

# Machine Learning Approach for Automated Legal Text Classification

Riya Sil<sup>1</sup>, Alpana<sup>2</sup> and Abhishek Roy<sup>3</sup>

<sup>1</sup> Department of Computer Science and Engineering, Adamas University,  
Barasat, Kolkata 700126, India  
*riyasil1802@gmail.com, dr.aroy@yahoo.com*

<sup>2</sup> Department of Computer Science and Technology, Manav Rachna University,  
Faridabad, Haryana - 121004, India  
*alpana.srk@gmail.com*

**Abstract:** As a developing nation, India is dealing with the huge burden of population explosion. Due to this reason, there persists a large number of issues amongst the citizens where all the contesting parties seek legal help to obtain the desired justice. This results in the filing of legal cases at several courts of our country, which remains pending for a longer duration due to lack of sufficient manpower and infrastructure. Moreover, in these recent days of the global pandemic of Coronavirus and subsequent phases of economic lockdown have adversely affected the socio-economic balance within our society. This incident had created various types of violence within our society particularly targeting the weaker sections i.e., domestic violence against women. To save the weaker section of society it has become the need of the hour to resolve all the pending legal cases related to domestic violence against women. To assist legal professionals in this process, Machine Learning based advanced techniques may be applied within the legal domain. To achieve this objective, authors have proposed a judgment classification model related to 'Domestic Violence Against Women Act'. Our proposed model will support legal professionals in resolving these cases in a structured manner through classification and identification of legal datasets (collected from various court proceedings of West Bengal) to identify the accused person. For classification of the offender, six classifiers namely, Multi-Layer Perceptron Classifier, Decision Tree Classifier, K-Nearest Neighbour Classifier, Naïve-Bayes Classifier, Support Vector Machine Classifier, and Ensemble Classifier have been applied to obtain the desired output.

**Keywords:** Machine learning, Document Analysis, Classification, Legal Dataset.

## I. Introduction

India, as a developing nation is dealing with a huge population thereby leading to an enormous number of issues. As a result, people seek help from the judicial system to obtain legal relief from their conflicting issues. Advanced technologies such as Artificial Intelligence, Data Analytics, Machine Learning, etc. may be applied to resolve various issues like legal disputes, social disputes, economic disputes, etc. so that justice can be delivered to the beneficiary within the proper time [1]. As an emerging technology, Artificial Intelligence (AI) has immense potential to manage various issues (including legal issues) in a structured manner, which was conceptualized in

the early 1990s with the help of LexisNexis and Westlaw [2]. Using Artificial Intelligence (AI), one can analyze semantic-based legal information and also make predictions from a given dataset. Nowadays, it has expanded to extraction, classification, statistical data analysis, and summarization of any raw dataset, including legal documents which are mostly unstructured in nature [3]. Presently many countries have started to follow the directive to encourage convenience and re-use of public information through virtual medium thus exploring new dimensions for automatic legal data analysis [4].

Analysis of language has also been used for years in the legal domain including criminology and forensic linguistic [5]. With the gradual evolution of science and technology, Machine Learning algorithms has helped legal professionals, starting from finding the name of an individual to age, traits, and even authors. Moreover, in this recent situation of global pandemic of coronavirus, excessive adjournments of legal proceedings are causing a huge pile of pending cases thereby lengthening the justice delivery system. Due to this delay, complainant, accused and witness may become medically unfit for trial, hostile, etc. which may jeopardize the effectiveness of legal proceedings and justice delivery system as a whole. To resolve the issue, Machine Learning based models can be proposed to optimize the workload of legal professionals, so that they can handle multiple pending cases simultaneously.

To implement Machine Learning based models, computers are trained to predict the decision of legal cases on the basis of quantitative analysis of words and phrases and their inter-relationships which are retrieved from unstructured raw legal dataset, collected from hard copy legal documents of several courts. The accuracy of prediction obtained through this approach will help to find the word with the most impact on final decision thereby resulting in the identification of the most significant factor for every judgement. Our proposed model will assist the legal professionals to resolve huge pending cases especially for the soft targets, like domestic violence cases against women. It includes structuring of case documents, classification and identification of judicial verdicts thereby recognizing the offender. To accomplish this

objective, authors have designed a Machine Learning-based automated legal classification system that will structure the case document by substantiating performance and accuracy of our legal classification model using six classifiers namely, Multi-Layer Perceptron Classifier, Decision Tree Classifier, K-Nearest Neighbour Classifier, Naïve-Bayes Classifier, Support Vector Machine Classifier and Ensemble Classifier to predict the accurate output.

This paper emphases on the argument-based legal classification system. Precisely, we have applied the Supervised Machine Learning technique over the cases related to ‘Domestic Violence Against Women’ and have proposed a model to predict offence of the accused person. The following steps have been followed in a sequential manner within our proposed legal model: (i) Hard-copies of argument-based legal documents are collected from various courts of West Bengal, India. (ii) Datasets are manually created on the basis of specific parameters, namely ‘dowry taken within a span of 7 years of marriage (Yes / No)’, ‘tortured after marriage (Yes / No)’, ‘incident taken place within a span of 7 years of marriage (Yes/ No)’ and many more, which act as deciding factors for the prediction of accusation. (iii) The Supervised Machine Learning algorithm is used to train legal datasets for prediction of the desired output. Finally, we have validated the performance and accuracy of our proposed legal prediction system using Standard Classifiers that are Multi-Layer Perceptron Classifier, Decision Tree Classifier, K-Nearest Neighbour Classifier, Naïve-Bayes Classifier, Support Vector Machine Classifier, and Ensemble Classifier.

Section 2 states the relevant literature survey. Section 3 discusses the basic concept of Machine Learning and its various Classification methods. Section 4 explains the Machine Learning based proposed Legal Analytics model. Section – 5 discusses the Classification of our proposed Legal Analytics Model. Section 6 shows the experimental results of our proposed model and its accuracy. Conclusion of this work is mentioned in Section – 7.

## II. Literature Survey

For years, researchers have applied different principal methods in the field of law that discuss various problem-solving skills, description of law and other details related to legislation [6]. From the doctrinal analysis related to legal cases, it is witnessed that all the work process is done through manual collection of data, summarization, comments are made and read and finally everything is placed over the legal system [7]. Currently, as there are enormous number of published legal cases [8], it is difficult for the legal researchers to analyse, deliver and structure the documents related to Legal verdicts [9]. The merging of empirical quantitative methods with doctrinal legal methods [10] has helped the researchers to study legal cases [11].

