Indian Penal Code Recognition Using Multiclass Classification Algorithms in Machine Learning

Prakash Aryan
Department of Computer Science and Engineering
Purnima University
Jaipur, Rajasthan

Abstract:- **As of May 2022, over 4.7 crore cases are pending in courts across different levels of the judiciary. This has resulted in a ponderous burden on the Indian judiciary system. According to Experts the reason why we have this problem is due to lack of infrastructure in the Indian Judiciary System. With this study I build a machine learning model that can determine the applicable Indian Penal Code according to the description provided by the user. I have used two machine learning approaches, one vs rest and one vs one multi-class classification. The one vs rest classification using Logistic Regression has a 99% accuracy on training data.**

Keywords:- Machine Learning, Natural Language Processing, Multiclass Classification.

I. INTRODUCTION

Between 2015 and 2019, the average number of pending cases was approximately 1.8 million per year, but the High Court currently lists 5.8 million pending cases. In most cases, the number of cases closed is lower than the number of cases opened, which continues to exacerbate the problem.[1] This has resulted in a huge backlog of cases and has in turn resulted in delay of justice. Experts have looked into this problem and determined the various reasons as to why this is happening. The foremost reason is the pending vacancies in the judiciary sector. One of the ways we can solve this is by encouraging digital courts in our country. Raj said there is a need to "expand and strengthen the system of digital courts, i.e. online courts, so that majority of the cases can be decided through the online system, which will popularize the judiciary, decentralize it and expedite the process of adjudication, across the country."[2]

Justice has often been delayed for people even for the fundamental rights. A lot of farmers have been affected due to this slump in our judiciary system. People have to be released after 20 to 30 years because enough evidence could not be gathered to acquit them of charges. Then we come across cases where people have waited years and even passed away after which their descendants had to fight the cases on their behalf and they were granted justice after 20 years. Such incidence puts the trust of people on democracy on a thin line and people lose hope.

With the digitalization of the Indian Judiciary the process of judgment can be enhanced and greatly reduced. The Indian Penal Code is the official penal code of the Republic of India. It is a complete code designed to cover all aspects of criminal law. The draft of the Indian Penal Code was prepared by the First Judicial Council, chaired by Thomas Babington Macaulay. The draft was based on a simple codification of English law, but at the same time borrowed elements from the Napoleonic Code and the Louisiana Civil Code of 1825.The IPC has been successful by and large in its attempt to prosecute and punish individuals who commit the crimes that are defined in this Code, but like Sedition there have been certain other provisions that have invited scrutiny time and again. IPC as a statute has survived and flourished over the last 160 years, which speaks volumes about its effectiveness as a penal code of high stature. However, over the years he has failed to get rid of some positions that smell of colonialism, for example. Rebellion. The Malimat Commission's report advocating criminal justice reform provided an opportunity for Congress to update the Code and other criminal laws. Seventeen years have passed since the report was submitted, but no concrete action has been taken in this regard. It's time for the legislature to intervene to make the code more modern than the British colonies. High courts stepping in and breaking the law does not reflect well on the legislature. Because that's the legislature's job in the first place.[3]

In this study, I aim to reduce the time taken on determining the applicable Indian Penal Code and the confusion around it. For a common man finding the appropriate section that is applicable in a given situation is very complicated and the person will have to seek a legal consultation which costs a lot both financially and timely. With this model we can predict the applicable IPCs in a particular situation based on the description provided by the user. This can help an individual get an idea of the seriousness of the situation and where do they stand in the entirety. The problem involves classifying the correct IPC applicable based on the user input. For this we will be using two of the machine learning algorithms which are multi class classification algorithm. One vs Rest and One vs One is used and the accuracy for both of these models are compared on the training set.

II. RELATED WORK

The use of machine learning in the field of legislation has been on the uprise. This is because it has made the process of legislation very simple for both the working professionals (lawyers and judges) and for the general public. Application of artificial intelligence in the legal world various areas of AI in the legal sector can be discovered with adequate research and careful comprehension of the legal industry by IoT app development companies. The current Artificial Intelligence applications in the industry can be categorized into six main parts:

➢ *Due Diligence:*
Lawyers use Artificial Intelligence tools to perform due diligence and uncover background information. In light of the current scenario, developers have opted to integrate a slew of new features, including agreement review, legal inquiry, and electronic media for this section of the industry.

➢ *Prognostication Technology:*
Artificial Intelligence (AI) aids in the generation of outcomes for legal investigations and agreement evaluations. This characteristic of AI programming appears to be extremely beneficial to legal firms and industries.

➢ *Legal Mechanism:*
Lawyers can use artificial intelligence technology to obtain information from previous or past instances. This data can also be used to track referee instructions and predictions. This technology will become increasingly important worldwide in the near future.

➢ *Documentation Mechanisms:*
Various types of software are used in the legal industry to develop documents that support the collection of data and information. There are many useful documents in the field of law firms.

As a result, it is really profitable.

➢ *Intellectual Property:*
AI algorithms show lawyers how to sift through massive IP files and extract meaning from a multitude of powerful texts.

