An Artificial Intelligence based Analysis in Legal domain

V.Vaissnave, P.Deepalakshmi

Abstract: Around the world, legitimate information and common laws are available in raw form, but hard to understand and not in organized form. All legitimate information is nowadays computerized since the legal information gets generated on a regular basis in a huge volume due to increase of maritime (law) courts. The automation tool to analyse this legal data can serve effectively for lawyers and law students, which can address a lawyer's role and can even become powerful to release such a role in future. The machine learning and deep learning algorithms-based analysis systems apply these methods mainly for document classification. Legal document translation, text classification, summarization, data forecasting and data obtainment are part of the goals got from research charity. In this study, we review about the different methods of deep learning used in legal tasks such as Legal data search, Legal document analytics, and Legal perspective interface. To solve aggregate tasks, one can use the deep learning methods like, Recurrent Network Networks (RNN), Gated Recurrent unit network (GRU), Long Short Term Memory networks (LSTM), Convolutional neural network (CNN). Through this review, we instituted that deep learning models are giving advanced performance.

Keywords: Legal system, Deep learning, Data analytics, Classification, Prediction, Artificial Intelligence

I. INTRODUCTION

1.1 Artificial Intelligence

Artificial Intelligence (AI) is the facility of a computer program otherwise machine to think and experience. It is a knowledge domain and makes computers intelligent. Most of the things that can be done by humans require intelligence. The AI implementations are commonly used for medical science, education field, automobile industry, protection, cultivation, power engineering, life sciences, commercial, craft, and judgement. Role of deep learning involves AI models that create good results based on the artificial neural network in the implementation sector.

Using the Natural Language Processing (NLP) models, we could read millions of files loaded into the system, and the systems learns relationships between the strings from the documents, and classify single sentences from the long paragraph. AI predicts that the categories are like prediction technology, legal analytics, document automation, and intelligent interfaces in legal industry.

The legal adviser can pick-up any case in the world after AI's introductory in legal domain. For a legal adviser, it takes lots of time to do analysis manually, but AI can help to get quick response [17]. The recent survey indicated that accuracy of an adviser prediction was nearly 60%, but AI techniques can perform successfully with 80% or more accuracy. Average prediction time required for a legal adviser is nearly one hour, and AI can take 26 seconds for completing the same task.

Many organizations have engaged themselves in developing software products which mainly concentrate on extracting information from the legal documents. The availability of official papers are in arithmetical structure. Intelligent models and applications are developed based on the intense logic from research area. Different algorithms and approaches have been developed for facing this problem. It involves more number of staff, effort, and the high cost of time, so the automation represents best savings in business world over 30 years. The legal system issues are, including translation, classification, text analytics, reasoning and summarization. Artificial intelligence is the computer based intelligent support for legal research community; via automation it can reduce the time and human task.

The inclusion of different Machine learning, deep learning techniques are most important in this process. Highly relevant tasks are translation, classification, contract reviews, and summarization in public legal area. Deep learning, a subdivision of machine learning is also a particular part of Artificial Intelligence. Deep learning involves, several structures of non-sequential processing, and it's a form of hierarchical learning for high level abstractions in data. Deep learning technique can be applied through supervised and unsupervised methods in legal tasks. Unsupervised methods generate word embeddings from non-systematic data and it brands to achieve the advanced results. The AI technology assists not only the experimental but learns what law is, where the complication occurs, how it should be cleared up, which technology have to be used to solve issues, and how the arithmetic data could help for users.

