Indonesian Journal of Electrical Engineering and Computer Science
Vol. 26, No. 2, May 2022, pp. 1103~1108
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v26.i2.pp1103-1108

Application of machine learning algorithms for predicting outcomes of accident cases in Moroccan courts

Aissa Haidar, Tarik Ahajjam, Imad Zeroual, Yousef Farhaoui
L-STI, T-IDMS, Faculty of Science and Technics, University of Moulay Ismail, Errachidia, Morocco

Article Info

Article history:

Received Jan 6, 2022
Revised Feb 11, 2022
Accepted Mar 11, 2022

Keywords:

Court judgments
Decision making
Legal cases
Machine learning
Text classification

ABSTRACT

Due to the large number of legal cases, the processing of them by the courts is generally very slow. Among these cases, we find accidents cases, which require a great speed of judgment to compensate the victims of those accidents. To this end, we thought of exploiting the possibilities offered by machine learning in order to simulate the work of judges and contribute to speeding up the time of decision. Further, we applied different machine learning algorithms, such as linear regression, decision trees, and random forests. According to the results achieved, the Random Forest is the most perfect model for with the utmost accuracy about 91.05%.

Corresponding Author:

Aissa Haidar
L-STI, T-IDMS, Faculty of Science and Technics, University of Moulay Ismail
Errachidia, Morocco
Email: a.haidar@edu.umi.ac.ma

1. INTRODUCTION

Today world is very well connected and there are so many technologies emerging every day. Human beings are depending a lot on machines, robot and computers to mitigate their workload, and to make the tasks quick, easy and effective. In fact, artificial intelligence and machine learning have revolutionized the world because they provide users with intelligent systems and interfaces in many domains including civil engineering [1], [2], education [3]-[7], tourism [8]-[10], medicine [11]-[15] and also justice.

Morocco, like many developing countries, is currently converting from paper to digitization. This digital transformation is going ahead, but not at full speed. The country's Justice courts are among the public sector institutions that are aware of digital transformations. Almost 80,000 of road accident in Morocco in each year. It is obvious that victims of those accidents will wait for a kind of compensation. The courts in Morocco are still working in a traditional way and really put a huge psychological and mental pressure on those victims. Implement artificial Intelligence and machine learning in the field of justice will surely create a positive environment and make the legal cases be solved quickly and effectively.

This work's primary motivation is taking advantage of intelligence artificial to develop decisionmaking systems and facilitate such laborious tasks for different parts of justice. To this end, we have compiled and processed the Errachidia court data to build a model to predict the outcome of accident cases. We trained and evaluated the performance of three different machine learning algorithms namely linear regression, decision tree, and random forests. In addition to the introduction, this work is divided into four main sections. In section 2, we provide an overview of relevant works related to the topic of the paper. Section 3 describes the methodology followed to process the collected dataset. Further, we introduce the three machine learning algorithms implemented in this study. In section 4, we present the results obtained. Finally, we conclude this paper in section 5.

2. RELATED WORK

Generally, processing legal data to automatically retrieve valuable information has a vital role in the legal field [16]. Numerous works have dealt with legal data and the use of information technology in the field of justice. However, working on legal Arabic textual data is little compared to other languages especially if those data are from Moroccan justice institutions. To the best of our knowledge, very limited studies have been worked on similar data in Morocco (e.g., [17] and [18]). However, we built the first of its kind predictive model that predicts outcomes of legal cases in Moroccan courts. In this section, we shed light on relevant studies related to the topic of the paper.

A study, conducted by Medvedeva *et al.* [19], aimed to automatically process the court proceedings to predict the likely verdict of new cases. In this study, the data was compiled from the European Court of Human Rights (ECtHR), which is publicly available. The support vector machine (SVM) linear classifier was used as a supervised machine learning that has been trained on many court cases associated with their judgments. Besides, the model was trained to predict only two verdicts, violation vs. no violation. The prediction model was evaluated using 9 articles of the ECtHR and the accuracy achieved was 75%.