In case laws of the United States of America, quantitative analysis has a lengthier period of tradition than that of the rest of the world. American court’s case laws are available in numerous quantitative datasets. Manual collection and coding of case laws are used in most studies. Databases from the Supreme Court are also used in many studies that consist of manually collected data and expert code on the behavior of the

US Supreme Court over the past 200 years [12]. Another major part of this study analyses decision making and relationship between political background and gender of the judges [13-15].

For other countries except the United States of America, quantitative analysis of case law is not a regular practice [16]. In few of the research works like that of Hunter et al. [17] states that due to insufficient number of male and white judges to provide sensible statistical analysis, United Kingdom has still not established any tradition. Several researchers from Belgium [18], Czech Republic [19], France [20], Germany [21,22], Israel [23], Latvia [24], Netherlands [25], Slovenia [26], Spain [27], and Sweden [28] had worked on the application of quantitative methods over legal datasets.

Many researchers have specifically worked on quantitative analysis related to international case law. Behn and Langford [29] accumulated and coded around 800 cases regarding Investment Treaty Arbitration [30]. Other researchers have analyzed case law of the International Criminal Court [31] by using quantitative methods [32], the Court of Justice of European Union [33-34] or the European Court of Human Rights [35-38].

Some of the research areas focus over analysis of citation regarding USA courts’ case laws that analyses citation patterns with reference to documents, impact and numbers related to case law [39-40]. In other countries, this process has been implemented by researchers from Sweden, European countries, the CJEU [41]. Some studies reveal that the main focus of research work is basically related to legal case-based text mining arguments [42-43].

Involuntary legal data analysis for prediction of court decisions requires identification of arguments. Due to lack of data type [44], data-driven approach without argument mining is used where unprocessed data is used for prediction of court decisions. Other studies involve court arguments for identifying the verdict [45-46]

Based on the knowledge gathered through this brief literature review, in this paper authors have focused on Dowry Prohibition Act of Indian Penal Court. The study of Riya et al [47] is pertinent to the research work done in this paper. However, in the cited paper, the process has been performed based on a single classifier and with lesser number of cases has been taken for prediction whereas in this paper, six classifiers namely, Multi-Layer Perceptron Classifier, Decision Tree Classifier, K-Nearest Neighbour Classifier, Naïve-Bayes Classifier, Support Vector Machine Classifier and also Ensemble Classifier has been applied for classification of the offender.

## III. Fundamentals of Machine Learning

Machine Learning, a subfield of Artificial Intelligence, focuses on developing a customized operational model with reference to relevant parameters [48]. It is a data analysis process that helps in the automation of analytical building of model. Machine Learning technique basically depends on the pattern of enormous amount of data [49-51]. These patterns are then classified and used in various tasks. It trains model to get the most appropriate result. With the help of this concept, several problems can be solved by broadly categorizing them into classification and regression problems. Using machine learning, system can easily and automatically learn and

improve from its previous experience without any explicit program. On the basis and ways of learning techniques, classifications of Machine Learning can be done namely, Semi-supervised Learning, Reinforcement Learning, Supervised Learning, and Unsupervised Learning.

In Supervised Learning, sample inputs and their respective labels are provided. It depends on the training of the model by comparing pre-learned output with actual output and finding its errors for further enhancements. In Unsupervised Learning, the dataset is unlabelled therefore, the model finds similarities among the dataset by itself. Unsupervised Learning may be correlated to search for hidden patterns in the input data and also focuses on learning features by itself. Semi-supervised Learning aims to train by using unlabelled data. It has been found that using both labelled and unlabelled data simultaneously leads to improved accuracy of the algorithm. And, in the case of Reinforcement Learning, proper actions are required to elevate the reward to its maximum level for a particular situation. In Supervised Learning, training procedure is totally reliant on existing labelled data-set whereas in Reinforcement Learning labelled datasets are unavailable and hence the output is obtained based on previous dataset.

#### IV. Proposed Legal Analytics Model

When it comes to the legal field, there are numerous ways of dealing with legal cases. Most of the processes taken are mainly to systemize data and automate the process. In this section, authors have discussed few ways of processing legal texts automatically. To assist the legal professionals to balance their overload of pending cases, computer science oriented advanced scientific techniques like Artificial Intelligence, Machine Learning, Data Analytics, etc may be blended with legal documents (mostly available in unstructured or semi structured hard copy format) for proper classification, prediction and other purposes. This approach will help to make law more accessible, manageable, predictable, understandable and useful. Legal information is mostly unstructured as it is mostly written in natural languages. Therefore, to process this legal information different techniques have been developed using Natural Language Processing. The main focus of this research work is to develop a system that will assist legal professionals to predict the verdict of a legal case automatically. To accomplish this objective, authors have designed a supervised machine learning-based automated legal classification system where argument-based legal text documents along with copy of the judgments are collected from different trial courts of West Bengal, India. By providing any such information about different cases, the computer gets to identify the patterns in the cases based on certain parameters. This phase is also known as the Training Phase. For evaluating the performance of the proposed Machine Learning-based classification model, data are provided into the system without the verdict (i.e. in the testing phase) to find the most likely verdict. Structuring of the case information is done by substantiating the performance obtained from the accuracy of the legal classification model.

In this paper, authors have used various Classification Methods, namely Multilayer Perceptron Classifier, Decision Tree Classifier, K-Nearest Neighbour Classifier, Naïve-Bayes Classifier, Support Vector Machine Classifier and Ensemble Classifier to predict and classify the most accurate result of an

accused person being offender or not from a set of hard copy legal documents that has been collected from several trial courts of West Bengal, India.

