

Review Paper on Court Decisions Using NLP

¹Bhavesh Shah, ²Yash Nalawade, ³Prachi Gijare, ⁴Hrushikesh Akhade

¹Asst.Professor, ^{2,3,4}Final Year Student

Department of Computer Engineering,

Suman Ramesh Tulsiani Technical Campus FOE, Pune, India

Abstract: Developments in Machine learning and prior work in the science of judicial prediction, the proposed system constructs a model designed to predict the behavior of the Supreme Court of India using Natural Language Processing. Proposed System intends to work upon the court document containing all the decisions and proceedings. This system can be useful, for both lawyers and judges, as an assisting tool to rapidly identify cases and extract patterns which lead to certain decisions. NLP here identifies each word and relates the meaning. On the other hand, machine learning here experiences each case and acquires knowledge. To the best of our knowledge, several NLP tasks have been carried out on legal texts. However, the use of text classification to predict court rulings is an under-explored area.

Keywords: NLP, Machine Learning, Prediction

I. INTRODUCTION

There are a number of lawsuits that come to the court per day for justice but merely some are taken for solving in the least amount of time. It is undisputedly important to improve the performance of the Indian judicial system in order to ensure that decisions given by the Judicial system are accurate and coherent, and to be made in a short span of time[3]. Moreover, lawyers are required to study those lawsuits and represent their client in the court proceedings. The extensive study of law in support of client's lawsuit takes excessive time, and it is a tedious job for a lawyer. In the modern age of computerization; if the lawyer will be assisted by the computer-aided system to give suitable law applicable to the case, the efforts of any lawyer will be lesser. Contrary to the judge can also use this system for decision making. The outcome of this system will be solving court cases within a short span of time. Any plaintiff can get justice within a reasonable period. Time and money involved in court matters will be lesser. Machine Learning is a field of Artificial Intelligence that uses various statistical techniques that helps computer systems to learn from various datasets. Natural Language Processing (NLP) is one of the fields of artificial intelligence. NLP involves interactions between computer and human language i.e. (natural language). NLP deals particularly with programming computers to process and analyze a large amount of natural language data. In this paper, the proposed model presents a model which predicts decision making for the Indian Court by using NLP[2][5]. This problem is mainly considered as a binary classification problem[4][5]. Model for this prediction is given in section 2. A Support vector machine(SVM) algorithm is used for classification or regression problem [1][2]. A support vector machine is one of the supervised classifiers of Machine learning which is defined by separating hyperplane. Training data points are support vectors which define hyperplane for classification. The aim of the SVM is to construct a hyperplane or set of a hyperplane in N-dimensional space. A good classification is achieved by the hyperplane with the largest distance with a nearest training data point of any functional margin or class. An SVM is used for text classification. The task of the SVM is that the classification of natural language or text or hypertext into a fixed number of categories based on their content. Hence, this system will be used as an assisting tool for quickly identifying lawsuits and to extricate patterns which yield the final decision for the same.