Wu *et al.* [20] attempted to recognize predictors of drug court graduation among amphetamineusing participants. To this end, the database used includes data of 540 participants where 341 are amphetamineusing. Moreover, the study used multivariate binary regression as predictive models, where chi-square and t-tests are performed to compare the outcomes and amphetamine-use groups. The results reported that having kids and the interaction of using amphetamine and being employed were predictive of graduation.

The use of artificial intelligence and machine learning techniques is increasing in the public sector. For instance, Lima and Delen [21] conducted a study to predict corruption-related issues inside government institutions. The study worked on massive datasets collected from several sources like the human development reports of the United Nations Development program. The final database comprised of information on 117 variables across 132 nations from different world regions (Americas, Asia Pacific, Europe, Central Asia, Middle East, North Africa, and Sub-Saharan Africa). Then, predictive models were built using popular machine learning algorithms namely SVM, artificial neural networks (ANN), and random forests. The results revealed that the random forests achieved the highest accuracy 85.77%, followed by SVM with 76.15%, and ANN with 73.84%.

Metsker *et al.* [22] dealt with Russian court decisions using machine learning. They used Spark for data processing and decision trees for analysis. They developed methods of extracting and structuring knowledge taking into account the specificities of the legislation of the Russian Federation.

Bozkir and Sezer [23] were able to predict the actual consumption in food demand. They did that using three main methods (CART, CHAID, and Microsoft Decision Trees). The study was done in food courts of Hacettepe University. And prediction accuracy was up to 0.83 in R2.

According to the study [24] they tried to implement technology in the legal practices. Because the traditional way of doing things, the legal practices takes so much time and procedures. So, the study developed what they call “virtual courtroom” to solve cases and problem that might occur.

Gomes *et al.* [25] investigated the effects of investments in information and communication technologies on the productivity of courts in Brazil. Organizations in the Brazilian justice system are seeking solutions to many of the challenges, such as limited access to justice and delays in resolving cases. The results confirm four of five hypotheses, showing that investment in information and communication technologies has a direct positive effect on the productivity of courts in Brazil.

Public data, once processed, can be used to improve society and policy making, but personal information must be removed from the data. Sharafat *et al.* [26] made an intelligent system to extract personal information from legal documents before mining them in searches. To automatically extract these entities, the first requirement is to construct a dataset using legal judgments. Thus, annotation guidelines are first prepared, followed by the preparation of an annotated dataset for extraction of various legal entities. Experiments with various datasets, several algorithms, and annotation schemes resulted in a maximum F1 score of 91.51% using conditional random fields.

3. METHOD

In this section, we describe our methodology which is composed of the following steps. First, extract the most representative characteristics. Second, divide the collected data into two parts (training, test). Third, start the learning process for each algorithm and build a model, and finally, test and evaluate each model built. Figure 1 describe the workflow of our research methodology.

