

This research has three main foundations: first, an assessment of the technological concepts necessary to implement the proposed solution; then, an analysis of the laws that were used as evaluation references and; finally, an implementation of the solution with accuracy experiments to validate the tool.

2.1 Technological Concepts

For the development of the proposed solution, the following concepts were evaluated for the technological foundations necessary to its implementation:

- UML (Unified Modeling Language): For the development of the solution, the standardization of its elements was in accordance with UML, a language that regulates

the representations of each element of a system from its conception to its deployment (Hidayat & Utomo, 2016)

- Cloud Computing: cloud computing resources were used to host and provide the service for recognizing laws, which according to Attaran & Woods (2019) are resources available on the internet that support demands for solutions through availability of them, as a service, by specialized providers.
- Natural Language Processing (NLP): it is a subarea of Artificial Intelligence (AI) that studies the ability and limitations of a machine to understand the language of human beings. The purpose of NLP is to provide computers with the ability to understand and compose texts. "Understanding" a text means recognizing the context, performing syntactic, semantic, lexical and morphological analysis, creating summaries, extracting information, interpreting the senses, analysing feelings and even learning concepts from the processed texts (Hardeniya et al., 2016).
- Natural Language ToolKit (NLTK): a set of open source program modules, tutorials and problem sets that provide ready-to-use computational linguistics courses. NLTK is a platform used to build Python programs that work with human language data for application in NLP. The NLTK defines an infrastructure that can be used to build NLP programs in Python; provides basic classes to represent data relevant to natural language processing; standard interfaces to perform tasks such as tokenization, Part-Of-Speech, parsing and text classification; and standard implementations for each task that can be combined to solve complex problems. (Hardeniya et al., 2016)
- Machine learning: Machine learning algorithms are structures with the ability to learn a problem and predict classifications through prior training. Such learning can be supervised or unsupervised. The algorithms used for this experiment were: Naive Bayes and SVM (Support Vector Machines) (Alpaydin, 2020).
- ROC curve and AUC: According to Silva et al. (2020), for measuring the accuracy of a solution based on machine learning, a reference widespread in the literature is the ROC curve and its AUC (Area Under the Curve). A value between 0 and 1 that is measured in relation to the rate of TP (True Positives) and FP (False Positives) that were obtained before the classification process.

These concepts were used in the solution, developed in the Phyton® programming language.

2.2 Law Concepts

The tool has an evaluation of Brazilian laws that focus on the themes of violence against women and racial prejudice. During the experiments, the following laws were used as a reference:

Law Nº 11.340, of August 7, 2006. Aggression against women:

- *Art. 1 This Law creates mechanisms to restrain and prevent domestic and family violence against women, under the terms of § 8 of art. 226 of the Federal Constitution, the Convention on the Elimination of All Forms of Violence against Women, the Inter-American Convention to Prevent, Punish and Eradicate Violence against Women and other international treaties ratified by the Federative Republic of Brazil; provides for*

the creation of Courts for Domestic and Family Violence against Women; and establishes assistance and protection measures for women in situations of domestic and family violence.

- *Art. 2 Every woman, regardless of class, race, ethnicity, sexual orientation, income, culture, educational level, age and religion, enjoys the fundamental rights inherent to the human person, being assured the opportunities and facilities to live without violence, preserve their physical and mental health and their moral, intellectual and social improvement.*
- *Art. 3 Women will be guaranteed the conditions for the effective exercise of the rights to life, security, health, food, education, culture, housing, access to justice, sport, leisure, work, citizenship, freedom, dignity, respect and family and community coexistence.*
- *§ 1 The public power will develop policies that aim to guarantee the human rights of women in the context of domestic and family relations in order to protect them from all forms of negligence, discrimination, exploitation, violence, cruelty and oppression.*
- *§ 2 It is up to the family, society and the government to create the necessary conditions for the effective exercise of the rights set out in the caput.*
- *Art. 4 In the interpretation of this Law, the social purposes for which it is intended and, especially, the peculiar conditions of women in situations of domestic and family violence will be considered* (Brazil, 2006)