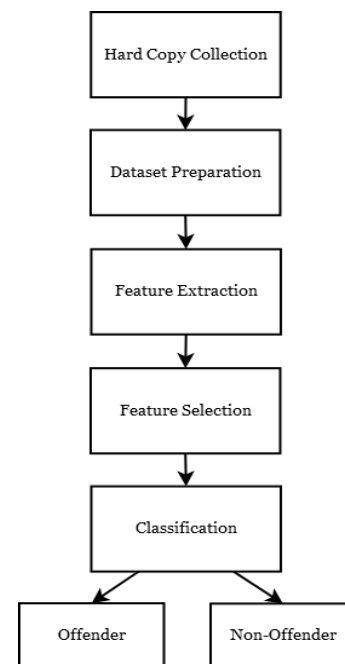


Figure 1. Proposed Legal Analytics Model

##### A. Hardcopy Collected and Dataset Preparation

Based on the characteristics of texts in a case, the Machine Learning algorithm takes input and learns the characteristics of legal cases. The 'Dowry Death Act' under the 'Indian Penal Code' includes multiple complaints in a single case related to similar Sections of the Dowry Death Act. However, in this paper, authors have conducted a binary task to classify only two possible decisions: whether an accused person is 'Offender' or 'Non-Offender' in a particular case. Though in several cases there are multiple offenses that lead to a charge of multiple sections, however in this paper authors have focussed on the final judgment in a case of a person being an offender or not. The main objective of this paper is to regulate the pattern explicit to an accused person being 'offender' or 'non-offender'. A large amount of data during the training phase will provide accuracy and better performance of the program. However, it is significant to learn what kind of data is fed into the machine to learn. In our proposed legal model, we have used the legal dataset related to dowry death cases only to predict the offense of the accused person based on certain parameters of Indian Penal Code (IPC) 498A and 304B. Selection of these parameters are taken with the help of a legal professional that includes: (i) Dowry taken within a span of 7 years of marriage (Yes: 1 / No: 0) (ii) Torture after marriage (Yes: 1 / No: 0) (iii) Incident took place within a span of 7 years of marriage (Yes: 1 / No: 0) and many more. A legal dataset is created by including these parameters for training and testing of the machine.

In this section, authors have explored the procedure of creating a legal dataset from argument based legal documents by extraction of important deciding features that are related to 'dowry cases' under the 'Dowry Prohibition Act'.

Dataset is a collection of data. This dataset is created before the machine is trained. It may be structured, semi-structured, or unstructured. As no existing, pre-structured argument

based legal judgment dataset is available, it should be created using Comma-Separated Value (CSV) file format to retrieve data in a structured manner. The hard copies of case documents and judgments related to ‘Dowry Death Act’ of ‘IPC Section 498A and 304B’ are collected from several trial courts of West Bengal, India. Due to the huge number of case documents, the data related to the parameters are stored manually in MS Excel file format for training using the Machine Learning model. Figure 2 shows the structured data set related to the Dowry Death case in India.

Victim	Tortured and assaulted by	Dowry (Yes: 1 / No: 0)	Torture after marriage (Yes: 1 / No: 0)	Incident Within 7yrs (Yes: 1 / No: 0)	Post Mortem Report (Unusual: 1 / Usual: 0)	If person is Offender (Yes: 1 / No: 0)
Sonerhar Biliti	husband	0	1	1	1	1
Sonerhar Biliti	husband	0	1	1	1	1
Sujata Dutta	ww,father in law, nanad,materi	1	1	0	1	1
Sujata Dutta	ww,father in law, nanad,materi	1	1	0	1	1
Sujata Dutta	husband, mother in law	1	1	0	1	1
Sujata Dutta	husband and in-laws	1	1	0	1	1
Sujata Dutta		1	1	0	1	1
Sujata Dutta	9 by in-laws and physically tor	1	1	0	1	1
Sujata Dutta	husband and in-laws	0	1	0	1	1
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	Husband	0	1	0	1	0
Bharti Majumdar	none	0	1	0	1	0
Ajiba Biliti	Kalo Dafader (Parents-in-law)	1	1	0	1	1
Ajiba Biliti	Family members	1	1	0	1	1
Ajiba Biliti	Kalo Dafader (Parents-in-law)	1	1	0	1	1
Ajiba Biliti	Kalo Dafader (Parents-in-law)	1	1	0	1	1
Ajiba Biliti	Kalo Dafader (Parents-in-law)	1	1	0	1	1
Ajiba Biliti	Kalo Dafader (Parents-in-law)	1	1	0	1	1
Champa Hembram	Husband	1	1	0	1	1
Champa Hembram	Husband	1	1	0	1	1

**Figure 2.** Sample legal dataset collected from several trial courts of West Bengal

### B. Feature Extraction and Selection

Features are mainly independent variables that are fed as input into the machine learning model to provide results related to classification and prediction. These are attributes that consist of values for a specific dataset. Few of the features extracted from the evidential hard-copy documents are: (i) ‘Victim’s Name’, (ii) ‘Tortured and assaulted by’ (iii) ‘Dowry taken within the span of 7 years of marriage (Yes / No)’ (iv) ‘Years married (More than 7 years) (Yes / No)’. The case filed will be known as a ‘dowry case’ only if it is filed and dowry is taken by the defendant within 7 years of marriage. (iv) ‘Incident took place within a span of 7 years of marriage (Yes: 1 / No: 0)’. These parameters determine whether the case will be referred to as a ‘dowry death case’ or not depending upon the death of the victim has taken place within the span of 7 years of marriage. (v) ‘Post-mortem Report (Usual: 0 / Unusual Death: 1)’ and many more parameters are included. Only the significant parameters or features are considered for data analysis and remaining parameters are excluded from the Feature Selection process. The selected features play a vital role in generation of desired output. To get the predicted value of the accused person being an offender or not, our proposed legal model is trained based on these selected features.

### C. Classification

Legal cases can be processed in many probable ways. Researchers have applied various processes to systemize and automate the process. In this section we have discussed Machine Learning and Deep Learning algorithms which are used within our proposed argument based legal judgement prediction system.

As most of the trial court legal documents are unstructured in type, they should be structured properly. The prosecution witness statements and argument based legal documents of India are type-written, and hence soft copy versions of Indian legal documents should be generated using Supervised machine learning algorithms to generate the desired output.

To calculate the result of our argument-based proposed legal model, data is provided to the system without the final judgement. This phase is called the Testing Phase, which provides an accurate output.

#### 1) Decision Tree

As a Supervised Machine Learning technique, the Decision Tree algorithm organizes data continuously on the basis of selected attributes or parameters to predict the outcome of a test. Hence, it represents either class labels or class distribution scores. Decision Tree consists of three main components, namely (i) Node (ii) Edge and (iii) Leaf Node.

(i) Nodes: Nodes are tested for the value of a specific attribute. (ii) Edges: Edges connects nodes of a tree for representation of the test outcome. (iii) Leaf Nodes: Leaf Nodes are also known as Terminal Nodes. Decision Tree consists of three types of nodes that include ‘Leaf Node’, ‘Internal Node’ and ‘Root Node’. The main objective of Decision Tree is to break the group of data into smaller subsets and simultaneously build an associated Decision Tree. The topmost node in a decision tree is the Root Node or Initial Attribute, which is the top predictor to make decisions. It has no incoming or outgoing edge unlike Internal Nodes. As the Decision Tree gets complex by getting deeper, its decision-making accuracy depends over its complexity.