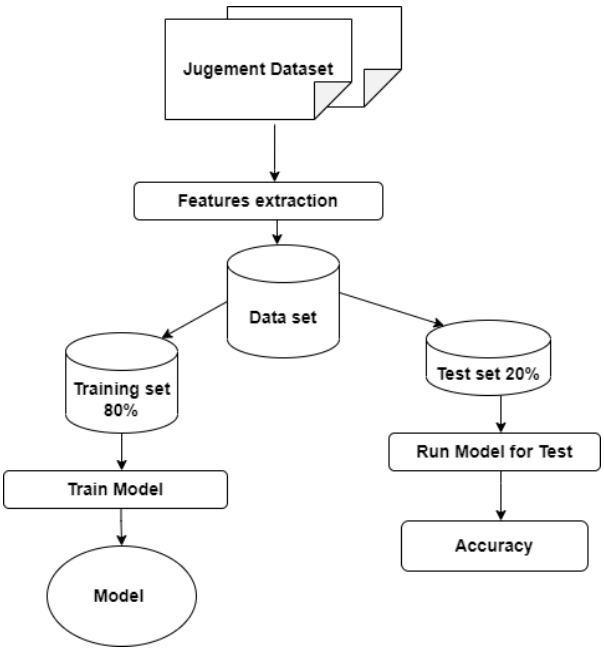

Figure 1. Research method

3.1. Description of the dataset

The dataset includes the judgments of Errachidia court in Morocco. The cases treated are accident cases. The judgments selected are from 2017 to 2019. Note that these judgments are written in Arabic because it is the official language of the courts in Morocco. Although the number of Arabic corpora available [6], none of them are similar to this dataset. Therefore, we believe that the current version and all updates to this dataset will be a valuable resource for computational linguistics and natural language processing alongside machine learning applications such as decision making.

3.2. Identification of characteristics

After having analyzed numerous judgments and treated the subject with different parts of the field as well as reviewing the codes and jurisprudence that organizes this type of cases. We have come to identify the most important characteristics that control the results of these judgments. These characteristics are presented in Table 1.

Table 1. Description of characteristics

	Charateristic	Description
1	Age	Age of the victim
2	Job	Employed, student, Not employed
3	Salary	Salary if he has a job, if not we use the minimum salary
4	Partial disability ratio	Ratio defined by expert
5	Total defict ratio	Ratio defined by expert
6	Physical pain	No important, Important…
7	Distoration of congential	No important, Important…

There are different criteria that control the prediction of the amount of compensation given to accident victims such as age, salary, disability rates defined by expert doctors and the rate of responsibility for each part of the accident. The outcome variable to be predicted is the amount of compensation for the victim of an accident. These seven characteristic variables are appropriate for predicting the final amount of compensation because each of them can modify this amount.

- Example 1: Two people have the same characteristics of salary, disability rates and one of them is very old, of course the amount of compensation is lower than the other who is younger.
- Example 2: Two people have the same age, salary but one of the disability rates is different, the amount is greater for the person with the highest disability rates.

So each variable of these seven variables can modify the final amount of compensation.

3.3. Algorithms used

As we have inputs and an exact result to predict, we are in a case of supervised learning. And as the compensation to predict is a continuous number, we treat a regression problem. For that we tried to work with three regression algorithms which are ”linear regression”, ”decision tree”, and ”random forest”.

3.3.1. Linear regression

Linear regression attempts to model the relationship between two variables by fitting a linear equation to the observed data. One variable is considered an explanatory variable and the other is considered a dependent variable. For example, a modeler may want to relate apartment prices to their area using a linear regression model. Before attempting to fit a linear model to the observed data, a modeler must first determine whether or not there is a relationship between the variables of interest. This does not necessarily imply that one variable causes the other, but that there is a significant association between the two variables. A linear regression line has an equation of the form $Y = a + bX$, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intersection (the value of y when x=0).

3.3.2. Decision tree

Decision tree learning is one of the predictive modeling approaches used in statistics, data mining, and machine learning. It uses a decision tree (as a predictive model) to go from observations on an item (represented in branches) to conclusions about the item’s target value (represented in leaves). Tree models in which the target variable can take a discrete set of values are called classification trees; in these trees, leaves represent class labels and branches represent conjunctions of entities that lead to those class labels. Decision trees in which the target variable can take continuous values (usually real numbers) are called regression trees. Decision trees are among the most popular machine learning algorithms due to their intelligibility and simplicity.

3.3.3. Random forests

Random forests or random decision forests is a set learning method for regression, classification, and other tasks that work by building a multitude of decision trees at the time of learning and pulling out the class which is the mode of classes (classification) or the mean/mean of the prediction (regression) of individual trees. Random decision forests correct the habit of over fitting decision trees to their training set. Random forests typically outperform decision trees, but their accuracy is lower than gradient-boosted trees. However, the characteristics of the data can affect their performance.