Law Nº 7.716, of January 5, 1989. Racism:

- *Art. 1 Crimes resulting from discrimination or prejudice based on race, colour, ethnicity, religion or national origin will be punished under this Law.*
- *Art. 3 To prevent or impede the access of someone, duly qualified, to any position of the Direct or Indirect Administration, as well as of the public service concessionaires.*
- *Art. 4 Deny or hinder employment in a private company.*
- *Art. 5 Refuse or prevent access to a commercial establishment, refusing to serve, serve or receive a customer or buyer.*
- *Art. 6 Refuse, deny or prevent the enrolment or entry of a student in a public or private educational establishment of any degree.*
- *Article 7 - To prevent access to or refuse accommodation in a hotel, pension, inn, or any similar establishment.*
- *Art. 12. Prevent access or use of public transport, such as airplanes, barge ships, boats, buses, trains, subways or any other means of transport granted.* (Brazil, 1999)

Such laws were the references of the proposed classifications.

2.3 Implementation Methods

During the implementation process, three steps were followed, where:

- firstly, the tool's use cases were developed in UML, followed by its implementation in Phyton®;

- a process of adaptation of the functioning between the classification and NLP technologies was necessary, with an alignment between the contexts translated by the NLP and evaluated later by the machine learning algorithms;
- finally, an accuracy assessment was carried out: for this process, the database was divided into test and validation sets. The learning involved 50 runs of the machine learning algorithm for classification with three options dividing the database: 50/50, 65/35, 75/25. In this case, for example in 75/25, 75% of the base was designated for learning (tests) and 25% was designated for validation.

As assertive as the tool is intended to be, it was implemented in such a way to be easy to be understand and used by the user.

3. Results

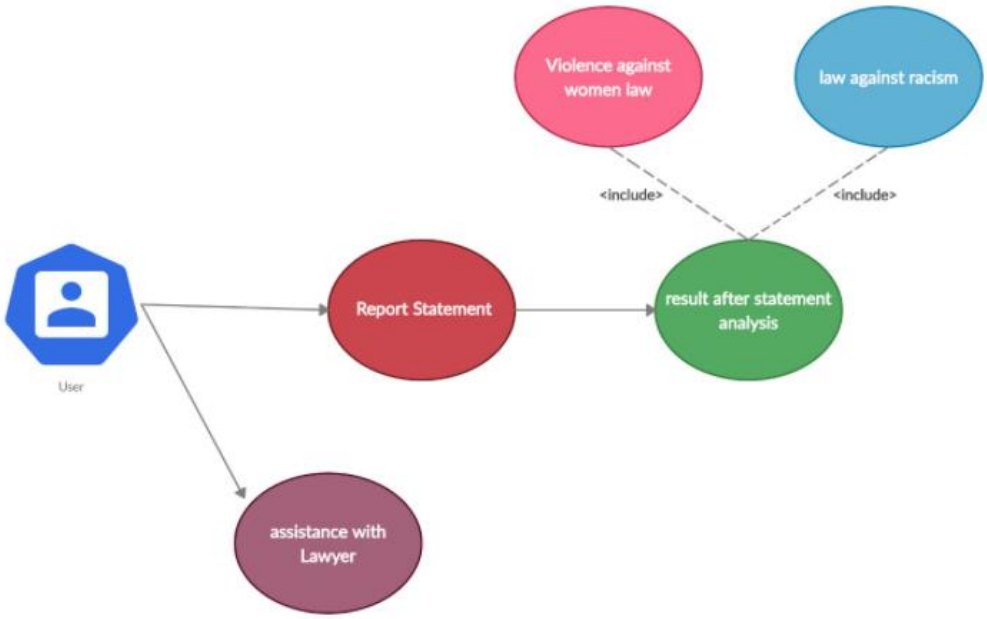
3.1 Structure Development

3.1.1 Web Structure Modelling and Simulation

To model the functionalities of the proposed system, a diagram was developed via UML (Unified Modeling Language). Figure 1 presents the use case diagram, highlighting the functionalities of the implemented system. As can be seen, possible use cases are reported:

