

Predicting Outcome of Judicial Cases and Analysis using Machine Learning

Neha Parab¹, Namrata Soni², Bhakti Thakur³

^{1,2,3}Student, Dept. of Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India

Abstract – Today, the world is on the verge of cultivating the most advanced and modernized civilization of mankind. The progress in terms of architecture and technology is enormous. The society has greatly evolved and the humankind is availing the most progressed, enlightened, cultured, refined and sophisticated way of life. This remarkable attainment is largely attributed to the field of technology and law. Law and society form the most strong and crucial establishment. The Judicial System has outreached every aspect of the society and this inception is largely expanding. But, in the recent times, the challenges faced by the civilians with the judicial system is becoming more outrageous. The non-transparency and lack of knowledge of the judicial proceedings amongst the civilians has led to great disturbances and uncertainty amongst the communities confined to a certain legislature. This insufficient understanding of the legal system has caused extensive deprivation in terms of remuneration and time. With the advancement in technology there has been multiple attempts to assist the civilians in the judicial matters. This Project is in the similar endeavor to assist the civilians with the legal matters and bridge the increasing gap between the judiciary and the society. The emergence of Artificial Intelligence and relative technologies made prediction of the decision of a judicial system possible. This Project aims to predict the outcome of a legal case with the maximum accuracy.

Key Words: Artificial Intelligence, Machine learning, Judicial Case, Outcome, Prediction, Civilians, performance comparison.

1. INTRODUCTION

Law and society are the most inseparable attainments of mankind. With the advent of human civilization began the formulation of the judicial system. The legal system and jurisprudence have the greatest significance in the development of a society. The enforcement of law is pivotal in modulating the growing communities. Due to the prolonged expansion and modernisation in the society, the theory and philosophy of law has also been subjected to many curtailments and advances. Many communities and societies do not include jurisprudence as a part of their primary education. Because of this a large number of civilians remain unaware to the legal proceedings.

A lawsuit can be very challenging especially for those who are unfamiliar with the legal process which includes the following major acts: meeting with the potential lawyers, hiring a right lawyer, deciding in which court to file the lawsuit and considering the applicable remuneration and time period to sue. An uninformed civilian may fall prey to any false claims proposed to him and may endure intense expense in terms of time and funds. This system is propounded with a motive to provide right guidance by providing the appropriate prediction of the outcome of a lawsuit or litigation in accordance to the judiciary system.

The system formulates this objective of legal foresee by acquiring and working on the principles of Artificial Intelligence. The beginning of Artificial Intelligence was endowed with many myths, stories and was once quoted by Pamela McCordick as he writes, "I began with 'a ancient wish to forge the gods' ". With the extensive research and study being conducted, the world has onlooked the astonishing applications of artificial intelligence.

Today, the applications of machine learning have outstretched in all aspects of human life including businesses , finance and medicine. This system is also the follow up of data analysis and machine learning. There are multiple research and implementations carried out in this domain. Artificial Intelligence, especially in Natural Language processing (NLP) and machine learning is gaining quiet a attention of various problems. These technologies provide us with tools to analyse the text and summons up the semantics of the text involved. Overall, we believe that building a text-based predictive system of judicial decision can offer the civilians as well as lawyers the insight of their law cases and help them analyse the outcomes of the case.

2. Predictive Models

A. Logistic Regression

Logistic regression algorithm can be used for both classification as well as regression. Logistic regression is similar to the linear regression with it logistic score. We use independent feature of the data to predict or classify the dependent target value. This algorithm is similar to perceptron neural algorithm where we update the weights in order to obtain the predicted output

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Here, x_1, x_2, \dots, x_n (independent variable) are the features of the data and $b_0, b_1, b_2, \dots, b_n$ are the weights that are updated and y is the output (dependent variable)

B. K-nearest neighbor

It is a supervised learning algorithm where value of k is

predefined and changes in the value of k will give rise to variations in the performance metrics. In this algorithm, data points are plotted according to their independent features and then based on the value of k , a radius is formed around the datapoint which is to be classified so as to include k nearest classified datapoints around the datapoint to be classified. Value of k has to be an odd number for classification into two groups.

C. Naive Bayes

Naive Bayes is probabilistic classifier algorithm which make use of Bayes theorem of probability or Bayes rules. Naive Bayes algorithm is called naïve because it makes assumption that the occurrence of certain feature is independent of occurrence of other features and it is called Bayes since it uses Bayes rule or law for the classification. This algorithm predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is considered as the most likely class. This is also known as Maximum A Posteriori (MAP). Mathematically,

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

where,

$P(A)$ is the probability of hypothesis A being true. This is known as the prior probability.

$P(B)$ is the probability of the evidence (regardless of the hypothesis).

$P(B|A)$ is the probability of the evidence given that hypothesis is true, which is called Posterior.

$P(A|B)$ is the probability of the hypothesis given that the evidence is there, which is called likelihood.