4. RESULTS AND DISCUSSION

It is very important to evaluate models in order to know which methods are the best. Indeed, there are several metrics that can be utilized to measure the performance of a regression models; among these criteria: R-squared, root mean squared error, residual standard error, mean absolute error. In our work, we utilize the first criteria in order to compare selected algorithms. Bellow a general definition of this metric:

R-squared (R2): is defined as (1 - U/V), where U is the residual sum of squares and V is the total sum of squares. The best possible score is 1.0 and it can be negative (because the model can be arbitrarily worse).

$$U=\sum_{i=1}^{N}(true_i - pred_i)^2 \quad (1)$$

$$V=\sum_{i=1}^{N}(true_i - mean(true))^2 \quad (2)$$

The test was carried out for each model on the same collected data and under the same conditions. Table 2 and Figure 2 show the accuracy value for each algorithm. The results show that the use of Machine Learning and exactly the algorithm of “random forest” in the field of justice in Morocco is very effective. This is proved by the high percentage (91.05%). If we compare this number to the other works and experiences, we can say that is can be higher or slow depending on the context and the circumstances. This work, as it is the first of its kind in Morocco, will greatly help judges in decision-making in a minimum time to predict compensation for victims of road accidents.

Table 2. Accuracy of algorithms

Algorithm	Accuracy
Linear regression	63.56%
Decision tree	82.23%
Random forests	91.05%

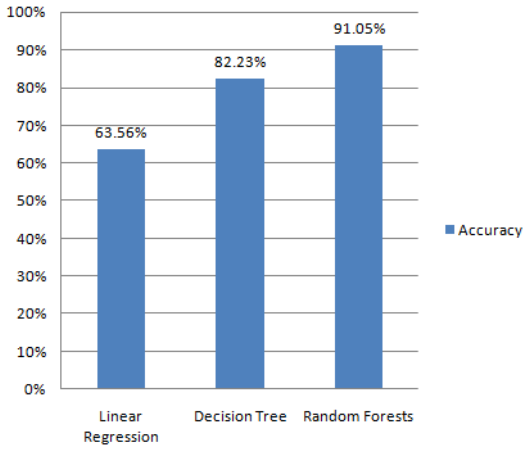

Figure 2. Research method

5. CONCLUSION AND PERSPECTIVES

In this paper, we have attempted to perform the first-of-its-kind study in Morocco by working on data from Errachidia Court Accident Cases to build a predictive model which will help us solve many court cases. The results revealed that the random forests model obtained the highest accuracy 91.05%, followed by decision tree with 82.23%, and linear regression with 63.56%. Other research and studies as it has been discussed earlier have indicated the efficacity of machine learning in the field of justice. We have other projects that we are going to work on in the field of justice in Morocco. But the focus is on other types of legal cases like cromes.

REFERENCES

[1] J. Naranjo-Pérez, M. Infantes, J. Fernando Jiménez-Alonso, and A. Sáez, "A collaborative machine learning-optimization algorithm to improve the finite element model updating of civil engineering structures," *Engineering Structures*, vol. 225, p. 111327, 2020, doi: 10.1016/j.engstruct.2020.111327.

[2] S. R. Vadyala, S. N. Betgeri, J. C. Matthews, and E. Matthews, "A review of physics-based machine learning in civil engineering," *Results in Engineering*, vol. 13, p. 100316, 2022, doi: 10.1016/j.rineng.2021.100316.

[3] O. Embarak, "A New paradigm through machine learning: a learning maximization approach for sustainable education," *Procedia Computer Science*, vol. 191, pp. 445-450, 2021, doi: 10.1016/j.procs.2021.07.055.

[4] A. Tarik, H. Aissa, and F. Yousef, "Artificial intelligence and machine learning to predict student performance during the COVID-19," *Procedia Computer Science*, vol. 184, pp. 835-840, 2021, doi: 10.1016/j.procs.2021.03.104.

[5] I. T. Sanusi, S. S. Oyelere, and J. O. Omidiora, "Exploring teachers' preconceptions of teaching machine learning in high school: a preliminary insight from Africa," *Computers and Education Open*, p. 100072, 2021, doi: 10.1016/j.caeo.2021.100072.

