Intelligent Legal Advisor

Ankit Barsainya
Senior Associate, Cognizant Technology Solutions Pvt Ltd
Pune, India
Ankit.Barsainya@cognizant.com

Manish Naik
Senior Associate, Cognizant Technology Solutions Pvt Ltd
Pune, India
Manish.Naik@cognizant.com

Himmat Patil
Senior Associate, Cognizant Technology Solutions Pvt Ltd
Pune, India
Himmat.Patil@cognizant.com

ABSTRACT

Citizens are usually unaware of the laws, procedures and amendments (e.g. in case of domestic LPG cylinder explosion, consumers are entitled up to 40lakh) hence, they don't have clarity on the strength of their case. Also, there is limited knowledge on expected expenditure, procedures and the associated hassles (Court cases often take ~10 years) before filing a case. This could potentially disrupt an individual's life.

In this paper, we propose Artificial Intelligence based solution that provides realistic approximation of all the expenses and procedures that needs to be followed and the timeline for judgement (more than 30 million cases are still pending) along with the chances of favorable ruling. Such a system will be beneficial in multiple ways. First, Citizens will be well aware of the processes and procedures involved before moving to court, reducing time wasted owing to procedural illiteracy. Second, drastic reduction in the manual research times. Third, cost savings for legal counsel in the initial stages. Fourth, and the most important benefit, end user can prepare themselves mentally and monetarily keeping in mind the nature of the case and expected timeline for resolution. An indirect benefit will be reduction of the burden of Courts as citizens might refrain from filling unnecessary cases and also complete documentation before moving the court rather than moving the court and then looking for relevant documents.

Using Machine learning, we need to build a model that can refer to all the relevant laws, latest amendments, regulations along with all previous court cases that are relevant to current one. This model will study all the above documents, case history and come up with key pointers that had the highest impact on the final judgement. This will be tallied against the key points in current case to predict the case outcome hence enabling the citizens to help with DSS (decision support system) that will reduce logger head of court proceedings. The proposed solution will be to build a dashboard with a 360-degree view of legal proceedings. It will also incorporate chat-bot to help the end user with relevant queries.

Data will be ingested by retrieving historical cases from lower/high/supreme courts and central repository for law maintained by Govt. of India. This information will be consolidated, parsed and stored in a local repository. Machine learning algorithms like Clustering, NLP & Neural Networks will be implemented on this information to predict the outcome.

Keywords: NLP, Machine Learning, Artificial Intelligence, Decision Support System, Indian Law, Text Analytics

INTRODUCTION

According to data available on government websites and various surveys done over the years, roughly 28 million court cases are in a pending status across different courts in India. More than half of these 28 million cases are pending for at least more than 2 years, along with more than 8% have been unresolved for more than 10 years. These numbers keep increasing every year.

SUMMARY REPORT OF INDIA AS ON DATE :-- 13/07/2018

PARTICULARS	CIVIL CASES	CRIMINAL CASES	TOTAL CASES	PERCENTAGE
Pending Cases				
Cases Pending over 10 years	595385	1689325	2284710	(8.26%)
Cases Pending (Between 5 to 10 years)	1236937	3192301	4429238	(16.02%)
Cases Pending (Between 2 to 5 years)	2544192	5384431	7928623	(28.68%)
Cases Pending less than 2 years	3946132	9055829	13001961	(47.03%)
Total Pending Cases	**8322646**	**19321887**	**27644533**	**(100%)**

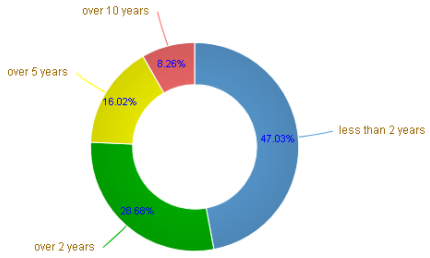

Fig 1. Total Pending cases

Below is the breakup of the approximate number of cases pending in Supreme Court, High Courts and District and Subordinate courts in India.

- Around 60,000 cases are pending in Supreme Court.
- Around 42 lakh cases are pending in different High Courts.
- Around 2.7 crore cases pending in District and Sub-ordinate Courts.

Also various survey suggests that at least 5 crore cases are filed every year and judges dispose of only 2 crore cases.