D. Random Forest

Random forest is supervised classification as well as regression algorithm. This algorithm creates a number of decision tree using the rule-based system. As the name suggests it is a forest that is more the number of trees better is its performance measure. This algorithm first creates a bootstrapped data from the original data sets. Then it creates a decision tree on the basis of the bootstrapped data. We can create any number of divisions of feature either by selecting one feature, two or n features. This divided feature data is then converted to decision tree and these trees are parsed for the data to be classified and a decision is arrived on the basis of maximum votes for the class

E. Support Vector Machine

SVM is a supervised algorithm. In this algorithm the datapoints are separated into a class with the hyperplane which is as wide as possible from every class. This hyperplane is constructed with the help of margins of every class which is created from support vectors, that give rise to the margins. Hyperplane is constructed using kernel trick which describe whether the problem stated is nonlinear classifier or linear classifier. SVM algorithm gives good results with even large datasets as well as large feature space.

Hyperplane in the case of binary SVM classifier has the equation as:

$$g(\vec{x}) = \vec{w} \vec{x} + \omega_0$$

$$\text{If } g(\vec{x}) \geq 1, \forall \vec{x} \in \text{class1}$$

$$\text{If } g(\vec{x}) \leq -1, \forall \vec{x} \in \text{class2}$$

where $\|\vec{w}\|$ is normal vector to the hyperplane, θ_i denotes classes & x_i denotes features. The Distance between two hyperplanes is $\frac{2}{\|\vec{w}\|}$, to maximize this distance denominator value should be minimized i.e, $\|\vec{w}\|$ should be minimized.

3. Methodology

The main aim of this system is to predict a judicial cases verdict based on the verdict of previous Indian judicial cases. Various prediction model is introduced in previous section with their fair shares of pros and cons and thus we strive to perform the performance analysis of the all the mentioned algorithms. Feature extraction for the Indian judicial case is carried out by textual analysis of the case description in the prepared case database.

3.1 Datasets and Features

A wide range of data is collected for this system from the various legal Indian judicial sites (ref). This collected data is then prepared according to the need of the algorithms for the training and testing. Prepared data consists of corporate case description with their judgement in the corresponding column. This data prepared is then fed for splitting into training and testing set. We split our prepared data in the training and testing set of particular size ranging from 80%-20% to 70%-30%

To perform the predictive models on the data we have to described its feature. This step of feature extraction is performed by using the textual information of the cases and removes the stopping words or words irrelevant to describe the feature of the case, thus we use vectorization technique to define the weightage of a word in the description of the case. Apart from giving this information, vectorization says how important that word is to that document with respect to the corpus.

Vectorization technique is also used to convert our feature extracted into a numerical value which is then can be used for the evaluation of the models and can help us to plot the point graphically.

3.2 Performance metrics

We have various measure to analyze the performance measures of all five algorithm. Before we dive into the performance metrics term we need to know about true positive, true negative, false negative and false positive value.

True positive(tp): values which are correctly predicted positive value

True negative(tn): values which are correctly predicted negative value

False positive(fp): when values predicted are positive but target is to be negative

False negative(fn): when values predicted are negative but target is to be positive

a. Accuracy:

It is one of the measures for the performance analysis and it is the ratio of correctly predicted results to the total dataset.

$$\text{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn}$$

b. Precision:

It is the ratio of correct predicted positive result to the total positive predicted result.

$$\text{Precision} = \frac{tp}{tp+fp}$$

c. Recall or sensitivity

It is the ratio of correct predicted positive result to total result in that class

$$\text{Recall} = \frac{tp}{tp+fn}$$

d. F1 score:

F1 score is used when their uneven class distribution feature where accuracy is not as useful as f1 score. it is the weighted average of precision and recall.

$$\text{F1score} = \frac{2 * (\text{recall} + \text{precision})}{\text{recall} + \text{precision}}$$

3.3 Parameters

We have implemented all the five-prediction algorithm in python using scikit library of the python. We have evaluated performance analysis of all the algorithm on the basis of 80% of training dataset and 20% of testing dataset. Various parameters involved in the algorithm are described here.

In SVM we use kernel trick as linear(hyperplane) for the classification of the data into winning and losing

In random forest we create 20 number of decision trees for the classification of the dataset. as the number of trees increase performance increase but the complexity and processing time also increases.

In KNN, we have to define the value for the k which is used to form the radius for its decision to predict the output hence we used k=5 as the parameter for KNN.

In Naïve Bayes, we have three types of prediction model, i.e., gaussian, multinomial and Bernoulli. We have used multinomial type since it uses discrete counts and used in text classification which is our base of the judicial prediction system.

4. Experimentation and Results

In order to gain the maximum accuracy of prediction, the above predictive models are tested with the same data set. The accuracy of each predictive model is noted and the predictive model giving maximum accuracy will be implemented in the proposed system. The Fig-1 typically represents the report that is generated after testing is performed on each prediction or classification algorithm.

	precision	recall	f1-score	support
0	0.83	1.00	0.91	5
1	1.00	0.83	0.91	6
micro avg	0.91	0.91	0.91	11
macro avg	0.92	0.92	0.91	11
weighted avg	0.92	0.91	0.91	11

Fig -1: Performance of SVM.

Thus, a classification report will be generated for every algorithm. This classification report can be obtained by using in python. The below is the classification result of SVM algorithm. It is very crucial to understand the terminologies used in classification report, they are explained in the above section.