[6] H. Vartiainen, M. Tedre, and T. Valtonen, "Learning machine learning with very young children: Who is teaching whom?," *International Journal of Child-Computer Interaction*, vol. 25, p. 100182, 2020, doi: 10.1016/j.ijcci.2020.100182.

[7] C. S. A. Assuad, N. Tvenge, and K. Martinsen, "System dynamics modelling and learning factories for manufacturing systems education," *Procedia CIRP*, vol. 88, pp. 15-18, 2020, doi: 10.1016/j.procir.2020.05.003.

[8] J. Yu and R. Egger, "Color and engagement in touristic Instagram pictures: A machine learning approach," *Annals of Tourism Research*, vol. 89, p. 103204, 2021, doi: 10.1016/j.annals.2021.103204.

[9] Y. Luo, J. He, Y. Mou, J. Wang, and T. Liu, "Exploring China's 5A global geoparks through online tourism reviews: A mining model based on machine learning approach," *Tourism Management Perspectives*, vol. 37, p. 100769, 2021, doi: 10.1016/j.tmp.2020.100769.

[10] G. Xie, Y. Qian, and S. Wang, "Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach," *Tourism Management*, vol. 82, p. 104208, 2021, doi: 10.1016/j.tourman.2020.104208.

[11] S. de S. Santos, M. Torres, D. Galeano, M. del M. Sánchez, L. Cernuzzi, and A. Paccanaro, "Machine learning and network medicine approaches for drug repositioning for COVID-19," *Patterns*, p. 100396, 2021, doi: 10.1016/j.patter.2021.100396.

[12] A. Verma, V. C. Chitalia, S. S.Waikar, and V. B. Kolachalama, "Machine learning applications in nephrology: a bibliometric analysis comparing kidney studies to other medicine subspecialities", *Kidney Medicine*, vol. 3, no. 5, 2021, doi: 10.1016/j.xkme.2021.04.012.

[13] C. L. Brouwer *et al.*, "Machine learning applications in radiation oncology: Current use and needs to support clinical implementation," *Physics and Imaging in Radiation Oncology*, vol. 16, pp. 144-148, 2020, doi: 10.1016/j.phro.2020.11.002.

[14] E. Arslan, J. Schulz, and K. Rai, "Machine learning in epigenomics: Insights into cancer biology and medicine," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1876, no, 2, p. 188588, 2021, doi: 10.1016/j.bbcan.2021.188588.

[15] J. Wilkinson *et al.*, "Time to reality check the promises of machine learning-powered precision medicine," *The Lancet Digital Health*, vol. 2, no. 12, pp. 677-680, 2020, doi: 10.1016/S2589-7500(20)30200-4.

[16] T. Gonçalves and P. Quaresma, "Is linguistic information relevant for the classification of legal texts?," in *Proceedings of the 10th international conference on Artificial intelligence and law*, New York, NY, USA, 2005, pp. 168–176, doi: 10.1145/1165485.1165512.

[17] A. Y. Ikram and L. Chakir, "Arabic text classification in the legal domain," in 2019 Third *International Conference on Intelligent Computing in Data Sciences (ICDS)*, 2019, pp. 1-6, doi: 10.1109/ICDS47004.2019.8942343.

[18] H. Aissa, A. Tarik, I. Zeroual, and F. Yousef, "Using machine learning to predict outcomes of accident cases in moroccan courts," *Procedia Computer Science*, vol. 184, pp. 829-834, 2021, doi: 10.1016/j.procs.2021.03.103.