- Report Statement: in this use case, the user reports, in their own way, what happened in the situation that they are seeking help. It describes the situation or occurrence;
- Assistance with Lawyer: once the classified demand of the user is recorded, according to the laws learned by the system, the user is shown the law and its terms that guarantee their rights, whether facing an action of racism or violence against women. At this time, lawyers previously registered in the system are recommended to assist the victim.
- Result after statement analysis: In this use case, two others are included, “violence against women law” and “law against racism”. At this moment, the user's demands are evaluated via NLP and the law that meets the user's demand is classified, via SVM.

Figure 1: Case Use Diagram



Source: Authors, 2020.

Thus, the interfaces that are used for interaction with the user were developed, as described in Figure 2. It should be noted that a system has been developed so that it is as simple as possible, as it is intended that the tool is accessible regardless of educational level.

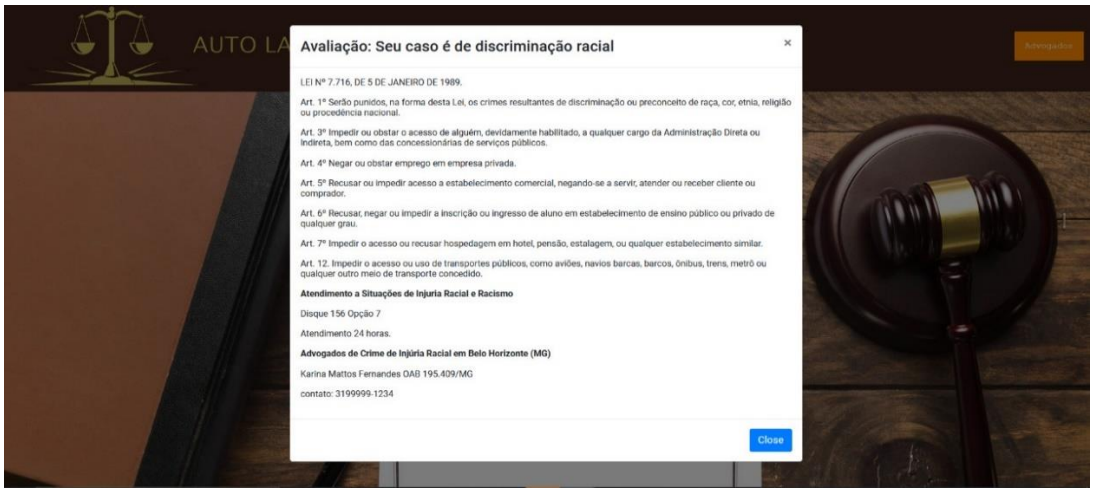
The first screen is where the user enters his demand or occurrence. There is a simple "Write your testimony" command (in Brazilian Portuguese) so that the user can perform the action.

On screens 2 and 3, the results of violence against women and racism are shown respectively. In this case, after the user enters their demand through screen 1, upon confirming their request, it is started the process of interpreting the situation via NLP and later the classification via SVM. After this classification process, there is the answer screen, which presents two options, demonstrating the law and its articles, in addition to a range of lawyers previously registered by the Autolaw administrative team.

Figure 2: Auto Law (Web Interface).

Screen 1 - Type the tesmonial

Screen 2 – Evaluation about the case related to Violence against woman



Screen 3 – devaluation of the case about racial discrimination

Source: Authors, 2020

3.1.2 NLP and Classification Structures Development

As already mentioned, the treatment of the user's demand is done by the NLP, for abstraction of the machine on the process that the user needs assistance and then a classification process via recommendation SVM. For initial tests, reports of violence against women and race prejudice were inserted to train the machine. Through the training process the machine would recognize whether the report proposed in the test would be Violence against women or Race prejudice (Racism).