Creation of Classifier: We have used the DecisionTreeClassifier module from the Scikit-Learn package of Python. The equation for creation of classifier is stated below:

$$dtClassifier = DecisionTreeClassifier()$$

Fitting inputs to Classifier: We have used fit ( ) function of DecisionTreeClassifier class for input (i.e. insertion of data from data set into the created Classifier), which is also stated below:

$$dtClassifier.fit(XTrain, YTrain)$$

Finding the Predictions / Classes: After fitting inputs into the Classifier model, we obtain classes predicted by our classifier. This task of prediction is done using predict() function of DecisionTreeClassifier class, which is stated below:

$$YPred = stClassifier.predict(XTest)$$

#### 2) Naïve Bayes

The Bayesian classification method is a supervised statistical classification method that is based on Bayes Theorem and is used for Pattern Recognition. The principle of Bayes’ decision theory is to select the most probable one.

Independent features within classes are classified using Naïve Bayes classifier.

The Naive-Bayes Classifier, a special kind of Supervised Classifier, belongs under a Probabilistic Machine Learning model, which is stated below:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

where,  $P(A)$  is the prior probability of class,  $P(B)$  is the prior probability of predictor

$P(B|A)$  is the likelihood which is the probability of the predictor given class

$P(A|B)$  is the posterior probability of class (A, target) given predictor (B, attributes)

Above equation helps to predict the probability for happening of event A only if event B is given. Here, B is the evidence and A is the hypothesis. It is also assumed that the features or attributes are independent i.e., naive to each other and hence it is known as Naive-Bayes Classifier. Naive Bayes Classifier may be further stated as Multinomial Naive-Bayes and Bernoulli Naive-Bayes, among which we have used Multinomial Naive-Bayes Classifier in our proposed legal model. Creation of the classifier: We have used the MultinomialNB module from the Scikit-Learn package of Python, which is stated below:

```
nbClassifier = MultinomialNB()
```

Fitting the inputs to the classifier: We have used the fit ( ) function of the MultinomialNB class for input (i.e. data from the data set into the created classifier), which is shown below:

```
nbClassifier.fit(XTrain, YTrain)
```

Finding the predictions / classes: After fitting inputs into Classifier Model, we get the desired classes using predict ( ) function of MultinomialNB class, which is stated below:

```
YPred = nbClassifier.predict(XTest)
```

### 3) Support Vector Machine

Support Vector Machine (SVM) is another example of Supervised Machine Learning algorithm that can be used both for Classification and Regression, but it is generally used for classification related problems. It is used in both linear and non-linear processes using Kernel. Support Vector Machine (SVM) can discover linear separation and works best when there are a limited number of points for numerous measurements.

In addition, Support Vector Machine (SVM) is also known to be universally learned. Support Vector Machine (SVM) basically learns linear threshold function in its basic form. Nevertheless, they can level-up their function to alternative future function with the help of easy plugin appropriate kernel. However, Support Vector Machine (SVM) is computationally exorbitant and mathematically complicated.

Support Vector Machines (SVM) are classifiers that distinctly classify a set of points or data using a N-dimensional Hyperplane. A Hyperplane is nothing but a separating n-dimensional line which separates data belonging to two different classes. This classifier is used in case of both regression and classification problems.

Creation of the classifier: We have used the Support Vector Machine (SVM) module from the Scikit-Learn package of Python. The statement of classifier creation is as shown below:

```
svmModel = svm.SVC()
```

Fitting the inputs to the classifier: We have used the fit ( ) function of the svm.SVC ( ) class which feeds in the inputs, i.e. data from the dataset into the created classifier. The statement of fitting the inputs to the classifier is as shown below:

```
svmModel.fit(XTrain, YTrain)
```

Finding the predictions / classes: After fitting the inputs to the classifier model, we get the classes that are predicted by our classifier from the testing input data. The task of prediction is done using the predict ( ) function of the svm.SVC ( ) class. The statement of predicting the output classes is as shown below:

```
YPred = svmModel.predict(XTest)
```

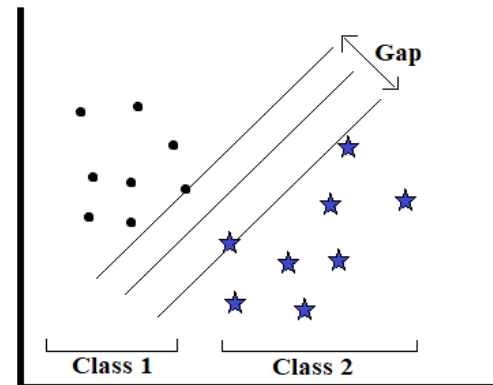


Figure 3. Support Vector Machine

Figure 3 shows an example of a working system [47]. The best hyperplane is decided by the algorithm for the separation of data. The hyperplane is the centre line that separates the two symbols. The nearest symbols to the lines are the support vectors. The objective of the Support Vector Machine algorithm is to fix the hyperplane such that the most likely margin is achieved with respect to the points. Once the training is completed, a new and different case set is used for performance evaluation of the Machine Learning approach. Each case is evaluated to check if the accused person is an offender or not, which is thereafter compared with the original output of the court.

### 4) K-Nearest Neighbour

It is a simple algorithm classification method with a good amount of highly competitive output of classification accuracy. It is a direct classifier that is based on the nearest neighbour class. It works best on large amounts of dataset. In K-Nearest Neighbour (KNN) classifier, every unknown test sample gets assigned with a specific class belonging where most of its K-Nearest Neighbour (KNN) belong. That is, it is based on the greatest number of votes of its most related k-nearest neighbours to which the object will be assigned to. K-Nearest Neighbour (KNN) algorithm is also known as: (i) Memory-Based Reasoning (ii) Instance-Based Learning (iii) Case-Based Reasoning (iv) Example-Based Reasoning (v) Lazy Learning. It is basically non-linear and can differentiate whether the dataset is linear or nonlinear.

As K-Nearest Neighbour (KNN) Classifier assumes similar things reside in close proximity, in our proposed legal model words are represented as things.