The classification report is generated for Naive Bayes, Logistic Regression, Random Forest, k-nearest and SVM algorithms. While Fig1 gives the classification report for SVM. Fig-2, Fig-3, Fig-4 and Fig-5 represent Naive Bayes, Logistic, Random Forest k-nearest respectively.

	precision	recall	f1-score	support
0	0.80	0.80	0.80	5
1	0.83	0.83	0.83	6
micro avg	0.82	0.82	0.82	11
macro avg	0.82	0.82	0.82	11
weighted avg	0.82	0.82	0.82	11

Fig -2: Performance of Naive Bayes.

	precision	recall	f1-score	support
0	0.67	0.40	0.50	5
1	0.62	0.83	0.71	6
micro avg	0.64	0.64	0.64	11
macro avg	0.65	0.62	0.61	11
weighted avg	0.64	0.64	0.62	11

Fig -3: Performance of Logistic Regression.

	precision	recall	f1-score	support
0	0.33	0.20	0.25	5
1	0.50	0.67	0.57	6
micro avg	0.45	0.45	0.45	11
macro avg	0.42	0.43	0.41	11
weighted avg	0.42	0.45	0.43	11

Fig -4: Performance of Random Forest.

	precision	recall	f1-score	support
0	0.67	0.80	0.73	5
1	0.80	0.67	0.73	6
micro avg	0.73	0.73	0.73	11
macro avg	0.73	0.73	0.73	11
weighted avg	0.74	0.73	0.73	11

Fig -4: Performance of K-nearest neighbor.

After studying each predictive model and determining their respective accuracy, the precision value of each predictive model is plotted against each other.

Thus, it can be clearly interpreted that the predictions of SVM algorithm is more accurate than the rest predictive model.

6. CONCLUSIONS

The advent of Artificial Intelligence has remarkably changed the perspective towards the applications of computational power in day to day life. With the robust collaboration of Statistics, Computer Science and Engineering, the machines can now adapt to the changing environment and easily optimize the problems that were previously non-computational.

The frequent developments in the field of data mining and Big Data has further given impetus to the applications of computational power to more diverse fields. Thus, the ever-evolving technology has led to prodigious innovations in the

society. This System has been largely inspired by the recent developments made in technology and aims at trussing the gap between Man and the Legal System. In present-day, the field of law has touched every aspect of human living. The field of law largely regulates the philosophy of Life. This has caused frequent acquittance of citizens with the judicial system. Thus, the dependence on the legal systems is tremendous. Also, the rapid change in the ideology of living and evolution, the judicial system is facing immense pressure. All this have largely penalized the civilians in terms of remuneration and time.

This system is with a strong motive to provide legal assistance to the civilians by predicting outcome of the case that the citizen will approach the legal system. This system is estimated to provide the prediction of outcome win or loss with an accuracy of 78%. Thus, the system will oblige, encourage civilians in making credible decisions with their legal proceedings and outreach the citizens with the legislature for the betterment of the society.

REFERENCES

1. Stefaine Brninghaus and Kevin D. Ashley Predicting Outcomes of Case- based Legal Arguments, Learning Research And Development Center and Intelligent System Program, School of Law University of Pittsburgh, PA
2. Daniel Martin Katz, Michael J. Bommarito II, Josh Blackman A general Approach for predicting the behaviour of Supreme Court of the United States, April 12, 2017..
3. Zhenyu Liu, Huanhuan Chen A Predictive Performance Comparison of Machine Learning Models for Judicial Cases, IEEE, 2017.
4. Octavia-Maria ulea , Marcos Zampieri, Mihaela Vela, Josef van Genabith Predicting the Law Area and Decisions of French Supreme Court Cases,2017.
5. Derek Yan ,Tiyangi Wang , Patrick Chase Personal Legal Counselor and Interpreter of Law via Machine Learning, FINAL REPORT FOR CS229,FALL 2014.
6. Sushanta Kumar, P. Krishna Reddy, V. Balakista Reddy, AdityaSingh Similarity Analysis of Legal Judgments, ACM, 2011.
7. Top 20 Landmark Judgements of Corporate Law. (2019, Feb 05) Retrieved from<https://lawnn.com/top-20-landmark-judgements->
8. 40 Latest Corporate Law Cases Law in Simple Format. (2019, Feb06) Retrieved from <https://www.caclubindia.com/forum/40-latest-corporate-laws-case-laws-in-simple-format-124405.asp>
9. Case laws on company law. (2019, Feb 10). Retrieved from<https://indiankanoon.org/search/?formInput=case%20laws%20on%20company%20law>
10. Top 25 Cases of Tax & Corporate Laws Reported in 2017. (2019, Feb15) Retrieved from <https://www.taxmann.com/blogpost/2000000282/top-25-cases-of-tax-corporate-laws-reported-in-the-year-2017.aspx>
11. CaseMine: The Most Granular Mapping of US Case Law. (2019, Feb15) Retrieved from <https://www.casemine.com/home>.
12. Kanoon(2019, Feb17) Retrieved from <https://indiankanoon.org/>