As with any machine learning model, an existing set of examples (training set) was required for each category (class). The classification of sentences was considered to classify a sentence as "aggression" (violence against women) or "racism" (it is in Portuguese due to the target audience). In this case, there are two classes ("aggression" and "racism"). With the training set, a Naive Bayes classifier was trained to automatically categorize a new statement. An example, for demonstration, is in Figure 3 with a training set consisting of five testimonies of aggression and ten on racism.

Figure 3: Dataset Example.

	depoimento	categoria
0	“Ontem, ele [seu companheiro] me deu um soco n...	agressao
1	“Uma vez, ele falou que ia me dar umas facadas...	agressao
2	“As brigas vêm acontecendo constantemente. Há ...	agressao
3	“Ele me vê como se eu fosse um animal. Eu tenh...	agressao
4	“bom, virei estatística de novo. queria chamar...	agressao
5	“Eu comecei a apanhar sem motivos, por eu gost...	agressao
6	“Festa da agência que trabalhava em 2016. Um r...	racismo
7	“Já aconteceu de policiais pararem o carro do ...	racismo
8	“Eu branca, olhos claros. Meu pai, negro, cabe...	racismo
9	“Em 2011 eu trabalhava pra uma família como ba...	racismo
10	“Eu tava no salão agora, e chegou uma senhora ...	racismo
11	“Uma vez estava num restaurante com uns amigos...	racismo
12	“Na alfabetização, uma garota olhou pra mim e ...	racismo
13	“Uma vez estava com um amigo negro indo pra ca...	racismo
14	“Eu estava andando na rua e um menino olhou pr...	racismo
15	“…E o pior dia foi quando eu estava fazendo co...	racismo

Translation:

0 “Yesterday, he [companion] punched me ..” aggression

...

15 "... And the worst day was when I was doing ..." racism

Source: Authors, 2020

Thus, it is necessary to find out what is the probability of the class 'aggression' given a new statement and the probability of the class 'racism' given a new statement. Then the testimony is

divided into words and it is assumed that each word in a testimonial is independent of the others. The entire testimony is no longer evaluated, but individual words.

To find the total number of times a word appears in a class, CountVectorizer was used in sklearn. CountVectorizer provides the Term Document Matrix (TDM) for each class. A thermo-document matrix (TDM) consists of a list of word frequencies that appear in a set of documents. Then, the term document (TDM) matrix is calculated for the class "aggression", as shown in Figure 4. After the words are counted, each one is evaluated in terms of probability for validation via Nayve Bayes, as shown in Figure 5.

Figure 4: TDM Example.

	acontecendo	acontecido	acordado	acordei	...	vou	vulnerável	vê	vêm
0	0	0	0	0	...	0	0	0	0
1	0	0	1	1	...	0	0	0	0
2	1	0	0	0	...	0	0	0	1
3	0	0	0	0	...	0	0	1	0
4	0	1	0	0	...	1	1	0	0
5	0	0	0	0	...	0	0	0	0

[6 rows x 215 columns]