* Correspondence Author
V.Vaissnave*, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, India. Email: vaissnave@gmail.com
P.Deepalakshmi, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, India. Email: deepa.kumar@klu.ac.in

An Artificial Intelligence based Analysis in Legal domain

Artificial Intelligence involves more data and work speed for human lawyers, and it shows that cooperation of human and AI. Efficient technologies with portable designs are provided for different experimental environments. Recent works of the legal domain includes applying a variety of common machine learning techniques and suggesting future event regulation.
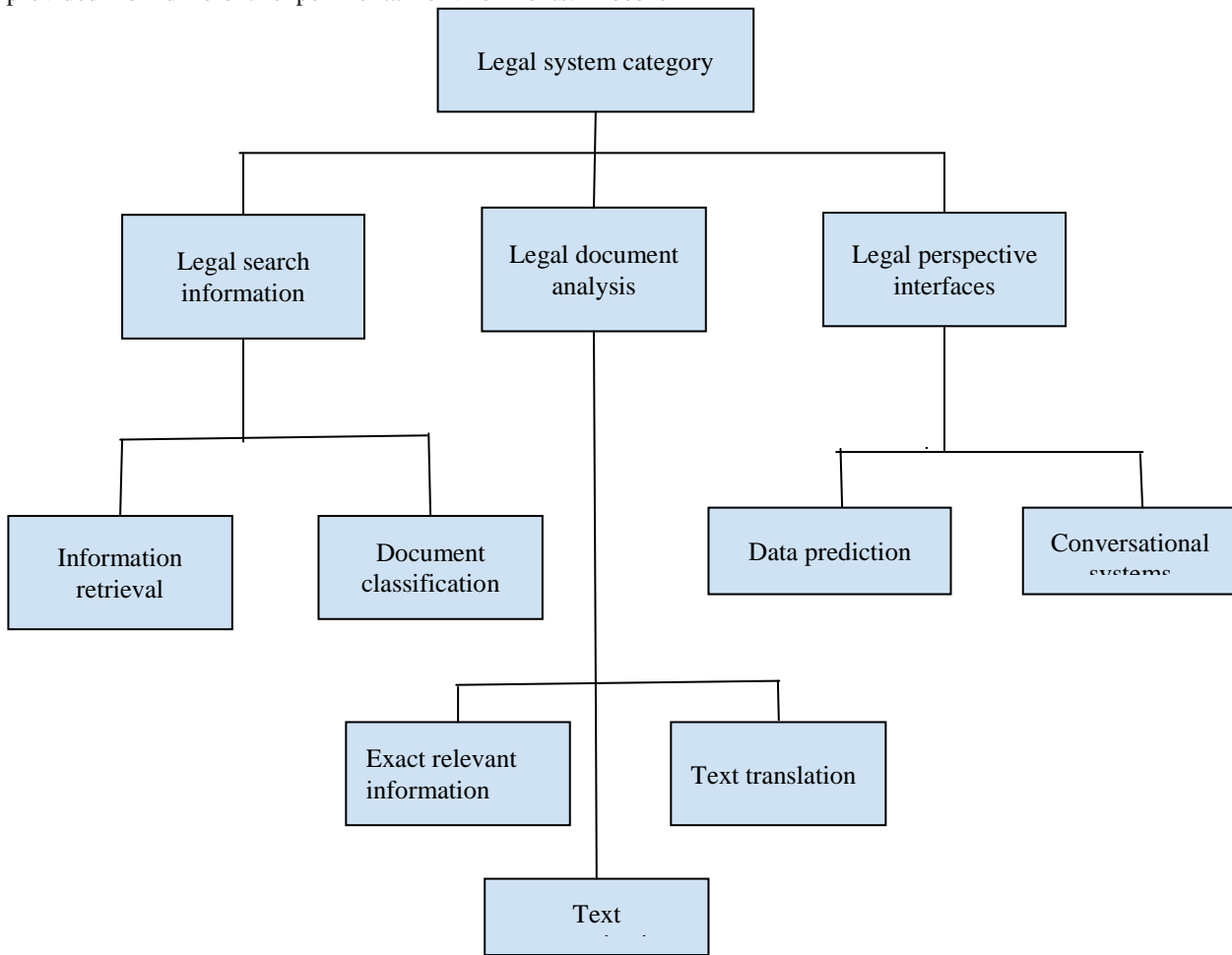

Fig. 1. Legal Domain categories

II. BACKGROUND

Here, we present a brief discussion about various legal task applications that have been implemented by deep learning algorithms, to the legal domain which can be divided into 3 categories; 1. Legal information search, 2. Legal document analytics, 3. Legal perspective interfaces as shown in Fig.1. The first list defines developed systems used to retrieve and classify the applicable document. The second list defines the Natural Language Processing (NLP) analysis such as legal text translation, classification, case prediction, text extraction from document, and summarization. The third one assists the tasks like conversational systems and judgement prediction systems.