Creation of the classifier: We have used KNeighborsClassifier module from Scikit-Learn package of Python, which is stated below:

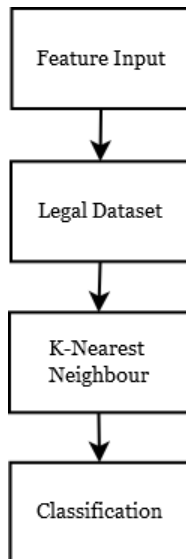
`knnClassifier = KNeighborsClassifier()`

Fitting the inputs into classifier: We have used `fit()` function of `KNeighborsClassifier` class for input of data into the created classifier, which is stated below:

`knnClassifier.fit(XTrain, YTrain)`

Finding the predictions / classes: After insertion of inputs to classifier model, we obtain the classes predicted by our classifier using `predict()` function of `KNeighborsClassifier` class, which is also stated below:

`YPred = knnClassifier.predict(XTest)`



**Figure 4.** K-Nearest Neighbour

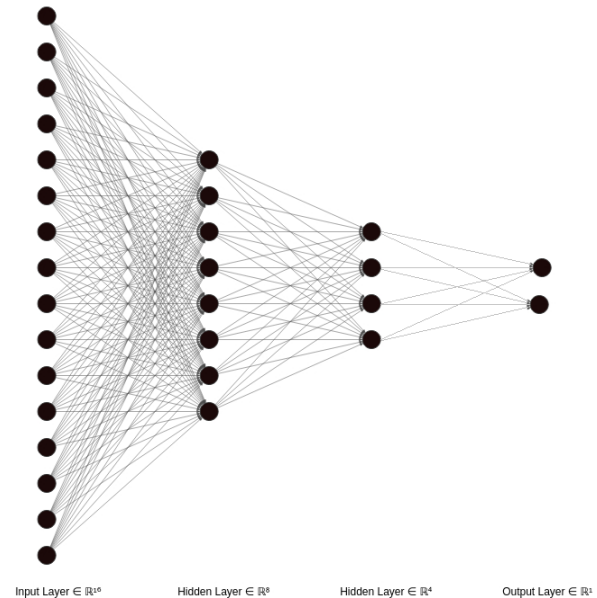
### 5) Multi-Layer Perception

As a specific type of Machine Learning model, Artificial Neural Network simulates the structure of the human brain and neurons for classification of input data set within a machine. As shown in Figure 5, in this paper authors have built a multi-layered feed forward network.

Our model involves 4 Dense Layers namely: (i) Input Layer (ii) First Hidden Layer (iii) Second Hidden Layer (iv) Output layer. Every node of one layer is connected to each node of next layer, which are mentioned below:

- Dense Layer:** It is a fully-connected layer where every neuron is connected to each other through the layers. It uses all learning features that come from combinations of all features of previous layers. In the first Dense Layer, 16 nodes take one-dimensional input whose shape is (4, 1). The activation function used in this layer is ReLU i.e., Rectified Linear Unit. This activation function transforms the weighted sum of inputs to output nodes as the output. The number of nodes in Second, Third and Fourth Dense Layers of our model are taken to be 8, 4 and 1 respectively. ReLU activation function is used in Second and Third Dense Layers and softmax is used in Fourth Dense Layer to get the Class Probability Scores.
- Dropout Layer:** In between each successive dense layer, we have added Dropout Layer to prevent overfitting with a dropout rate of 0.2, i.e., one out of five inputs will be removed at random in each epoch.

One of the utmost popular supervised neural classifiers is Multilayer perceptron (MLP). Numerous learning paradigms are developed till date and can perform nonlinear mapping as well. MLP networks consist of nonlinear activation functions. This non-linear mapping is done in the MLP network by hidden layers and connected synaptic weights. A general supervised method known as Backpropagation is used for iterative calculation of biases and weights of the MLP network.



**Figure 5.** Architecture of our MLP model

After obtaining predicted class labels, those are analysed using Classification Report and Accuracy Score, which are shown below in tabular format.

### 6) Ensemble Classification

Over the past few decades, Ensemble-based systems, also known as Multiple-Classifiers Systems have been well known and therefore drawn quite a substantial amount of researchers' attention. Single-classifiers are less preferred as compared to multiple-classifier systems for numerous reasons. The implementation details of several other circumstances where favourable results have been displayed are shown in Kuncheya's work. Out of the different properties of multiple classifiers, the most desirable property is that of Classifier diversity. Being one of the desirable properties of multiple classifier systems, classifier diversity can be achieved through numerous probable ways. The most common ways of achieving diversity are by combining different classifier sets, using different datasets and training parameters that will be used for training of individual members. A classifiers' set with sufficient decision boundaries are required for a diverse ensemble classifier to be built. For the combination of outputs of each distinct classifier and to design an ensemble classifier such that the potential incorrect decisions are ruled out and the correct ones are amplified, an appropriate strategy is required to be framed. It is one of the approaches for application to only class labels and is built on classifying of decision output. Class labels are combined by multiple classifier systems that

are attained from specific ensemble members to predict the ultimate class-label [52-55].

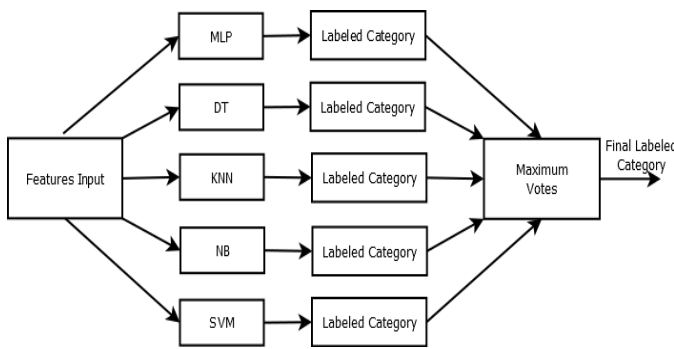


Figure 6. Ensemble learning model for legal classification

### V. Experimental Results

Our proposed Machine Learning algorithm based legal model has been projected and executed using Python 3.5 and AMDA Ryzen 5 with RAM 16 GB has been used for the processed model and executed on Windows 10 professional operating system. The final output is classified based on whether the accused person is an offender or non-offender. In this paper, authors have used different classifiers namely, Multi-Layer Perceptron Classifier, Decision Tree Classifier, K-Nearest Neighbour Classifier, Naïve-Bayes Classifier and Support Vector Machine Classifier that analyse. An appropriate and distinguishable Ensemble classifier has been suggested for classification to get better accuracy.