TDM Women aggression

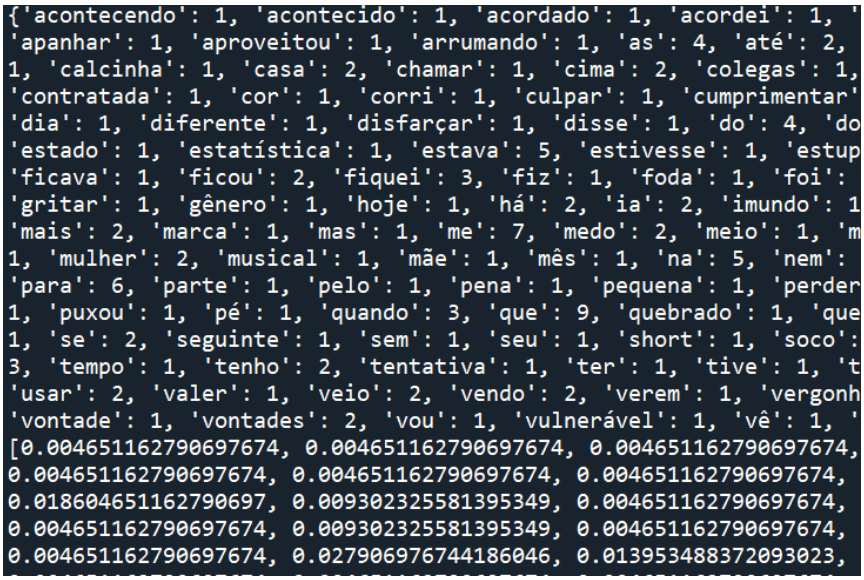
	2011	2016	acasalando	achavam	...	zoológico	época	éramos	única
0	0	1	0	0	...	0	0	0	0
1	0	0	0	0	...	0	0	0	0
2	0	0	0	1	...	0	0	0	0
3	1	0	0	0	...	0	0	0	0
4	0	0	0	0	...	0	0	0	0
5	0	0	1	0	...	1	1	1	0
6	0	0	0	0	...	0	0	0	1
7	0	0	0	0	...	0	0	0	0
8	0	0	0	0	...	0	0	0	0
9	0	0	0	0	...	0	0	0	0

[10 rows x 307 columns]

TDM Racism

Source: Authors, 2020

Figure 5: Terms Probability.



Source: Authors, 2020.

There is still one more aspect to be addressed. As the algorithm Naïve Bayes above involves probabilities for each word of a new testimony with respect to a class, if a word of the new testimony does not occur in the class within the training set, the equation will become zero. For example, the word "price" did not occur in the training set in any of the classes. Then $P(price \mid aggression) = 0$ and $P(price \mid racism) = 0$, which will cancel the P value. In summary, for the above P value to work, it is necessary all words from each new statement to appear in the training set, which is not possible. To solve this problem, Laplace Smoothing was used, where the value 1 is added to each word of the sentence.

With this apparatus, the initial dataset of the experiments was developed with a previous classification already analysed.

However, due to the number of variables, the dataset, for the classification, demanded an algorithm with greater processing capacity for a regression analysis. Then the SVM algorithm was implemented with the simple linear kernel for demand classification. In the following topic, accuracy experiments are demonstrated.

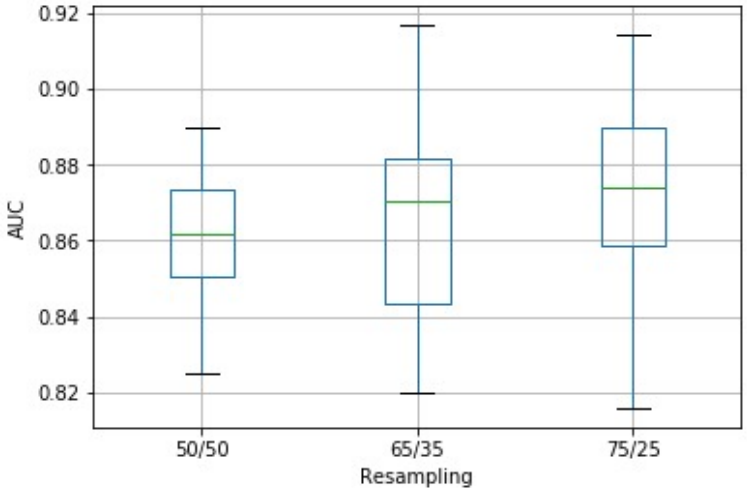
3.2 Accuracy Evaluation

During experiments with SVM, an experimentation process was carried out with 700 testimonials. The test was carried out with the separation of the base in 3 options, using the technique of test sets and validation as a reference. In this case, the separation was 50/50 (50% for testing and 50% for validation), 65/35 and 75/25. The accuracy was measured by the AUC obtained by the algorithm during the classification with 50 repetitions of each test.