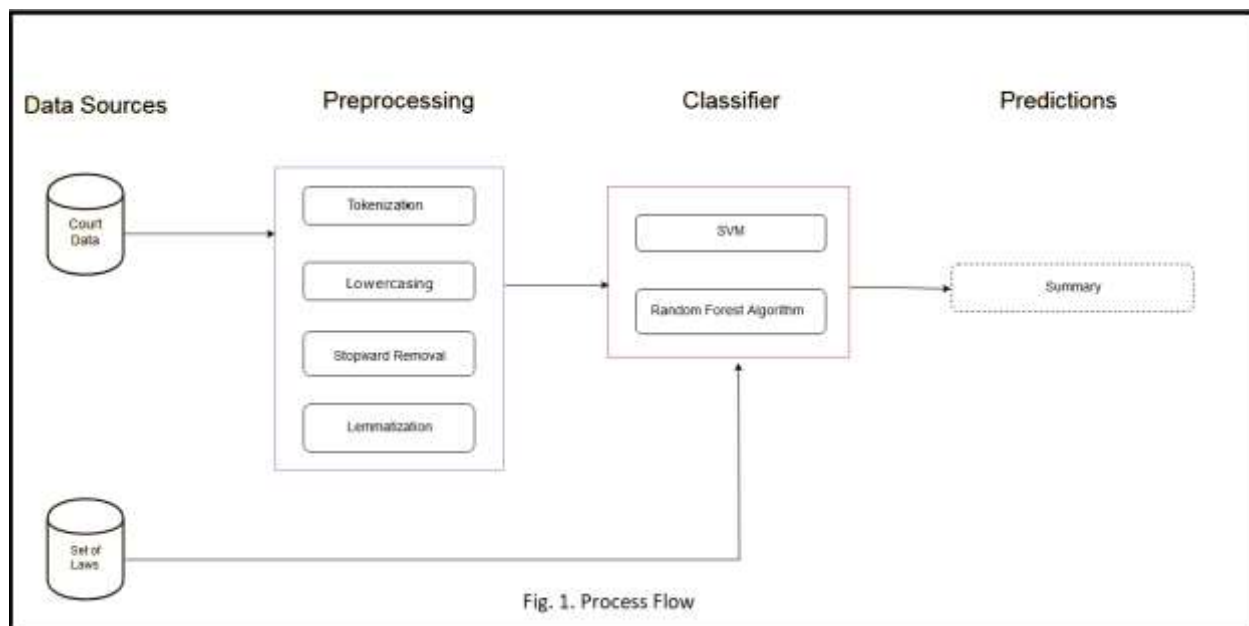
II. LITERATURE SURVEY

In [1], The authors proposed a system for prediction of court decisions. Text classification methods have been used in a wide range of NLP tasks. In [2], The authors have proposed a system which takes the Philippine Supreme Court case decisions, the only readily available online data as input data. They used Natural Language Processing, particularly the bag-of-words model to represent the case text into n-grams.

In[4], The authors used the extremely randomized tree, a method similar to the random forest approach. Applying the extremely randomized trees approach to each case from 1953-2013, their model correctly forecasts 69.7% of Case Outcomes and 70.9% of Justice Level Vote Outcomes over the sixty-year period which are close and better for the justice votes than on the cases themselves. In[5], The authors formulate a binary classification task where the input of their classifiers is the textual content extracted from a case and the target output is the actual judgment as to whether there has been a violation of an article of the convention of human rights.

For development in Machine learning and prior work in the science of judicial prediction, proposed system is going to implement a system that emulates the behavior of the Supreme Court of India moreover, predicts decisions of lawsuits of the Indian Court using only the textual information extracted from relevant resources by using Natural language processing (NLP)[2]. The problem of predicting the decisions is defined as a binary classification task. Our goal is to predict if, in the context of a particular lawsuit, there is a violation or non-violation in conformity with laws[4].

The proposed system intends to work upon the transcript that usually contains a record of all decisions of the judge and the spoken arguments by the litigants' lawyers. The output is the summary or the actual decision to be made by the judges and area of the law. Proposed system prefers the model to have a consistent performance so it intends to train the application over time using different cases and conditions.



The above diagram depicts an explanation of the process flow. The input sources are the textual data(transcript). The various preprocessing algorithms that have been used here are tokenization, lowercasing, stop word removal and lemmatization.

Tokenization:

The process of breaking sequence of strings data into the smallest unit possible. In this process, some characters like punctuation marks are discarded. The tokens become the input for another process like parsing and text mining.

Lowercasing:

Formatting data in lowercase letters. This process will simplify the searching process. Although lowercasing can decrease the reliability of searches in some cases.

For example, MIT is an acronym which when lowercased becomes mit which is a word in German.

Stop Word Removal:

The process of removing useless words such as “the”, “a”, “an”, “in”. In natural language processing, useless words (data), are referred to as stop words. These words consume space and increase processing time. The search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query.

Lemmatization:

Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma. For lemmatization it necessary to have detailed dictionaries which the algorithm can look through to link the form back to its lemma.

Form	Morphological Information	Lemma
studies	Third person, singular number, present tense of the verb study	study
studying	Gerund of the verb study	study
niñas	Feminine gender, plural number of the noun niño	niño
niñez	Singular number of the noun niñez	niñez

Table: Example for Lemmatization

Natural Language Processing :

Regular expressions can be used to remove punctuation marks, numbers, and any other characters that were not letters [2]. Parts-of-speech (POS) can be used to tag the part of speech of the remaining tokens. Also, remove proper nouns and stop words from the case decisions. Stop words are frequent words which do not add any significant meaning to the text [2]. The NLTK library has a corpus of proper nouns and English stop words [6].

III. CONCLUSION

The proposed system conducts an experiment on Indian Supreme Court case decisions, and investigated the application of text classification method to legal texts from Indian Supreme Court and applied Natural Language Processing approaches to predict the case outcome and the area of the law. To the best of our knowledge, several NLP tasks have been carried out on legal texts. However, the use of text classification to predict court rulings is still an under-explored area.

Text classification techniques can be used to provide valuable assistive technology for law professionals in obtaining guidance and orientation from large corpora of previous court rulings of Indian Supreme Court. Also, the semi-automation of case classification task will aid the Indian Judiciary in performing their duties.

REFERENCES

- [1] Predicting the Law Area and Decisions of French Supreme Court Cases 4th August 2017.
- [2] Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning July 2018.
- [3] Performance Evaluation of The Judicial System in Taiwan Using Data Envelopment Analysis and Decision Trees 10 April 2010.
- [4] Predicting the Behavior of the Supreme Court of the United States: A General Approach 12th April 2017.
- [5] Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective October 2016.
- [6] S. Bird, E. Klein, and E. Loper. Natural Language Processing with Python. O'Reilly Media, 2009.