On the appellant's part, a lot of these delays can be attributed to lack of knowledge of legal procedures, required documentation and processes. Additionally, a lot of these appellants file court cases without a proper understanding of the strength of their case in terms of legal validity. Often, appellants do not have a clear idea of the timelines for resolution or the monetary implications. This leads to wastage of potential productivity and in-efficiency of the judicial system. According to a report by DakshIndia.org, delays in court case resolution cause a loss of ~.5% of India's GDP in terms of productivity, which amounts to 11.3 billion dollars every year.

A simplistic solution is required to solve/reduce above mentioned issues and help Indian Judicial system function efficiently.

OBJECTIVE

In order to solve above mentioned issues we plan to build a Machine Learning/Artificial Intelligence based solution. This solution will take into account all the relevant historical case details, all the landmark judgments that pertains to the potential case at hand, details of which will be provided by the end user (potential appellant), along with the applicable laws and compatibility with the Constitution. Armed with this information, end user can make an educated choice on whether it's worth moving to court or make an out-of-court settlement. This solution will be supported by chat-bot for query resolution as well.

Upon processing of this information, the model will come up with details, like the potential

outcome of the case, along with the documents required, a realistic approximation of timeline for resolution, and the basic procedures that needs to be followed before and during the court case.

This solution will also provide a list of relevant laws, few of the closest matching cases and the pertinent landmark cases for further analysis by the appellant.

This will result in the common man being aware of the strength of their case beforehand, along with the documentation required and potential timeline for resolution which in turn helps the judiciary dispose-off cases in quick time.

RELEVANT DATASETS

Details of various cases, laws, constitution will be retrieved from the websites of various courts and the Government of India websites.

Below diagram depicts these different datasets used by the model to analyze and forecast judgment and further details on this also discussed below.

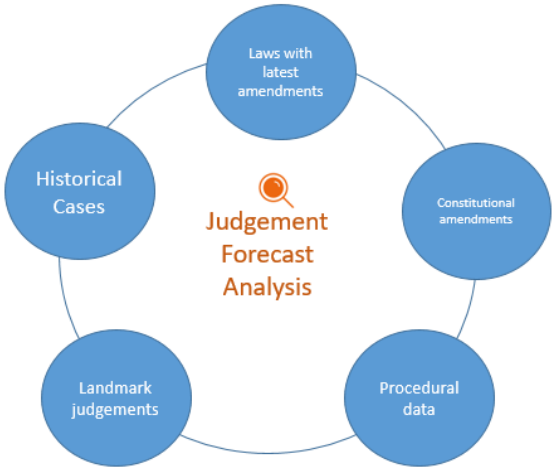

Fig 2. Datasets

- **Historical cases**
  This dataset will contain all the available historical cases and their pertinent details. This will be utilized as a basic training dataset to the model.

- **Landmark judgments**
  This will contain all landmark judgments which will hold a higher weightage. Also historical cases past such judgments will carry more weight. Timelines will be split on such cases using weightage calculation.

  e.g. For divorce related cases, last year Supreme Court delivered landmark judgment that cooling/waiting period of 6 months for passing decree of divorce isn't mandatory & can be waived by the Courts based on the facts & circumstances of the case. So accordingly the forecast of timeline to deliver judgment will be calculated based on the landmark judgment.

- **Laws with latest amendments**
  This will contain landmark changes to a law will hold even more weightage as it represents a paradigm shift.

- **Constitutional amendments**
  This will contain latest amendment in the constitution and will have even bigger impact on outcome of succeeding cases. Hence cases post amendments will have the highest weightage.

- **Procedural Data**
  Procedural information will be captured by crawling various legal advisory websites.

METHODOLOGY

Below is the model for proposed solution with more details.