III. LITERATURE REVIEW

Hereby, we represent the survey of deep learning techniques for legal framework published in, Journal of Machine learning, Journal of AI in Law, International Conference on AI Applications and Innovations, Conference on knowledge discovery and data mining, IEEE International Conference on Big Data. We searched the integration of keywords from legal system, document classification, legal data, neural networks, deep learning, and AI. Our review related papers are mainly published from 2017 to 2019, and we found a total of 50 papers.

After going through all the papers, here we limited our survey to 15 papers for this legal data survey.

A. Legal search data

Sugathadesa. K. et al. [1] described a legal document retrieval using document embeddings and deep learning. Here, the authors developed three major models, that is vector space representations of legal system by using Node2Vec algorithm for first model, sentence similarity for second model, and vector space for the third model. The three models have been implemented using Neural networks. The authors concluded that the group models showed higher accuracy level, and the approach could be used for developing retrieval data in other field of legal domain. But the traditional data search systems do not match the related string, like synonyms and related any other terms, just extract exact information of string. Landthaler et al. [2] described the data retrieval for legal system that searched not only exact information of single document or document collections, but retrieved some related patterns for search query.

The word embeddings are developed using

word2vec algorithm and they suggested for future work with other natural language pre-processing steps in word embedding, like stemming, stop word removal, part of tagging to achieve high performance and reliability. Undavia et al. [3] developed legal document classification via automation. Here they compared the machine learning algorithms with Neural Network algorithms, RNN and CNN.

Finally they concluded that word2vec and CNN algorithms are better algorithms giving best accuracy, nearly 72%. The popular machine learning algorithms used for text classification in legal domain are, logistic regression (LR), and support vector Machines(SVM). The authors, Wei et al. [4] reported the comparison of deep learning methods and SVM algorithm for four data sets, among which CNN is stated as a powerful tool which can produce better results for large volume of data sets. Silva et al. [5] classified the Brazilian court system documents from the PDF datasets and conveyed all challenges during the classification process. Here, the authors used simple CNN method and obtained an accuracy of 90% on this original datasets.

Kaiz Merchant et al [20] proposed automated text summarization system that can extract the necessary information from very long documents for civil and criminal cases using the NLP technique, semantic analysis. LSA is an automated unsupervised skill that can examine the document silent relationship among words and sentences. LSA reduces the similarity of words, and reduced noise. They performed two approaches. First one is with a single document untrained for civil cases, and second, a multi-doc trained for criminal cases. Drawback of the paper is that it was not effective because their model root on only similar words, not for whole conception, & breaks the final flow. Dipankar Chakrabarti et al [18] developed Risk-o-matter framework to find the risk in legal domain. The framework helped to decide whether the paragraph would be affected by risk or not. This framework takes out the fallible passages and they achieved 91% accuracy for risk classification.Guangyi Xiao et al [22] Purpose of the paper described classification of chinse legal question using CNN. The major task was coarse grained classification and the side task was fine grained classification. The authors developed ranking classification in which coarse gram classification output precedes the fine grained classification input. CNN model achieved good accuracy results in Chinese legal question. Afnan Iftikhar et al. [21] developed the automation system for extracting the relevant information to be used by Named Entity Recognition (NER). Authors described 9 entities including person name, date, ref num, FIR no., case no., etc. They developed three supervised algorithms (1) CRF, (2) MaxEnt, (3) TNT. CRF algorithm performance is found better than the other two algorithms. Octavia-Maria S, et al [24] developed SVM classifiers to find the high accuracy in document classification for legal professionals using the French Supreme court and legal area. The optimum results are forecast law area (98%), forecast court decision (96%) and forecast time for case decision (87.5%)[24].