➢ *Electronic Receipts:*
For a long time, lawyers created their own receipts. After these companies adopted AI software development technology, lawyers' accounts went electronic.[4]

These systems and the companies offering them are reshaping relationships between lawyers and clients, introducing new kinds of professionals into legal practice, altering the discovery process, and shaping how lawyers construct knowledge about their cases and professional obligations. In the midst of these shifting relationships—and the ways in which these systems are shaping the construction and presentation of knowledge—lawyers are grappling with their professional obligations, ethical duties, and what it means for the future of legal practice. Through in-depth, semi-structured interviews of experts in the e-discovery technology space—the technology company representatives who develop and sell such systems to law firms and the legal professionals who decide whether and how to use them in practice—we shed light on the organizational structures, professional rules and norms, and technical system properties that are shaping and being reshaped by predictive coding systems. [5]

In February 2018, an artificial intelligence system developed by Israeli legal technology firm lawgeex won more than 20 leading lawyers with rich experience in a standard business contract review competition. In addition, the average accuracy is more than 9% higher than that of humans, and in the whole process, the system that needs to complete tasks that humans need to complete in 92 minutes or less completes in 26 seconds, so speed and efficiency, he has committed "double murder" of people. [6]

III. MATERIALS AND METHODS

A detailed overview of how the project is carried out has been discussed in this section. Various important parts of building a machine learning model such as Data acquisition, Data preprocessing, feature selections and importance are also described.

➢ *Software Requirements*
- Operating System -Pop!_OS, Release –21.04
- Python 3.8.3
- Jupyter NotebookVersion –6.0.3

➢ *Hardware Requirements*
- CPU –Ryzen 5 3550H Quad core
- RAM –8gb ddr4
- Storage-20gb

➢ *Dataset*
We have used the Indian Penal Code dataset. To train our model better we have used 200 selected sections so that we can first test our model according to which algorithms suits it the most. The sections are up to 301. [7]

➢ *Algorithms*
Machine Learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without being explicitly programmed. Learning algorithms in numerous operations that's we make use of diurnal. Every time a web hunt machine like Google is used to search the internet, one of the reasons that work so well is because a literacy algorithm that has learned how to rank web runners. These algorithms are used for colorful purposes like data mining, image processing, prophetic analytics, etc. to name a many. The main advantage of using machine learning is that, once an algorithm learns what to do with data, it can do its

work automatically. In this paper, a brief review and unborn prospect of the vast operations of machine learning algorithms has been made.[8]

One vs. all provides a way to work double bracket. Given a bracket problem with N possible results, a one-vs. - All result consists of N separate double classifiers — one double classifier for each possible outgrowth. During training, the model runs through a sequence of double classifiers, training each to answer a separate bracket question [9]

Some classification algorithms permit the use of further than two classes by design. Others circumscribe the possible issues to one of two values (a binary, or two- class model). But indeed double bracket algorithms can be acclimated for multi-class bracket tasks through a variety of strategies. This element implements the one- versus- bone System, in which a double model is created per class brace. At vaticinator time, the class which entered the most votes is named. Since it requires to fit n_classes *(n_classes- 1)/ 2 classifiers, this system is generally slower than one- versus- all, due to its O (n_classes2) complexity. Still, this system may be profitable for algorithms similar as kernel algorithms which do n't gauge well withn_samples. This is because each individual literacy problem only involves a small subset of the data whereas, with one- versus- all, the complete dataset is used n_classes times.[10]

Logistic Regression is a classification algorithm that performs well for linearly separable classes. It is designed for two class classification but with the help of One-vs-Rest (OvR) classifier it can handle multi-class problem also. The output of logistic regression is binary dichotomous.[11]

Support Vector Machines are classification algorithms that are based on determining vectors that in turn creates a hyperplane. The classes are divided based on this hyperplane. [12]

Natural Language Processing - Natural languages are those languages that are spoken by the people. Natural language processing cinctures everything a computer needs to understand natural language and also generates natural language. Natural Language Processing is a subfield of Artificial Intelligence and verbal, devoted to make computers understand the statements or words written in mortal languages. A Natural language also known as ordinary language that's spoken or written by people (humans) for general purpose communication. Natural language came into actuality because when stoner wishes to communicate with the computer we can't force the druggies to learn machine specific language so this principally caters to directors or children's who don't have enough time to learn new specific languages or get professed in them. Languages can be any like Hindi, French, English, and Chinese etc. A language is a system, a set of rules or set of symbols. [13]

Accuracy of a model is one of the most used evaluation technique. It tells us how accurate our model is in making predictions. Accuracy is calculated by – Correct predictions/ Total predictions.