Architecture of the proposed solution

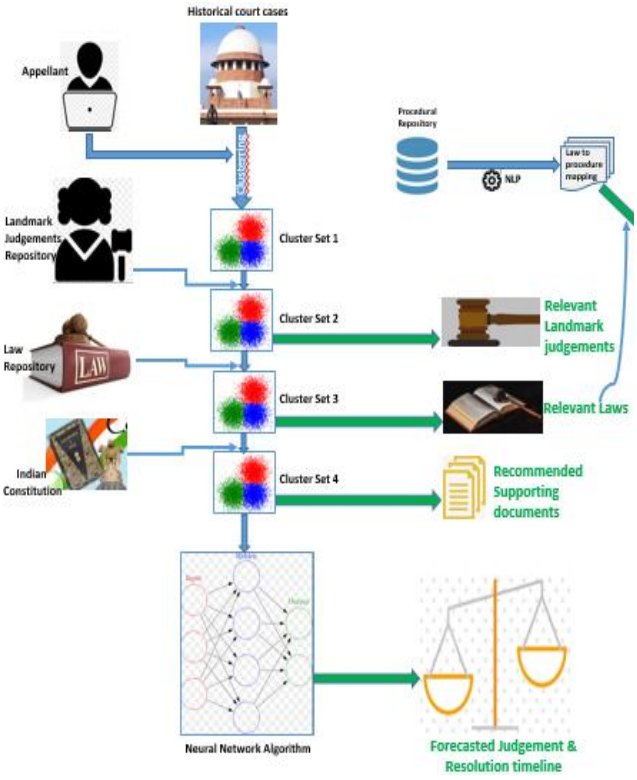

Fig 3. Architecture diagram

End user (Appellant) will provide the relevant case details using a web-based, interactive Q&A system. These details will be used as a seed for the basis of clustering of all the historical cases & filtering only the relevant cases for further analysis.

Dataset from above clustering's output (Cluster Set 1) will again be clustered & divided into mini clusters (Cluster Set 2) based on the judgments happening before and after landmark judgments. Landmark judgments will be used here as seed for clustering of Cluster Set 2. Thus creating a division of timeline and enabling calculation of weightage of all relevant historical cases. Data from cluster set 2 will be used to provide appellant details of 2 most recent and relevant landmark judgments.

Similarly, changes to law and constitutional amendments will be used as seed which will cause a split of timeline into smaller clusters (Cluster Sets 3 & 4) and weightage tuning for the final processing via a Neural Networking Model.

Dataset from Cluster Set 4 will be provided to a Neural Network along with the calculated initial weights for each sub-cluster. These weights will be further tuned using Backward Propagation to achieve higher accuracy.

This Neural Network will analyze the resolution timelines, documents, outcomes for each case and provide a reasonably accurate prediction of the outcome, approximate timeline of resolution and relevant documents required for case processing.

Additional Detail for Procedural Information

Using Natural Language Processing (NLP) on top of procedural dataset, each procedure will be mapped with its relevant law, thus generating a comprehensive dataset of law to procedure mapping.

The list of relevant laws, generated by cluster set 3, will be used as seed for clustering of procedural information and will generate new dataset (cluster set 5) of relevant procedures for further assistance to appellant/end users.

Experiment Details

Dataset & Model

In order to build our model, we obtained training dataset of 8,700 court cases which includes 600 types of petitioner, 600 types of respondent, 115 justices, and period of 60 years. It includes identification variable (case id, docket id, vote id etc.), background variables (case name, petitioner state, respondent state, origin of case), chronological variables (date of decision, term of court, natural court, chief justice, date of argument etc.), substantive variables (issue, issue id, decision descent direction), voting & opinion variables (vote not clearly specified, split vote, majority vote, minority vote, justice name etc.) and outcome variables (decision type,

declaration of unconstitutionality, disposition of case, unusual disposition, winning party etc.)

We have developed time evolving random forest classifier which leverages above features to predict more than 78,000 justice votes and 8,700 case outcomes. On data spanning over 60 years, we achieved **68%** of accuracy at case level outcome.

We converted all categorical variables into binary/indicator variables e.g. case disposition corresponds to DIG, no change in precedent or change in percent.

Random forest classifier model begins with selection of fixed term (e.g. 5 years) of data and trained the model over period of terms. To evaluate final prediction, we calculated prediction of each individual tree and then averaged it across entire forest. In our growing approach, it need train a small number of trees and most of the trees in the forest are stable for most years.

Dummy window = 10 years

Training years = 25

Trees per year = 5

Initial trees = Training years * Trees per year

Model Results

With the given data we have done case level prediction and below are the results.

Prediction of case outcomes, there are four case level results from our prediction model. Our model correctly predicts **68%** of the court decisions.