B. Legal document analytics

Elnaggar, A., et al. [6] proposed multi task learning approach to be used by the Multimodal algorithm. The algorithm was developed as a single deep learning model, for multi tasks in legal domain. The authors additionally suggested that this is an effective way to solve the data scarcity issues in the legal system. Wang. Y. [7] developed another

implementation, a novel model namely Word2sent, which can find similarity between the legal sentences through unsupervised machine learning. The advantage here is achieving a more in depth approach between the sentences.
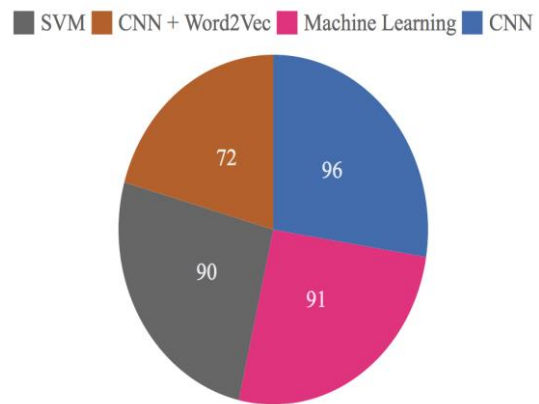

Fig2. Deep learning techniques for data classification

Li.S et al. [8] appraised the rationality of judicial decision to analyse the judgement logic depending on judgement results. A novel LSTM tool is used to confirm the reality of the case judgements. Son, N.T et al. [9] described that, there are four models based on RNN approach, that includes LSTM, Bidirectional LSTM, and the combination of LSTM-CRF, BI-LSTM-CRF. The NN mechanism resolves the task of recognising logical part of the legal documents or text. This mechanism achieved the best and most modern performance. Chalkidis et al. [10] described contract element extraction system that every organization having law form for contract, based on year payment and contain some terms and conditions to handle the legal problems. In case, the rule is changed, the contract would be expired. So here, the author developed Bi-LSTM model on token embedding, word, POS. For getting further improvement in results, they suggested adding additional LSTM on top of Bi-LSTM or adding CRF layer on the top of LSTM. Neill et al. [11] described some of the empirical techniques and extraction methods for single and multi-document summarization. Authors collected the legal documents that contain acts, directives, and regulations from the UK courts. Automatic text summarization consists of 2 steps : 1. Train the legal word embeddings using word2vec technique. 2. Select the CNN classifiers for document summarization. In this paper, authors improved the document summarization using CNN technique by focusing on text representation, sentence ranking, and sentence extraction for document summarization

C. Legal perspective interface

The author John et al.[13] developed a legal boat system using seq2seq conversational model. The machine responded to user questions and for retrieval method, the authors proposed a "productive model".

The model makes use of generating exact answer

of specific queries and the model was trained, the data come behind ques-ans pattern by using our dataset on legal issues. The authors concluded that the performance could be improved by more dataset deliver to the model.. P.K., Nguyen et al.[14] developed two concepts, information retrieval, ques-ans.

Lawyers can ask any kind of questions, which will be searched in the related documents by applying Ranking SVM and CNN algorithms. Hashianov et al,[16], QA system present legally reliable, found on NLP and IR.QA system educate the large documents of ques-ans pairs nearly 200k QA pairs. Lawyers learn answers for their questions using legal search engine that the name was Westlaw. The logical architecture

contained 4 components 1. Question analysis, 2. Query generation, 3.Ans analysis, 4.ques-ans pair score. GayleMcElvain et al.[19] developed bag of days in earlier days, but it offered low accuracy. So, here they implemented Bi-GRU based decision system. Developing RNN is used to find the original face of the criminal case, and traditional mechanisms were used to compare them against the legal distribution. Finally output model shows if the person is accused or not. Dipankar Chakrabarti, [18] implemented predicted possible decision and gave necessary information about law cases.The expert system "Robert Lawyer" provide advanced technology for legal users in which CNN algorithm is in answering to complex search questions fed by the lawyers.