Table-1 shows the performance measure of the classifiers. For the evaluation, in this research F1 (Macro), Precision and Recall has been used as a performance measure to check the feasibility of the machine learning models. The Harmonic Mean related to precision and recall is the F-1 score. The correct assigned label for the percentage related to case is called Precision. Correct identification of a certain label for the percentage related to case is called Recall.

Sl. No.	Classifier	F1 (Macro)	Precision	Recall
1	MLP	0.71	0.78	0.69
2	KNN	0.77	0.81	0.79
3	NB	0.82	0.83	0.88
4	DT	0.87	0.89	0.89
5	SVM	0.91	0.90	0.92
6	EC	0.93	0.94	0.93

Table – 1. Performance Measure

Table 2 shows the classification accuracy of Machine Learning classifiers using K-Fold cross validation technique. Application of cross-validation technique resolves the dearth of data for generation of validation sets. In this paper, k-fold cross validation has been used for the resampling process. This process starts with the division of the original sample via stratified split into k equal size separate subsets namely S1, S2, S3.....Sk. After this, k-times execution of the training sample test is conducted in a sequential manner. Here, k has

been considered as 5 on the basis of which all the sample sets are divided into 5 parts. Therefore, the execution of the entire process involves 5 epochs consisting of various grouping of training and testing samples. Finally, the approximate average of 5 operations have been done to deliver accuracy and computational time on the whole classifier recitals' approximation.

Sl. No.	Class ifier	Fold1	Fold2	Fold3	Fold4	Fold5	Accu racy (%)
1	MLP	70.00	72.21	76.61	74.02	67.26	72.02
2	KNN	76.78	80.67	82.59	81.35	80.16	80.31
3	NB	82.39	85.37	84.79	81.09	85.66	83.86
4	DT	83.34	88.01	86.62	91.51	90.72	88.04
5	SVM	88.68	92.89	91.33	90.97	89.48	90.67
6	EC	91.76	94.59	92.07	93.18	93.20	<b>92.96</b>

Table – 2. Classification Accuracy

For a better visualization, the classification accuracy of Machine Learning models for legal text classification has been shown in Figure 7.

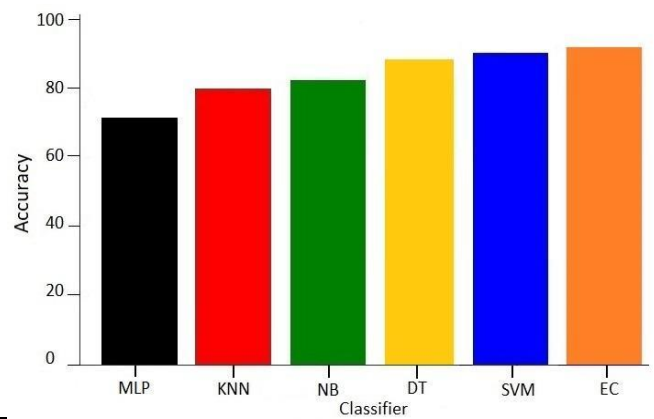


Figure 7. Accuracy of classical ML algorithms are compared

In Figure 8, the computational time taken by different classifiers for legal text classification is shown.

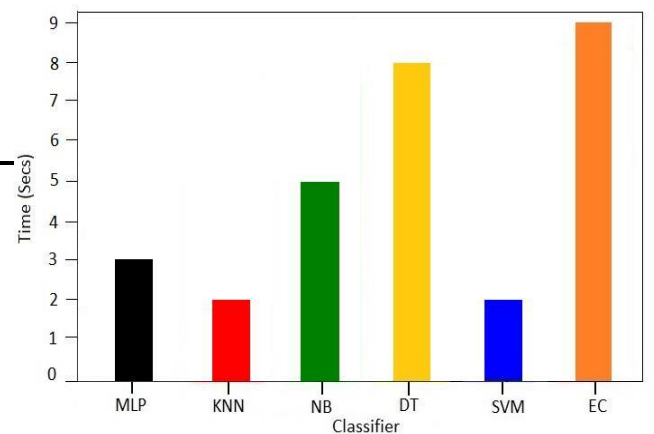


Figure 8. Computational Time for different classifiers

## VI. Conclusion

In this paper, authors have proposed a Machine Learning algorithm-based judgment classification model to support legal professionals to resolve the huge pile of cases that are pending for decades and deliver justice to the victims. Classification and Identification of legal datasets have been done in a structured manner to identify the offender. For classification of the offender, six classifiers namely, Multi-Layer Perceptron Classifier, Decision Tree Classifier, K-Nearest Neighbour Classifier, Naïve-Bayes Classifier, Support Vector Machine Classifier, and Ensemble Classifier have been applied and analyzed thoroughly in tabular fashion.

## Author Contributions

All authors contributed to the study conception and design. Riya Sil: conceptualization, formal analysis, material preparation, data collection, coding, writing-original draft, reviewing and editing. Alpana: conceptualization, coding, validation, supervision, draft-editing and reviewing. Abhishek Roy: conceptualization, draft-writing, reviewing and editing, supervision.