Feature Importance

feature	importance
previous_lc_direction_diff	0.038185
cumulative_lc_direction_diff	0.037549
decision_delay	0.030489
previous_court_direction_diff	0.018167
previous_direction	0.017817
cumulative_court_direction_diff	0.017548
cumulative_direction	0.017173
previous_action	0.015901
previous_court_action_diff	0.015796
cumulative_court_action_diff	0.015395

Here, feature importance suggests us which of our variables have the most effect in this model. Feature and corresponding importance (score) suggest us how important each feature was in classifying. This is most important part of random forest because we can clearly see that few features are more important in classification than others.

Model Accuracy

Random Forest model
==========
precision recall f1-score support
0 0.77 0.20 0.31 2019
1 0.68 0.97 0.79 3488
mean/total 0.71 0.68 0.62 5507
[[399 1620]
[121 3367]]
0.6838569093880515
==========

Dummy model
==========
precision recall f1-score support
0 0.00 0.00 0.00 2019
1 0.63 1.00 0.78 3488
mean/total 0.40 0.63 0.49 5507
[[0 2019]
[0 3488]]
0.6333757036499001
==========

A confusion matrix allows the visualization of the performance of a random forest algorithm. Here, first part shows confusion matrix & accuracy of actual random forest model while second part shows confusion matrix & accuracy of dummy model.

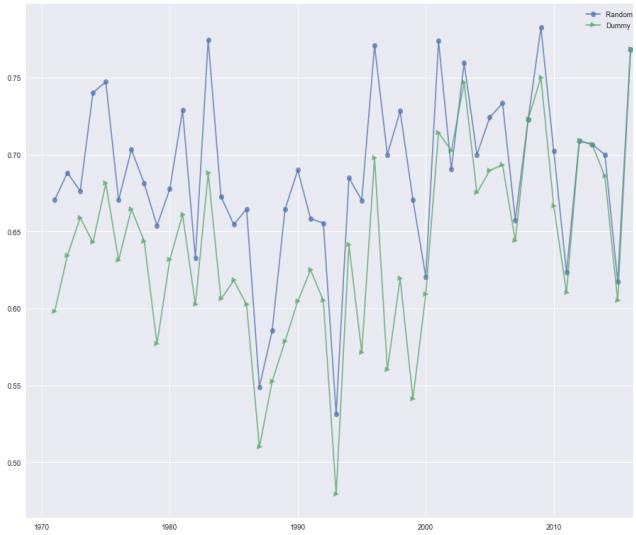

Fig 4. Random Forest vs. Dummy Model Training Pattern

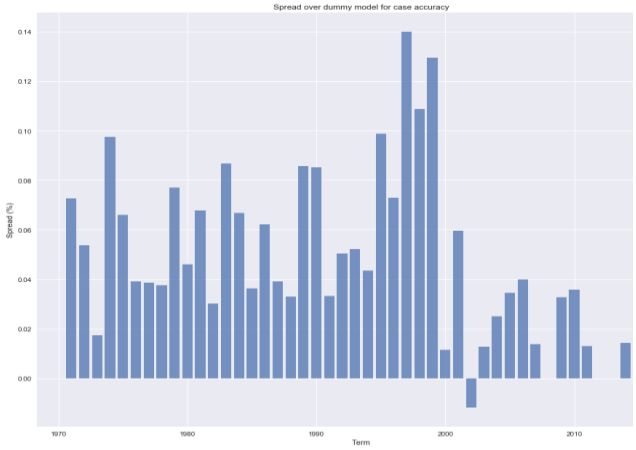

Fig 5. Dummy Model Case Accuracy

CONCLUSION

Based on the assumption of free availability of various types of legal data, we tried to build here a generalized machine learning model for helping out common citizens of India in knowing the potential outcome/judgment of their legal case & chances of win along with other useful information such as documentation, approx. case resolution timelines, details on the previously filed similar cases with their judgment outcomes and existing laws pertaining to the case. This model should be of interest to not only common citizens but to court observers, litigants and to market as well in future.

We believe the modeling approach undertaken in this paper can also serve as strong baseline for future research in the field of judicial prediction & legal informatics.

REFERENCES

[1] https://supremecourtofindia.nic.in/

[2] https://data.gov.in

[3] http://njdg.ecourts.gov.in/njdg_public/main.php

[4] http://dakshindia.org/state-of-the-indian-judiciary/08_contents.html#idTextAnchor06_002