Table 1. CATEGORIES OF LEGAL TASK, METHODS, DATASET & RESULT

Deep Learning categories	Sub category	Algorithm and Methods	Dataset	Results
Legal search data	Document retrieval	Node2Vec algorithm with Neural network. [1]	Got from the findlaw website 2500 law cases	NA
		Word embeddings. [2]	EU-Data production Directive, German rental contracts	NA
	Document classification	CRF,MAXEnt,TNT[20]	Lahore high court dataset	TNT was outperformed the other 2 algorithm.
		Recurrent neural network and Convolutional neural network[3]	University School of Law Supreme Court Database (SCDB)	CNN with Word2Vec= 72%
		Convolution neural network[4]	Real datasets four	CNN out performed SVM
		Word embedding with Neural network.[5]	Brazilian court dataset	90%
		Supervised learning algorithm (Risk-o-matter) framework.[18]	Largest dataset	91% accuracy.
		CNN[21]	Chinese legal dataset	CNN model got good accuracy.
		SVM[24]	French Supreme court	average 96% accuracy.
		Active learning algorithm(Top ranked & uncertain).[22]	1 lac to 5 lac documents	Top ranked 75%, uncertain 90% optimum performance.
Legal document analytics	Text translation, Summarization and classification.[6]	Multi task deep learning algorithm[6]	German legal corpora	Multi task deep learning better than modern algorithms
	Automated text summarization.	NLP technique(latent semantic analysis[19]	civil & criminal cases	LSA achieves 0.58 score.
	Extract relevant data	Word2sent algorithm. [7]	Lawsents (github)	Word Embedding results were better.
		LSTM Tool. [8]	24,000 case judgements from china online judgements	Artificial validation results good
		LSTM, BI-LSTM, LSTM-CRF, BILSTM-CRF. [9]	Vietnam legal dataset.	Good enough results.
		BI-LSTM model. [10]	3500 English contracts	By adding LSTM or CRF for improving results.
		ANN and Non-ANN models, with handle semantic modal.[11]		Results were improved by LSTM model.
Legalperspective interfaces	Conversational agent	Seq2Seq deep learning "productive model" [12]	Dataset build from ques-ans on some legal issues	Author suggested by performance could be improved by more dataset deliver to the model
		Convolutional Neural Network with Ranking SVM [13]	Legal dataset from Japon court.	Structured relationship between sentences and references between articles can be explored to improve upon results
		Westlaw search engine[19]	200k ques-ans pair.	NLP, structured semantic searches give better answers.
	Judgment prediction system	Traditional deep learning and Bi-directional GRU. [14]	Thai Supreme court criminal cases.	For better results Naïve Bayes and SVM
		SVM,CNN algorithm.[16]	Russia legal cases.	AI advanced technology provide good results

		NLP technique[25]	Singapore supreme court.	Recent NLP got higher accuracy
		SCALE methodology.[23]	1.8 million sentences	Reduce the manual effort.

Rishi Chhatwal et.al [12] developed Active learning, used to applied in the legal domain. The paper stated that active learning is superior methodology as well as a successful method in legal industry. Here they used three approaches namely (1)top ranked (2)uncertain (3)random sampling. Top ranked approach got excellent discharge at 75 and 90% recalls. When comparing top ranked with uncertain approach, uncertain approach got 90% cells. Random sampling was powerful method, when compared with top ranked. K.Brantinget al [23] described that legal decision support for magistrate are more compatible and systematic. Here they developed attention network for forecast, attention weight which can be used to highspot the important text. In this approach, they couldn't get accurate results so they implemented Semi supervised Case Annotation for Legal Explanations (SCALE). K.Jerrold et.al [25] described that comparative study about the machine learning techniques in legal domain for Singapore supreme legal court cases. They classified the legal documents and compared NLP methods with conventional methods for lengthy documents and could prove that recent NLP techniques can offer more accuracy in legal area.