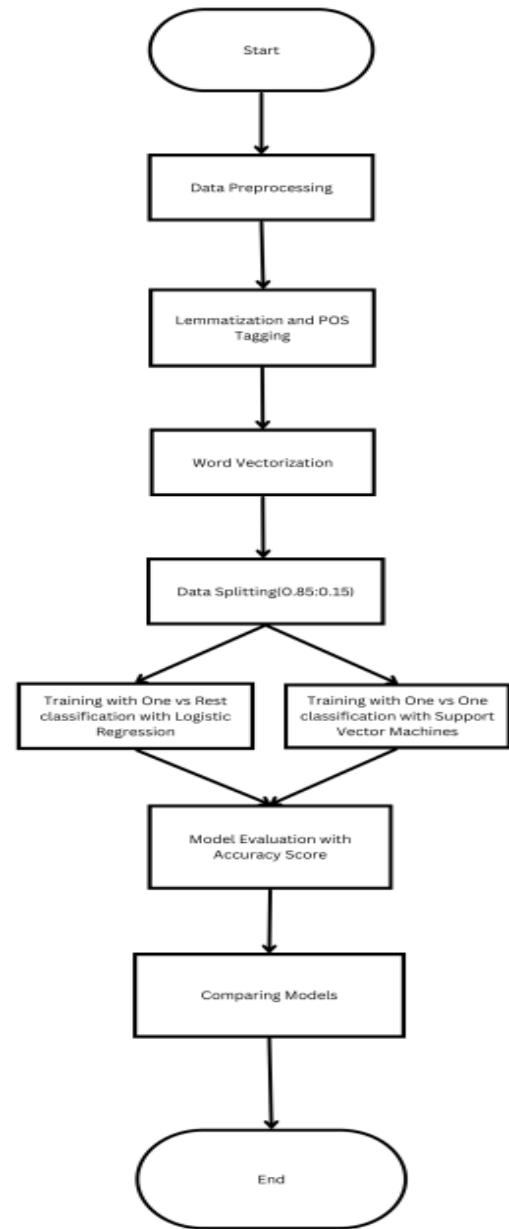$$\textbf{Accuracy} = \frac{TP+TN}{TP + TN + FP + FN}$$

Fig – 1: Flowchart of the Model Implementation

IV. RESULTS

We have used accuracy parameter to check performance of our model. This is done on the training set of our dataset. We observed that using one vs one classification using support vector machines gave us an accuracy score of 97% and when we used one vs rest with logistic regression we got an accuracy score of 93%. This is due to the fact that each data point on our dataset is comparable to each and every other dataset and cannot be grouped together. We had to vectorize our words and sentences into arrays because machine learning needs numerical data for processing.

V. CONCLUSION AND DISCUSSION

With this model we saw that using the correct methodology and the right dataset can aid the general public and the people working in legislation. The dataset which we have is not very sophisticated and thus we could not check accuracy on trainset. This is one of the potential we can work on. We can include the entire dataset of the Indian Penal Code instead of selective sections. We can also include more description of each dataset so that our model does not underfit.

REFERENCES

[1]. "How to Start Resolving the Indian Judiciary's Long-Running Case Backlog - Carnegie Endowment for International Peace." https://carnegieendowment.org/2021/09/09/how-to-start-resolving-indian-judiciary-s-long-running-case-backlog-pub-85296 (accessed Feb. 24, 2023).

[2]. "India's huge backlog of court cases delays justice: Experts⊞." https://www.aa.com.tr/en/asia-pacific/indias-huge-backlog-of-court-cases-delays-justice-experts/2638784 (accessed Feb. 24, 2023).

[3]. "Indian Penal Code (IPC) - History, Structure & Recent Developments⊞." https://byjus.com/free-ias-prep/indian-penal-code/ (accessed Feb. 27, 2023).

[4]. "Role of artificial intelligence in law - iPleaders." https://blog.ipleaders.in/role-of-artificial-intelligence-in-law/ (accessed Feb. 27, 2023).

[5]. "AUTOMATED DECISION SUPPORT TECHNOLOGIES AND THE LEGAL PROFESSION on JSTOR." https://www.jstor.org/stable/26954398 (accessed Feb. 27, 2023).

[6]. T. Hu and H. Lu, "Study on the Influence of Artificial Intelligence on Legal Profession," Jan. 2020, doi: 10.2991/AEBMR.K.191225.184.

[7]. "THE INDIAN PENAL CODE".

[8]. "Machine Learning Algorithms-A Review," 2019, doi: 10.21275/ART20203995.

[9]. "Multi-Class Neural Networks: One vs. All | Machine Learning | Google Developers." https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/one-vs-all (accessed Feb. 27, 2023).

[10]. "One-vs-One Multiclass - Azure Machine Learning | Microsoft Learn⊞." https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/one-vs-one-multiclass (accessed Feb. 28, 2023).

[11]. D. W. Hosmer and S. Lemeshow, "Applied Logistic Regression," Applied Logistic Regression, Jan. 2005, doi: 10.1002/0471722146.

[12]. A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Stat Comput, vol. 14, no. 3, pp. 199–222, Aug. 2004, doi: 10.1023/B:STCO.0000035301.49549.88.

[13]. "https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=eeace1d14e266a5cd44fe781a874c662928602fd⊞." https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=eeace1d14e266a5cd44fe781a874c662928602fd (accessed Feb. 27, 2023).

Author –
This Paper is submitted by Mr. Prakash Aryan. He is an undergraduate student pursuing his course in computer science and engineering at Poornima University Jaipur. His interest lies in Machine Learning, Natural Language Processing and Computer Vision.