## References

- [1] Makridakis, S. (2017). The forthcoming artificial intelligence (ai) revolution: Its impact on society and firms. *Futures*, 90, 46-60. doi:10.1016/j.futures.2017.03.006
- [2] Hirczy de Mino, W. (2019). How algo Westlaw and Algo Google Scholar failed THE Rohrmoos test, and how to IMPROVE case Law citation with or without artificial intelligence. *SSRN Electronic Journal*. doi:10.2139/ssrn.3416668
- [3] Sil, R., Roy, A., Bhushan, B., & Mazumdar, A. K. (2019). Artificial Intelligence and Machine Learning based Legal Application: The State-of-the-Art and Future Research Trends. 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS).<https://doi.org/10.1109/icccis48478.2019.8974479>
- [4] Saha, D., Sil, R., & Roy, A. (2020). A study on implementation of text analytics over legal domain. *Evolution in Computational Intelligence*, 561-571. doi:10.1007/978-981-15-5788-0\_54
- [5] Ali, J. H. (2020). Forensic linguistics: A study in criminal speech acts. *Beni-Suef University International Journal of Humanities and Social Sciences*, 2(1), 39-65. doi:10.21608/buijhs.2020.91333
- [6] Husa, J. (n.d.). Comparative law, legal linguistics and methodology of legal doctrine. *Methodologies of Legal Research : What Kind of Method for What Kind of Discipline?* doi:10.5040/9781472560896.ch-011
- [7] Epstein, L., & Martin, A. D. (2010). Quantitative approaches to empirical legal research. *Oxford Handbooks Online*. doi:10.1093/oxfordhb/9780199542475.013.0038
- [8] Nielsen, L. B. (2010). The need for multi-method approaches in empirical legal research. *Oxford Handbooks Online*. doi:10.1093/oxfordhb/9780199542475.013.0040
- [9] Derlln, M., & Lindholm, J. (2017). Serving two MASTERS: CJEU case law in Swedish first instance courts and national courts of precedence as gatekeepers. *SSRN Electronic Journal*. doi:10.2139/ssrn.2952783
- [10] Johns, F., & Grolman, L. (2019). Big data and international law. *International Law*. doi:10.1093/obo/9780199796953-0176
- [11] ŠADL, U., & OLSEN, H. P. (2017). Can quantitative Methods Complement Doctrinal legal Studies? Using CITATION network and Corpus linguistic analysis to understand international courts. *Leiden Journal of International Law*, 30(2), 327-349. doi:10.1017/s0922156517000085
- [12] Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PLOS ONE*, 12(4). doi:10.1371/journal.pone.0174698
- [13] Richter, H. (2018). The behavior of federal JUDGES. a theoretical and empirical study of rational choice. *Rabels Zeitschrift Für Ausländisches Und Internationales Privatrecht*, 78(4), 932. doi:10.1628/003372514x683800
- [14] Rachlinski, J. J., & Wistrich, A. J. (2017). Judging the judiciary by the numbers: Empirical research on judges. doi:10.31228/osf.io/cr3vz
- [15] Frankenreiter, J. (2018). Are advocates General Political? An empirical analysis of the voting behavior of the Advocates general at the European Court of justice. *Review of Law & Economics*, 14(1). doi:10.1515/rle-2016-0037
- [16] Spence, D. B., & Murray, P. (1999). The law, economics, and politics of federal Preemption Jurisprudence: A quantitative analysis. *California Law Review*, 87(5), 1125. doi:10.2307/3481041
- [17] Hunter, C., Nixon, J., & Blandy, S. (2008). Researching the judiciary: Exploring the invisible in judicial decision making. *Journal of Law and Society*, 35(S1), 76-90. doi:10.1111/j.1467-6478.2008.00426.x
- [18] Van Helden, A. (2019). De provinciale verordening. *RegelMaat*, 34(4), 253-272. doi:10.5553/rm/0920055x2019034004002
- [19] Bricker, B. (2017). Breaking the principle of Secrecy: An examination of Judicial dissent in the European constitutional courts. *Law & Policy*, 39(2), 170-191. doi:10.1111/lapo.12072
- [20] Ikram, A. Y., & Chakir, L. (2019). Arabic text classification in the legal domain. 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS). doi:10.1109/icds47004.2019.8942343
- [21] Dyeve, A. (2016). The promise and pitfalls of automated text-scaling techniques for the analysis of judicial opinions. *SSRN Electronic Journal*. doi:10.2139/ssrn.2739692
- [22] Bricker, B. (2017). Breaking the principle of Secrecy: An examination of Judicial dissent in the European constitutional courts. *Law & Policy*, 39(2), 170-191. doi:10.1111/lapo.12072
- [23] Doron, I. (., Totry-Jubran, M., Enosh, G., & Regev, T. (2015). An American friend in an ISRAELI COURT: An empirical perspective. *Israel Law Review*, 48(2), 145-164. doi:10.1017/s0021223715000047