IV. MAJOR FINDINGS AND SUGGESTIONS

Our objective in writing this article is to decide the best practice for applying automated document classification to legal texts with the aim of easy judge knowledge in their classification of law court records. We discussed different methods and tools for legal classification and summarization. The legal document classification has some particular issues like document structure, different aspects and assessments for the task. Even Though the legal summarization results are more competitive and comparable. In similar cases the multi document classification provide overall brief view of particular court cases for lawyers. In the multitasking, the experiments lead to two models, multimodal light version and base version. These models can be performed by different input values. In the first model, they used low input data and in the second model, they used bit high input data and finally find the percentage of score. This method can overcome the data security problem in legal system. However, the procedure need to test it on other tasks, datasets, languages and training combinations. The provided datasets could be used to test the approach on the remaining languages.

Robot Lawyer system, aims to help participants of the legal process, Russian Court judgement in criminal law cases. The system applies advanced technologies in the field of artificial intelligence to process and search information necessary for users.

V. CONCLUSION

The purpose of applying AI techniques and deep learning methods is to increase the overall performance of legal systems. The deep learning methods are in various tasks such as, legal data search, information retrieval, document classification, text translation, extract the relevant data, and conversational agents. From this review, we derived following outcomes. Legal data search, retrieval information applications are implemented by building domain specific (word2vec) word embeddings, and the document classification methods were majorly enhanced using convolutional neural networks. To work on perspective legal systems, implementing productive models from deep learning has provided the best results. Legal document analytics, relevant data extraction, data translation and summarization are performed mostly by recurrent neural network models and LSTM tools. Various countries developed legal datasets using deep learning intelligence and improved legal domain. Finally we conclude by saying that the implementation of Artificial Intelligence in legal framework has increasing rate in last 3 years and AI improves quality of documents, saves time, cost, assists other issues quickly and efficiently.

REFERENCES

[1]. Sugathadasa, K.,etal.:"Legal document retrieval using document vector embeddings and deep learning". In:Arai,K.Kapoor,S.,Bhatia,R.(eds)SAI 2018.AISC, vol.857,pp.160-175. Springer, Cham(2019).

[2]. Landthaler. J., Waltl,B.,Holl.P.,Mathes,F.":Extending full text search for legal document collections using word embeddings".In:JURIX, pp.73-829(2016)

[3]. Undovia,S.,Meyers,A.,J.E: "A Comparative study of classifying legal documents with neural networks". In:Federated Conference on Computer Science and Information Systems. (FedCSIS),pp. 515-522.IEEE,Poland (2018).

[4]. Wei,F.,Qin,H.":Empirical study of deep learning for text classification in legal document review". In: IEEE International Conference On Big Data(Big Data), pp.3317-3320. IEEE,USA(2018).

[5]. Da Silva, N.C.:Document type classification for Brazil's supreme court using a Convolutional Neural Network. In: The Tenth Conference on Forensic Computer Science and Cyber Law-ICoFCS, pp. 7-11.Brazil(2018).

[6]. Elnaggar, A., Gebendorfer,C.,Glaser,I., Matthes, F.: "Multi-task classification. Multi-task deep learning for legal document translation, summarization", andarXiv preprint arXiv:1810.07513(2018).

[7]. Wang, Y.: "An unsupervised approach to relatedness analysis of legal language". Master's thesis, University of Waterloo(2018).

[8]. Li, S.,Zhang, H., Ye,L.,Geo,X.,Fang,B.:"Evaluating the rationality of judicial decision with LSTM-based case modeling" In.: IEEE Third International Conference On Data Science in Cyberspace(DSC),pp.392-397,IEEE, China (2018).

[9]. Son, N.T., Nguyen, L.M., Quoe, H.B., Shumazu, A.:" Recognizing logical parts in legal texts using neural architectures". In: Eighth International Conference on knowledge and Systems Engineering (KSE), pp. 252-257. IEEE, Vietnam (2016)

[10]. Chalkidis, I., Androutsopoulos, I: "A deep learning approach to contract element extraction". In: JURIX, pp. 155-164 (2017)

[11]. Neil, J.O., Buitelaar, P., Robin, C., Brien, L.O.: "Classifying sentential modality in legal language: a use case in financial regulations acts and directives". In: Proceedings of the 16th Edition of the International Conference on Artificial Intelligence and Law, pp.159–168. ACM, USA (2017).