[19] M. Medvedeva, M. Vols, and M. Wieling, "Using machine learning to predict decisions of the European Court of Human Rights," *Artif Intell Law*, vol. 28, no. 2, pp. 237-266, 2020, doi: 10.1007/s10506-019-09255-y.
[20] L. J. Wu, S. J. Altshuler, R. A. Short, and J. M. Roll, "Predicting drug court outcome among amphetamine-using participants," *Journal of Substance Abuse Treatment*, vol. 42, no. 4, pp. 373-382, 2012, doi: 10.1016/j.jsat.2011.09.008.
[21] M. S. M. Lima and D. Delen, "Predicting and explaining corruption across countries: A machine learning approach," *Government Information Quarterly*, vol. 37, no. 1, p. 101407, 2020, doi: 10.1016/j.giq.2019.101407.
[22] O. Metsker, E. Trofimov, M. Petrov, and N. Butakov, "Russian court decisions data analysis using distributed computing and machine learning to improve lawmaking and law enforcement," *Procedia Computer Science*, vol. 156, pp. 264-273, 2019, doi: 10.1016/j.procs.2019.08.202.
[23] A. S. Bozkir and E. A. Sezer, "Predicting food demand in food courts by decision tree approaches," *Procedia Computer Science*, vol. 3, pp. 759-763, 2011, doi: 10.1016/j.procs.2010.12.125.
[24] T. K. Wah and M. a/p Muniandy, "Courtroom decision support system using case based reasoning," *Procedia - Social and Behavioral Sciences*, vol. 129, pp. 489-495, 2014, doi: 10.1016/j.sbspro.2014.03.705.
[25] A. O. Gomes, S. T. Alves, and J. T. Silva, "Effects of investment in information and communication technologies on productivity of courts in Brazil," *Government Information Quarterly*, vol. 35, no. 3, p. 480-490, 2018, doi: 10.1016/j.giq.2018.06.002.
[26] S. Sharafat, Z. Nasar, and S. W. Jaffry, "Data mining for smart legal systems," *Computers & Electrical Engineering*, vol. 78, pp. 328-342, 2019, doi: 10.1016/j.compeleceng.2019.07.017.

BIOGRAPHIES OF AUTHORS

Aissa Haidar is a PhD student at Faculty of Science and Techniques in Errachidia City, University of Moulay Ismail, Morocco with Master of Business Intelligence from University of Sultan Moulay Suliman, Beni Mellal, Morocco (2012). He obtained Bachelor Degree in Computer Science from University of Sidi Mohamed Ben Abdellah, Fes (Morocco) in 2010. His research is in fields of computer science, digital systems, and machine learning. Recently, machine learning application on legal cases has been tackled. He can be contacted at email: a.haidar@edu.umi.ac.ma.

Tarik Ahajjam is a Moroccan computer scientist, Data scientist, high school computer teacher and member of the L-STI T-IDMS team of the science and technology faculty of Errachdia. His research is in fields of computer science, digital systems and machine learning. Recently, machine learning application on education has been tackled. He can be contacted at email: t.ahajjam@edu.umi.ac.ma.

Prof. Dr. Imad Zeroual is currently an assistant professor with the department of computer science at Faculty of Sciences and Technics, Moulay Ismail University, Morroco. He holds a Ph.D. degree in Computer Science from Mohamed First University with specialization in artificial intelligence and Data Science. He primarily works on natural language processing, machine learning, information retrieval, and Language Teaching and Learning. He can be contacted at email: i.zeroual@umi.ac.ma.

Prof. Dr. Yousef Farhaoui is Professor at Moulay Ismail University, Faculty of sciences and Techniques, Morocco. Chair of IDMS Team, Director of STI laboratory. Local Publishing and Research Coordinator, Cambridge International Academics in United Kingdom. He obtained his Ph.D. degree in Computer Security from Ibn Zohr University of Science. His research interests include learning, e-learning, computer security, big data analytics, and business intelligence. Farhaoui has three books in computer science. He is a coordinator and member of the organizing committee and also a member of the scientific committee of several international congresses and is a member of various international associations. He has authored 6 Book and many Book Chapters with Reputed Publishers such as Springer and IGI. He is served as a Reviewer for IEEE, IET, Springer, Inderscience and Elsevier Journals. He is also the Guest Editor of many Journals with Wiley, Springer, and Inderscience. He has been the General Chair, Session Chair, and Panelist in Several Conferences. He is Senior Member of IEEE, IET, ACM and EAI Research Group. He can be contacted at email: y.farhaoui@fste.umi.ac.ma.