Figure 6 shows the results that were obtained. It was noted that option 65/35 showed the highest classification value of AUC 0.92, however a wide dispersion of results. This option

showed a relevant average of approximately 0.875, however, from the graph it can be seen that the set of responses are far apart.

Figure 6: Accuracy Evaluation.



Source: Authors, 2020

.

The 50/50 option, on the other hand, demonstrated the most concentrated behaviour of all, despite the fact that it did not present an outstanding AUC or an average that could be more relevant than the others, that is, despite being more controlled, the accuracy results would not be the best to be obtained.

Finally, option 75/25 draws attention to issues such as: it presents the worst of the lowest results but shows the highest possible average. Based on that it has a worst AUC value of approximately 0.818 (less than the worst result in 65/35 of 0.82) but it has an average of approximately 0.89 (greater than the average of 65/35). However, it does not show the best possible result. Finally, its dispersion is not as high as that of the 65/35. Thus, 75/25 was the option maintained for the functionality.

4. Conclusion

Humanity has its problems, since its conception, where different thoughts or characteristics are not always accepted. This implies segregation and sad chapters in history. This work constitutes an initial contribution for the population that seeks to know their rights.

The Autolaw solution, focused on Brazilian laws and terms, but this solution proves to be very promising for other nations.

During the tests it was noticed that the machine learning algorithms had a considerable performance, however the behaviour of the accuracy tests shows that evaluations with other techniques besides SVM can be of great value for the optimization of accuracy.

Finally, simplicity of use was maintained, precisely so that it was accessible to all social classes.

References

[1] Dave, D. M., Friedson, A. I., Matsuzawa, K., Sabia, J. J. and Safford, S. (2020). Black Lives Matter protests, social distancing, and COVID-19 (No. w27408). National Bureau of Economic Research.

[2] Peterman, A., Potts, A., O'Donnell, M., Thompson, K., Shah, N., Oertelt-Prigione, S., and van Gelder, N. (2020). Pandemics and violence against women and children. Center for Global Development working paper, 528.

[3] Barros, R., Peres, A., Lorenzi, F., Wives, L. K., and da Silva Jaccottet, E. H. (2018, June). Case Law Analysis with Machine Learning in Brazilian Court. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 857-868). Springer, Cham.

[4] Medvedeva, M., Vols, M. and Wieling, M. (2019). Using machine learning to predict decisions of the European Court of Human Rights. Artificial Intelligence and Law, 1-30.

[5] Brasil (1989). Lei nº 7.716 de 5 de janeiro de 1989. Define os crimes resultantes de preconceito de raça ou de cor.

[6] Brasil (2006). Lei nº 11.340, de 7 de agosto de 2006. Diário Oficial da União 2006; 8 ago.

[7] Hidayat, A., Utomo, V. G. (2016). Adaptive Online Module Prototype for Learning Unified Modelling Language (UML). International Journal of Electrical and Computer Engineering, 6(6), 2931.

[8] Attaran, M. and Woods, J. (2019). Cloud computing technology: improving small business performance using the Internet. Journal of Small Business & Entrepreneurship, 31(6), 495-519.

[9] Hardeniya, N., Perkins, J., Chopra, D., Joshi, N. and Mathur, I. (2016). Natural language processing: python and NLTK. Packt Publishing Ltd.

[10] Alpaydin, E. (2020). Introduction to machine learning. MIT press.

[11] Santos P H J, Costa Junior C A D F, Rodrigues, D S S, Carvalho L M, Souza F H B (2020) Information Security and Machine Learning: Encryption Allocation Based on Recognizing of Text Patterns. *In Proceedings of International Joint Conference on Industrial Engineering and Operations Management- ABEPRO-ADINGOR-IISE-AIMASEM (IJCIEOM 2020)*