[12]. Rishi chhatwal ESQ legal AT & T services Inc. "Empirical evaluation of Active learning strategies in legal document review" 2017 IEEE International conference on Big Data.(BIGDATA).

[13]. John, A.K., Di Caro, L., Robaldo, L., Boella, G.: "Legalbot: a deep learning-based conversational agent in the legal domain". In: Frasincar, F., Ittoo, A., Nguyen, L.M., Métais, E. (eds.) NLDB 2017. LNCS, vol. 10260, pp. 267–273. Springer, Cham (2017). https://doi. org/10.1007/978-3-319-59569-6_32 18.

[14]. Do, P.K., Nguyen, H.T., Tran, C.X., Nguyen, M.T., Nguyen, M.L.: Legal question answering using ranking SVM and deep convolutional neural network. arXiv preprint arXiv: 1703.05320 (2017).

[15]. Kowsrihawat, K., Vateekul, P., Boonkwan, P.: Predicting judicial decisions of criminal cases from thai supreme court using bi-directional GRU with attention mechanism. In: 5th Asian Conference on Defense Technology (ACDT), pp. 50–55. IEEE, Vietnam (2018).

[16]. Hashianov, A., Alomava, I., Nurhumbetvo, G.: lawyers intellectual tool for analysis of legal documents in Russian. (IC-AIAI) International Conference on Artificial Intelligence Applications and Innovations (2018)

[17]. https://interestingengineering.com/ai-vs-lawyers-the-future-of-artificial-intelligence-an.

[18]. Dipankar Chakrabarti Senior Member, "Use of Artificial Intelligence to Analyse Risk in Legal Documents for a Better Decision Support" Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference (Jeju, Korea, 28-31 October 2018)

[19]. GayleMcElvain∗ gayle.mcevlain@capitalone.com CapitalOne McLean,Virginia,USA:A "Non-factoid Question-Answering System for the Legal Domain"

[20]. Kaiz Merchant B.E in Computer Engineering Dwarkadas J. Sanghvi College of Engineering Mumbai, India merchant.kaiz@gmail.com,YashPande B.E in Computer Engineering Dwarkadas J. Sanghvi College of Engineering Mumbai, India yash3096@gmail.com. "NLP Based Latent Semantic Analysis for Legal Text Summarization".

[21]. AFNAN IFTIKHAR et al:"Information mining from criminal judgements of lahore high court".Received February 23, 2019, accepted April 21, 2019, date of publication May 8, 2019, date of current version May 20, 2019. Digital Object Identifier 10.1109/ACCESS.2019.2915352.

[22]. Guangyi Xiao et al: "Multitask CNN for classification classification of chinese legal question", The Fourteenth IEEE International Conference on e-Business Engineering.

[23]. K.BrantingB.WeissB.Brown "semi-supervised methods for explainable Legal Prediction". The MITRE Corporation McLean,VA,USA {lbranting,bweiss,bcbrown}@mitre. org.

[24]. Octavia-Maria S, et al:, "Exploring the Use of Text Classification in the Legal Domain" ASAIL 2017, June 2017, London, United Kingdom.

[25]. Jerrold shoLegal Area Classification: "A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments" Domains @agc.gov.sg @smu.ac.in.

AUTHORS PROFILE

Mr. V.Vaissnave is doing Research in the Department of Computer Science and Engineering, School of Computing, Kalasalingam Academy of Research and Education, Krishnankoil, Tamilnadu. she is pursuing his PhD in the area of legal domain classification using machine learning techniques.

Dr.P.Deepalakshmi is currently working as a Professor in Department of Computer Science and Engineering at Kalasalingam Academy of Research and Education (KARE), Virudhunagar, Tamilnadu, India. She is also serving as Dean, School of Computing. Her research interest includes Optimization Techniques, Network Routing, Distributed Computing, Network Security, Data Analytics, Machine Learning Techniques. She also takes care of KARE ACM student chapter as faculty mentor. Contact her at deepa.kumar@klu.ac.in