- [24] Bricker, B. (2017). Breaking the principle of Secrecy: An examination of Judicial dissent in the European constitutional courts. *Law & Policy*, 39(2), 170-191. doi:10.1111/lapo.12072
- [25] Vols, M., Tassenaar, P., & Jacobs, J. (2015). Anti-social behaviour and European protection against eviction. *International Journal of Law in the Built Environment*, 7(2), 148-161. doi:10.1108/ijlbe-08-2014-0021
- [26] Van Dijck, G. (2016). Victim oriented tort law in action: An empirical examination of catholic church sexual abuse cases. *SSRN Electronic Journal*. doi:10.2139/ssrn.2738633
- [27] Bruijn, L. M., Vols, M., & Brouwer, J. G. (2018). Home closure as a weapon in the Dutch war on DRUGS: Does judicial Review function as a safety net? *International Journal of Drug Policy*, 51, 137-147. doi:10.1016/j.drugpo.2017.08.003
- [28] Garoupa, N., Gili, M., & Gómez-Pomar, F. (2012). Political influence and Career Judges: An empirical analysis of administrative review by the Spanish Supreme Court. *Journal of Empirical Legal Studies*, 9(4), 795-826. doi:10.1111/j.1740-1461.2012.01270.x
- [29] Behn, D., & Langford, M. (2017). Trumping the environment? An empirical perspective on the legitimacy of investment treaty arbitration. *The Journal of World Investment & Trade*, 18(1), 14-61. doi:10.1163/22119000-12340030
- [30] Holá, B., Bijleveld, C., & Smeulders, A. (2012). Consistency of international sentencing: ICTY and ICTR case study. *European Journal of Criminology*, 9(5), 539-552. doi:10.1177/1477370812453112
- [31] Tarissan, F., & Nollez-Goldbach, R. (2014). The network of the International criminal court decisions as a complex system. *Emergence, Complexity and Computation*, 255-264. doi:10.1007/978-3-642-45438-7\_25
- [32] Tarissan, F., & Nollez-Goldbach, R. (2016). Analysing the first case of the International criminal court from a Network-science perspective. *Journal of Complex Networks*. doi:10.1093/comnet/cnw002
- [33] Infantino, M. (2021). The Italian Constitutional court in the European Space: An empirical approach. *Dialogues on Italian Constitutional Justice*, 221-243. doi:10.4324/9781003051121-18
- [34] Frankenreiter, J. (2016). The politics of citations at the Ecj: Policy preferences of EU member state governments and the CITATION behavior of members of the European Court of justice. *SSRN Electronic Journal*. doi:10.2139/ssrn.2778818
- [35] Bruinsma, F. J., & De Blois, M. (1997). Rules of law from Westport TO Wladiwostok. separate opinions in the European Court of human rights. *Netherlands Quarterly of Human Rights*, 15(2), 175-186. doi:10.1177/092405199701500204
- [36] Webb, T. E. (2019). *Hirst V United KINGDOM* [2005] ECHR 681, European Court of human RIGHTS (Grand Chamber). *Essential Cases: Public Law*. doi:10.1093/he/9780191868306.003.0014
- [37] Olsen, H. P., & Küçüküsu, A. (2017). Finding hidden patterns in Ecthr's case law. *International Journal of Discrimination and the Law*, 17(1), 4-22. doi:10.1177/1358229117693715
- [38] Madsen, M. R. (2017). Rebalancing European Human Rights: Has the BRIGHTON DECLARATION engendered a New deal on human rights in Europe? *SSRN Electronic Journal*. doi:10.2139/ssrn.2993222
- [39] Shulayeva, O., Siddharthan, A., & Wyner, A. (2017). Recognizing cited facts and principles in legal judgements. *Artificial Intelligence and Law*, 25(1), 107-126. doi:10.1007/s10506-017-9197-6
- [40] Frankenreiter, J. (2018). Are advocates General Political? An empirical analysis of the voting behavior of the Advocates general at the European Court of justice. *Review of Law & Economics*, 14(1). doi:10.1515/rle-2016-0037
- [41] Tarissan, F., & Nollez-Goldbach, R. (2016). Analysing the first case of the International criminal court from a Network-science perspective. *Journal of Complex Networks*. doi:10.1093/comnet/cnw002
- [42] Duarte d'Almeida, L., & Michelon, C. (2016). The structure of arguments by analogy in law. *Argumentation*, 31(2), 359-393. doi:10.1007/s10503-016-9409-3
- [43] Wyner, A., Mochales-Palau, R., Moens, M., & Milward, D. (2010). Approaches to text mining arguments from legal cases. *Semantic Processing of Legal Texts*, 60-79. doi:10.1007/978-3-642-12837-0\_4
- [44] Grabmair, M. (2017). Predicting trade secret case outcomes using argument schemes and learned quantitative value effect tradeoffs. *Proceedings of the 16th Edition of the International Conference on Artificial Intelligence and Law*. doi:10.1145/3086512.3086521
- [45] Ruppert, E., Hartung, D., Sittig, P., Gschwander, T., Rönneburg, L., Killing, T., & Biemann, C. (2018). LawStats – Large-scale German court DECISION evaluation using web Service classifiers. *Lecture Notes in Computer Science*, 212-222. doi:10.1007/978-3-319-99740-7\_14
- [46] Walzl, B., Bonczek, G., Scepankova, E., Landthaler, J., & Matthes, F. (2017). Predicting the outcome of appeal decisions IN germany's tax law. *Electronic Participation*, 89-99. doi:10.1007/978-3-319-64322-9\_8
- [47] Sil, R., & Roy, A. (2020). A novel approach on argument based legal prediction model using machine learning. *2020 International Conference on Smart Electronics and Communication (ICOSEC)*. doi:10.1109/icosec49089.2020.9215310
- [48] Manikandan, G., & Abirami, S. (2018). A survey on feature selection and extraction techniques for high-dimensional microarray datasets. *Knowledge Computing and Its Applications*, 311-333. doi:10.1007/978-981-10-8258-0\_14
- [49] Sil, R., Saha, D., & Roy, A. (2021). A Study on Argument-Based Analysis of Legal Model. *Advances in Intelligent Systems and Computing*, 449-457. https://doi.org/10.1007/978-3-030-73603-3\_42



- [50] Pavlyshenko, B. (2019). Machine-Learning Models for Sales Time Series Forecasting. *Data*, 4(1), 15. <https://doi.org/10.3390/data4010015>
- [51] Sil, R., Alpana, Roy, A., Dasmahapatra, M., & Dhali, D. (2021). An intelligent approach for automated argument based legal text recognition and summarization using machine learning. *Journal of Intelligent & Fuzzy Systems*, 1–10. <https://doi.org/10.3233/jifs-189867>
- [52] Buey, M. G., Roman, C., Garrido, A. L., Bobed, C., & Mena, E. (2018). Automatic legal document analysis: Improving the results of information extraction processes using an ontology. *Studies in Big Data*, 333-351. doi:10.1007/978-3-319-77604-0\_24
- [53] Alpana, & Chand, S. (2020). An intelligent technique for the characterization of coal microscopic images using ensemble learning. *Journal of Intelligent & Fuzzy Systems*, 38(5), 6257-6267. doi:10.3233/jifs-179707
- [54] Roy, P., Chakraborty, R., Chowdhuri, I., Malik, S., Das, B., & Pal, S. C. (2020). Development of Different Machine Learning Ensemble Classifier for Gully Erosion Susceptibility in Gandheswari Watershed of West Bengal, India. *Machine Learning for Intelligent Decision Science*, 1–26. [https://doi.org/10.1007/978-981-15-3689-2\\_1](https://doi.org/10.1007/978-981-15-3689-2_1)
- [55] Kadkhodaei, H. R., Moghadam, A. M., & Dehghan, M. (2020). HBoost: A heterogeneous ensemble classifier based on the Boosting method and entropy measurement. *Expert Systems with Applications*, 157, 113482. <https://doi.org/10.1016/j.eswa.2020.113482>

Cell: +91-9476134119  
Email: dr.aroy@yahoo.com.

## Author Biographies

**First Author:** Riya Sil has done her M.tech (Computer Science & Engineering) from Birla Institute of Technology, Mesra, Ranchi, India and is currently pursuing her Ph.D in Computer Science & Engineering from Adamas University, Kolkata, India. Her research interests include legal analytics, machine learning and artificial intelligence. She is presently working as an Assistant Professor in Department of Computer Science and Engineering, Adamas University, Kolkata, India.  
Email: riyasil1802@gmail.com.

**Second Author:** Alpana has completed her Ph.D in Computer Science & Engineering from Jawaharlal Nehru University, New Delhi, India and M.Tech (Information Technology) from Birla Institute of Technology, Mesra, Ranchi, India. Her area of interest includes image processing, artificial intelligence and machine learning. She is presently working as an Assistant Professor in Department of Computer Science and Technology, Manav Rachna University, Faridabad, Haryana, India.  
Email: alpana.srk@gmail.com.



**Third Author:** Abhishek Roy has completed his in Ph.D. (Computer Science), M.Sc. (Computer Technology) and B.Sc. Information Technology Honors from The University of Burdwan, West Bengal, India. His research area includes Artificial Intelligence, Machine Learning and Data Analytics. He is serving as resource person to various international and national research societies, conferences and journals. Currently he is serving in the capacity of Associate Professor and Ph.D. Supervisor (Technology) at Department of Computer Science and Engineering under School of Engineering and Technology at Adamas University, Kolkata 700126, India. For further details, please contact at: