# LEVERAGING AI IN THE KENYAN JUDICIARY-A CASE FOR UTILIZING TEXT CLASSIFICATION MODELS FOR DATA COMPLETENESS IN CASE LAW META DATA IN KENYA'S EMPLOYMENT AND LABOR RELATIONS COURT



*IDRC GRANT / SUBVENTION DU CRDI : - DEVELOPING RESPONSIBLE AFRICAN AI POLICY ENVIRONMENTS: POLICY CENTRE II*

# Leveraging AI in the Kenyan Judiciary: A Case for Utilizing Text Classification Models for Data Completeness in Case Law Meta Data in Kenya's Employment and Labor Relations Court.

**Florence Ongojo[1], CIPIT, Joseph Theuri Gitonga[1], CIPIT, and Angeline Wairegi[1], CIPIT**

[1]Center for Intellectual Property and Information Technology Law (CIPIT), Strathmore University, Nairobi, KE 00100

Corresponding author: Angeline Wairegi (e-mail: awairegi@strathmore.edu).

This work was supported by International Development Research Center (IDRC) under Grant 109307 – 002.

## Abstract

AI has been revolutionary in improving different professional fields and sectors. In the legal sector, AI is utilized, in a number jurisdictions, for different purposes both at the bar and bench level. The study investigates the efficacy of an AI algorithm in completing missing data in digitized documents, i.e. how AI can be utilized to achieve data completeness of precedents in the judiciary through text classification in order to achieve an optimal foundational basis for the creation of data sets that will facilitate the utilization of AI for different purposes. The Employment and Labor Relations court is used as a case study. The study analyzed the efficacy of 5 text classifier models: passive aggressive, linear regression, decision tree, random forest, and support vector machine (SVM) model. The results obtained from the study show that text classification can be automated successfully using machine learning techniques to generate case metadata. The accuracy of the text classifier methods utilized in the study range between 82% and 98%. Despite the data limitations faced in this study, the good results recorded help increase confidence that advanced NLP techniques have matured enough to be applicable to legal text in the Kenyan Judiciary. Findings from the study suggest that the success rates of the text classifier techniques are not merely dependent on text content, but the context of this content is also a determining factor - the nature of the cases and the structure of the legal system play an important role in the performance of text classifier models.

**Keywords**: AI, Text Data Mining, Digitization, Data Completion, Text Analysis

## Introduction

The judiciary plays a pivotal role in the structure of society as the judicial arm of the Kenyan government. The judiciary upholds the rule of law, protects and enforces legal and human rights, enforces the contracts of a laissez-faire economy, defines and interprets the law, imposes and releases individual actors of legal responsibility, and is free from political interference.[1] It's adjudicating function primarily serves the public to ensure justice is dispensed in a timely and efficient manner. It is, primarily, in this manner that the judiciary interacts with the public, and, on a broader scope, its immediate stake holders such as prosecutors and defense lawyers.

Judicial services are grounded on the principles of independence and accountability. These principles are structured to ensure that the administration of justice is delivered in an efficient, effective, inclusive and transparent manner.[2] Over the years, the judiciary in Kenya has been keen on improving its service delivery and building public confidence. The adoption and implementation of technology has been one of the areas which it has embraced to improve its service delivery which, in turn, impacts the overall perception of the judiciary. Digitization of the judiciary is a core component in adopting technology. It has been characterized by the move away from paper base procedures and physical courts.[3]

In 2010, the promulgation of the Constitution of Kenya included in it a bill of rights that informed labor laws, the rights of laborers, and the task of the judicial system to uphold the rights under Article 41. It provided for every person's right to fair labor practices; fair remuneration; fair working conditions; the right to, form, join and participate in programs of a trade union and go on strike; the right of every employer to form and join an employer's organization and participate in its activities, and the right of every trade union and employer's organization to engage in collective bargain. These rights form the basis for the nature of cases that are heard under the Employment and Labor Relations Court (ELRC) structured under the constitution, reconstituting the post-colonial structure of the Industrial Courts.[4] The ELRC is a specialized court with High Court Status established under the section 162(2) (a) of the Constitution of Kenya. The court was instituted with the primary reason to ensure, equitable and progressive access to judicial services. The ELRC primarily has jurisdiction to handle disputes arising and relating to employment and labor relations.

---

[1] Hon. Justice A. O. Muchelule, 'Rebuilding Confidence in the Kenyan Judiciary: Civic Relation, Public Communication and Accountability of the Judiciary.'

[2] 'Enhancing Judicial Transparency and Promoting Public Trust.' (IDLO, 2018) https://www.idlo.int/news/speeches-and-advocacy/enhancing-judicial-transparency-and-promoting-public-trust

[3] Leveraging on Digital Technology in Administration of Justice (KIPPRA, July 2021) https://kippra.or.ke/leveraging-on-digital-technology-in-administration-of-justice/

[4] O.J. Ochieng, L.K Waithaka, 'Evolution Of Labour Law In Kenya' (International Journal of Law and Policy, 2019) https://www.iprjb.org/journals/index.php/IJLP/article/download/1014/1125/3317

Employment relationships are a legal relationship between the employer and employee; these relationships impact the labor force in the different economic sectors, i.e., finance, health, agriculture, education, transportation, manufacturing, hospitality, retail, energy and mining, media, building and construction, and real estate. It is from these sectors that cases arise and are adjudicated upon in the ELRC. In this way, laborers and employees alike become key stakeholders and have vested interests in the nature of disputes that arise in the ELRC.

This study investigates the viability of utilizing AI in the digitization process of documents in the ELRC by developing a system of completing missing data using pre - existing open source case records. The ELRC is a court with specialized jurisdiction; meaning, it is a court that was created for a specific purpose with limited jurisdiction in one particular field of the law.[5] The specialized nature of this court informs the reason for adopting it under this case study, as the parameters and issues for each case can be clearly identified and streamlined across all cases, clarifying the areas where data completeness is required through better classification.

The judiciary can leverage AI in different capacities, one of them being in an administrative capacity to expand the scope of reporting through a broader assessment of its courts in this context the ELRC. This can only be achieved, however, by ensuring already existing data from the court cases is quality data and complete. Quality data is the cornerstone of developing data sets that would successfully leverage AI - without the right data sets AI technologies have nothing to work with. Digitization and the adoption of technology gives access to data. Digital transformation and adoption of technology is a recognized strategy in the Kenyan Judiciary as early as 2005. The first ever Judiciary Strategic Plan of 2005- 2008 began the process of integrating digital technology in the administration of justice in Kenya. The strategy led to the introduction and launch of the Court Records Management System (CRMS) and Digital Audio Recording as key components to improving service delivery.

The strategic plan, developed in 2009, further fortified the infusion of digital technologies in the judiciary. It highlighted plans for introducing appropriate information and communications technology (ICT) facilities in the court system to increase the efficiency of judicial operations. For example, the National Council for Law Reporting (NCLR), a semi - autonomous state corporation under the judiciary responsible for providing public legal information, benefited from the introduction of ICT facilities. An ICT department was established with one of the core outputs being the design and implementation of a system for monitoring, collecting, and tracking judicial opinions delivered by the Court of Appeal and the High Court.[6] The 2012-2016 Strategic Plan: The Judicial Transformation Framework further identified the harnessing of digital technology to expedite the delivery of justice as one of the key pillars of Judicial transformation. The strategic

[5] Specialized Courts/ Divisions (Judicial Reform and Institutional Strengthening Project) < http://juristproject.org/specialized-courts
[6] NCLR Strategic Plan 2009-2012 (NCLR, 2009) http://kenyalaw.org/kl/fileadmin/pdfdownloads/NCLR_Strategic_Plan_2009_2012.pdf/

plan mandated the judiciary develop and deploy an electronic case management system, undertake the digitization of court records, and adopt audio visual recording and transcription for court proceedings. In 2014, the Integrated Court Management System Committee (ICMSC) was established. The primary objective of the committee was to actualize an efficient and effective Court Management System and advice on best ICT practices.[7] The committee developed the 2018-2022 ICT Master plan which ensured all ICT projects and developments within the Judiciary aligned to the priorities outlined in the Judiciary Strategic Plan 2014-2018, the SJT (2017-2021), and plans for automating the Judiciary processes.[8]

Digitization is an ongoing and continuous process. One of the major challenges, at present, is ensuring data completeness in the ongoing process of digitizing judiciary records. Digitization builds sets of structured sequences and interdependence that regulate the execution of organizational procedures and processes.[9] Efforts to ensure data completeness are worthwhile as complete and accurate data would enable the judiciary to better leverage technology such as AI for different functions within the Kenyan court system.

**The State of Technology in the Kenyan Judiciary[10]**

The Judiciary Strategic Plan points out that ICT has enormous potential to provide a quantum leap in the administration of justice. Properly harnessed and deployed, ICT infrastructure would improve efficiency and effectiveness of both back office and court processes. The master plan identified over 30 potential ICT projects. It prioritized 6 as flagship projects. These 6 projects are (i) the Court Case Management System (CMS), (ii) the Court Recording and Transcription Services (CTRS), (iii) the Judicial Integrated Financial Management System (JIFMS), (iv) the Human Resource and Performance Management, (v) the ICT Connectivity Infrastructure, and (vi) the Judiciary Intranet to improve internal communication.[11]

The Court Case Management System, which encompasses e-filing and case tracking functionalities, was officially launched in 2020 - prompted by the COVID -19 pandemic and the need to ensure continued functionality within the courts. E - filing facilitates the filing of court documents for proceedings from litigants electronically. The case tracking system is an automated system that tracks the life cycle of cases; it indicates the status of a case, generates the cause list, keeps an e-diary, and generates reports.  The case management system was piloted in the Nairobi,

---

[7] Integrated Court Management Systems Committee ( ICMSC), https://www.judiciary.go.ke/about-us/our-programmes/icms/

[8] ICT Master Plan:  Enabling Justice through ICT 2018- 2022) (Judiciary, 2018)https://repository.kippra.or.ke/handle/123456789/557

[9] A. Cordella, F. Contini, 'Digital Technologies for Better Justice.' (Inter-American Development Bank, April 2020 Discussion Paper)

[10] Leveraging on Digital Technology in Administration of Justice (KIPPRA, July 2021) https://kippra.or.ke/leveraging-on-digital-technology-in-administration-of-justice/

[11] ICT Master Plan:  Enabling Justice through ICT 2018- 2022) (Judiciary, 2018)https://repository.kippra.or.ke/handle/123456789/557

Eldoret, Machakos, and Mombasa law courts between 2011-2012. These systems were only fully adopted to the majority of the court stations in 2020 due to various challenges which included, but were not limited to: the lack of ongoing developer support due to human and financial resource constraints; turnover of judicial officers and staff due to transfers; competing priorities such as the management of case backlogs, and a lack of enough resources such as computers.[12] Challenges notwithstanding, the electronic case management system is currently in use in the Supreme Court, Court of Appeal, all High Court Divisions, and other subordinate courts in Nairobi. The judiciary plans to expand its use countrywide. Further, the judiciary is also implementing the digital court recording and transcription system in 32 out of the 132 courts in Kenya.[13] The State of the Judiciary and Administration of Justice Annual Report 2020-2021 reports 144,000 cases were heard through the virtual courts; 356,997 new cases were filed; 295,837 cases were heard and determined, and 1,359,768 cases were processed through the Case Tracking System (CTS).

**Current Utilization of Artificial Intelligence (AI) in the Kenyan Judiciary.**

The judiciary's digital transformation sets a good foundation for the overall adoption of new and emerging technologies such as Artificial Intelligence (AI). It has already started leveraging AI in some form, particularly, with the automation of its registry systems through the development, implementation, and use of the CTS. The case tracking system has been in use from the year 2017 and, currently, almost 90% of all active cases have been captured on the system and can be tracked. Judges and judicial officers use the system to access documents filed by litigants through the e-filing system. As of 2021, a total of 1,359,768 cases had been captured into the CTS.[14] Further, the Judicial Financial Management Information System has been deployed in all court stations. The system is used for management of court revenue (court fees and fines), court deposits and expenditure. It has been integrated with the Case Tracking System and has fully automated all the processes from court fees assessment, and e-receipting across all court of stations in the judiciary. Within the reporting period of 2020 -2021, 26 court rooms were installed with CRTS equipment; the CRTS is designed to automate courtroom proceedings through digital recording of court proceedings and provision of transcription services.[15]

**AI in other Jurisdictions**

Jurisdictions across the world have incorporated various technologies into their court systems.[16] While there may be indigenous factors explaining why specific technologies have been adopted in

---

[12] Case Management Assessment - Kenyan Judiciary (USAID, 2016) <https://pdf.usaid.gov/pdf_docs/PA00X5P4.pdf>

[13] Leveraging on Digital Technology in Administration of Justice (KIPPRA, July 2021) https://kippra.or.ke/leveraging-on-digital-technology-in-administration-of-justice/

[14] The State of the Judiciary and Administration of Justice Annual Report 2020-2021 https://www.judiciary.go.ke/resources/reports/

[15] The State of the Judiciary and Administration of Justice Annual Report 2020-2021 https://www.judiciary.go.ke/resources/reports/

[16] Finucan L, Sierra E and Rajesh N, *Smart Courts: Roadmap for Digital Transformation of Justice in Africa*, 21 August 2018, 3-23.

a given jurisdiction, there is seemingly a common motivation to improve the delivery of justice. AI appears to be the next frontier for these initiatives. AI is a recently emerging system that uses computers and big data as a basis to simulate human behavior. It imitates human thinking by learning massive data knowledge and using algorithms to reason and analyse data.[17] Consequently, AI is changing the way in which digital transformation is perceived as it promises to have grand benefits which would impact the existing social, economic and even cultural norms.

Specifically, AI offers judiciaries around the world opportunities from either an administrative perspective or a procedural perspective. Opportunities from an administrative perspective would include automating filing systems for easy and fast search and retrieval, automated case appointment schedules, and reporting. Opportunities created under the procedural perspective would include preliminary case analysis, document analysis, speech translation and transcription systems, case search, case prediction, automated docketing, and legal advice.[18] With the sophistication that AI offers, movements toward AI-related technologies in court systems globally can therefore be understood as a logical manifestation of the aforementioned motivation and as a reasonable evolution of more conventional technologies. Indeed, many court systems across the world have adopted some form of AI to leverage these opportunities.[19] While a number of examined initiatives are at a pilot or trial phase, they are still useful in showcasing prevailing use cases of AI globally.

**AI Application in the Employment and Labor Relations Court: Establishing Data Completeness through Automatic Text Classification.**

With the increasing availability of electronic documents and the rapid growth of the internet, the task of automatic categorization of documents has become the key method for organizing the information and knowledge discovery. Proper classification of e-documents, online news, blogs, e-mails and digital libraries need text mining, machine learning and natural language processing techniques to get meaningful knowledge. The aim of this study is:

i) to determine the viability and efficacy of applying techniques and methodologies used for text documents classification in competing digitized records in the ELRC, and
ii) to create awareness of some of the interesting challenges that remain to be solved, focused mainly on text representation and machine learning techniques.

In this study, we apply machine learning methods for automatic case metadata classification. In this regard, we first exert language pre-process in employment and labour datasets, and then we extract a feature vector for each case text by using feature weighting and feature selection algorithms. Afterwards, we train our classifier algorithms. In experiments, the algorithms show good results with, the performance ranging between 82% and 98%.

---

[17] K. Zhua, L. Zheng, 'Artificial Intelligence in the Judicial Field Operation Status and Countermeasure Analysis.' (Journal of Mathematical Problems in Engineering , 2021) https://www.hindawi.com/journals/mpe/2021/9017181/

[18] 'Artificial Intelligence in the Judiciary.' (The Legal State, 2021)

[19] Introduction to AI for Courts,' (JTC Resource Bulleting, 2020) https://www.ncsc.org/__data/assets/pdf_file/0013/20830/2020-04-02-intro-to-ai-for-courts_final.pdf

## Procedure for Text Classification

### *Automatic Text Classification*

Automatic text classification has always been an important application and research topic since the inception of digital documents.[20] Online text documents have provided a large knowledge and information pool. To properly utilize this information, systematic organization is required to facilitates ease of storage, searching, and retrieval of relevant text content for the needed application.[21] Text classification, also known as text categorization, is the process of assigning text to one or more predefined category labels according to its content. Text classification has been successfully used in domains such as topic detection, spam e-mail filtering, news text classification, web page classification, author recognition, and sentiment analysis.[22] Traditionally, text categorization is done by human experts. Considering the great number of texts available, manually classifying text documents is time-consuming, expensive, and even impossible; therefore, it is better to use automatic classification techniques.[23] In this regard, there are two main approaches to classify documents automatically: rule-based approach and machine learning approach. In the rule-based approach, a set of rules are written by human experts, and the classification process is done according to these rules. In machine learning approaches, a classifier is built by learning from some pre-classified documents.[24]

A number of methods have been discussed in literature for document classification. These techniques include: the naïve Bayes classifier, decision trees, nearest neighbour classifier, linear discriminant analysis (LDA), logistic regression, and neural networks

The main steps of text classification are:

i) text pre-processing (remove stop-words, stemming);
ii) feature extraction/selection (the Term Frequency-Inverse Document Frequency (TF-IDF));
iii) model selection,
iv) and training and testing the classifier

The initial pipeline input consists of some raw text data set. In general, text data sets contain sequences of text in documents as $D = \{X_1, X_2, ..., X_N\}$, where $X_i$ refers to a data point (i.e., document, text segment) with s number of sentences such that each sentence includes $w_s$ words

---

[20] Farhoodi, M., & Yari, A. (2010). Applying machine learning algorithms for automatic Persian text classification. *2010 6th International Conference on Advanced Information Management and Service (IMS)*, 318–323.
[21] Tang, L., Rajan, S., & Narayanan, V. K. (2009). Large scale multi-label classification via metalabeler. *Proceedings of the 18th International Conference on World Wide Web*, 211–220.
[22] Mostafavi, S., Pahlevanzadeh, B., & Falahati Qadimi Fumani, M. R. (2020). Classification of Persian News Articles using Machine Learning Techniques. Computer and Knowledge Engineering, 3(1), 73–81.
[23] Aghila, G. (2010). A Survey of Na$\backslash$" ive Bayes Machine Learning approach in Text Document Classification. *ArXiv Preprint ArXiv:1003.1795*.
[24] Ibid-116

with $l_w$ letters. Each point is labelled with a class value from a set of *k* different discrete value indices.[25] Then, a structured set is created for training purposes which calls this section's feature extraction. The dimensionality reduction step is an optional part of the pipeline; often used to reduce computational time and complexity. The most significant step in document categorization is selection of the classification algorithm. Another part of the classification pipeline is the evaluation step. The evaluation step is divided into two parts: prediction the test set and evaluating the model. In general, the text classification system contains four different levels of scope that can be applied:

a. **Document level**: In the document level, the algorithm obtains the relevant categories of a full document.
b. **Paragraph level**: In the paragraph level, the algorithm obtains the relevant categories of a single paragraph (a portion of a document).
c. **Sentence level**: In the sentence level, obtains the relevant categories of a single sentence (a portion of a paragraph).
d. **Sub-sentence level:** In the sub-sentence level, the algorithm obtains the relevant categories of sub-expressions within a sentence (a portion of a sentence)

### *A. Text pre-processing*

Features useful in text classification are simple vocabulary words, user-specified or extracted keywords, multi-words, or metadata. In text classification, the steps involved in feature reduction are mainly applied in pre-processing, e.g., stop-word removal, stemming, etc.[26] Text documents generally use words from a large vocabulary, however, not all words occurring in a document are useful for classification. Researchers utilize feature reduction techniques like TF-IDF, multi-word removal, or a combination of such techniques to refine document data.[27] TF-IDF is a statistical technique to evaluate the importance of a word based on its frequency of occurrence in the document and in its relevant corpus. Latent Semantic Indexing (LSI) and multi-word techniques are semantics-oriented techniques which attempt to overcome the two basic problems in classification: 'polysemy', one word having many distinct meanings, and 'synonymy', different words having the same meaning. The LSI technique tries to use the semantics in a document structure using SVD (Singular Value Decomposition) matrix manipulations. A multi-word is a sequence of consecutive words having a semantic meaning, for example, "Information Technology", "Delhi Public School"," Computer Engineering Department", "State Bank of India".

---

[25] Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. In *Mining text data* (pp. 163–222). Springer.
[26] Kim, S.-B., Han, K.-S., Rim, H.-C., & Myaeng, S. H. (2006). Some effective techniques for naive bayes text classification. *IEEE Transactions on Knowledge and Data Engineering*, *18*(11), 1457–1466
[27] Zhang, C. (2008). Automatic keyword extraction from documents using conditional random fields. *Journal of Computational Information Systems*, *4*(3), 1169–1180

Multi-words are useful in classification as well as disambiguation. Several methods are used to extract multi-words from text such as the frequency approach mutual information approach.[28, 29]

### B. Feature Extraction

In general, texts and documents are unstructured data sets. However, these unstructured text sequences must be converted into a structured feature space when using mathematical modelling as part of a classifier. First, the data needs to be cleaned to omit unnecessary characters and words. After the data has been cleaned, formal feature extraction methods can be applied. A common feature extraction technique is Term Frequency-Inverse Document Frequency (TF-IDF). The mathematical representation of the weight of a term in a document by TF-IDF is given by:

$$W(d,t) = TF(d,t) * \log(\frac{N}{df(t)})$$

Here N is the number of documents, and $df(t)$ is the number of documents containing the term $t$ in the corpus.

Another common technique is Word2Vec. The Word2Vec approach uses shallow neural networks with two hidden layers, continuous bag-of-words (CBOW), and the Skip-gram model to create a high dimension vector for each word. The Skip-gram model gives a corpus of words, $w$, a context, $c$.[30] The goal is to maximize the probability:

$$\arg\max_{\theta} \prod_{w \in T} [\prod_{c \in c(w)} p(c \mid w; \theta)]$$

where $T$ refers to Text, and $\theta$ is parameter of $p\ (c \mid w; \theta)$.

### C. Dimensionality Reduction

As text or document data sets often contain many unique words, data pre-processing steps can be lagged by high time and memory complexity. A common solution to this problem is simply using inexpensive algorithms. However, in some data sets, these kinds of cheap algorithms do not perform as well as expected. To avoid the decrease in performance, many researchers prefer to use dimensionality reduction to reduce the time and memory complexity for their applications. Using dimensionality reduction for pre-processing can be more efficient than developing inexpensive classifiers.[31]

---

[28]Zhang, W., Yoshida, T., & Tang, X. (2007). Text classification using multi-word features. 2007 IEEE International Conference on Systems, Man and Cybernetics, 3519–3524.

[29] Zhang, W., Yoshida, T., & Tang, X. (2008). TFIDF, LSI and multi-word in information retrieval and text categorization. *2008 IEEE International Conference on Systems, Man and Cybernetics*, 108–113.

[30] Goldberg, Y., & Levy, O. (2014). word2vec Explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. *ArXiv Preprint ArXiv:1402.3722*.

[31] Ibid- 118

### D. Algorithms

The most significant step in document categorization is selection of the classification algorithm. Some of the techniques utilized for text classification are detailed below:

#### a. Logistic Regression

The most important step of the text classification pipeline is choosing the best classifier. One of the simplest classification algorithms is logistic regression (LR) which has been addressed in most data mining domains.[32] LR is a linear classifier with decision boundary of $\theta^T x=0$. LR predicts probabilities rather than classes.[33] The goal of LR is to train from the probability of variable $Y$ being 0 or 1 given $x$. Let us have text data which is $X \in R^{n\times d}$. If we have binary classification problems, the Bernoulli mixture models function should be used [121] as follows:

$$L(\theta|x) = p(y \mid x; 0) = \prod_{i=1}^{n} sigm(x_i)^{yi}(1 - sigm(x_i))^{1-yi}$$

where $x_i\, \theta = \theta_0 + \sum_{j=1}^{d}(x_{ij}\theta_j)$, and *sigm(.)* is a sigmoid function which is defined as:

$$sigm(n) = \frac{1}{1 - e^{-n}} = \frac{e^n}{1 - e^n}$$

#### b. Naïve Bayes Classifier

In the earliest history of information retrieval as a feasible application, The Naïve Bayes Classifier (NBC) was very popular because it is computationally inexpensive and also needs a very low amount of memory.[34] If the number of documents ($n$) fit into $k$ categories where $k \in \{c1, c2, \ldots, ck\}$, the predicted class as output is $c \in C$. The Naïve Bayes algorithm can be described as:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)}$$

where $d$ is document and $c$ indicates classes.

#### c. K-Nearest Neighbours

Non-parametric techniques, such as k-nearest neighbour (KNN), have also been studied and used for classification tasks.[35] Given a test document $x$, the KNN algorithm finds the $k$ nearest neighbours of $x$ among all the documents in the training set and scores the category candidates based on the class of $k$ neighbours. The similarity of $x$ and each neighbour's

---

[32] Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, *151*, 147–160

[33] Genkin, A., Lewis, D. D., & Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, *49*(3), 291–304.

[34] Palacios-Alonso, M. A., Brizuela, C. A., & Sucar, L. E. (2010). Evolutionary learning of dynamic naive Bayesian classifiers. *Journal of Automated Reasoning*, *45*(1), 21-37

[35] Li, L., Weinberg, C. R., Darden, T. A., & Pedersen, L. G. (2001). Gene selection for sample classification based on gene expression data: Study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics*, *17*(12), 1131–1142.

document could be the score of the category of the neighbour documents. Multiple KNN documents may belong to the same category; in this case, the summation of these scores would be the similarity score of class *k* with respect to the test document *x*. After sorting the score values, the algorithm assigns the candidate to the class with the highest score from the test document *x*.[36] The decision rule use is given by:

$$f(x) = \arg\max_{j} S(x, C_j)$$
$$= \sum_{d_i \in KNN} sim(x, d_i)y(d_i, C_j)$$

where *S* refers to score value with respect to $S(x, C_j)$, the score value of candidate *i* to class of *j*, and output of $f(x)$ is a label to the test set document.

### d. Support Vector Machine (SVM)

Support Vector Machine (SVM) is another popular technique which employs a discriminative classifier for document categorization. This technique can also be used in all domains of data mining such as bioinformatics, image, video, human activity classification, safety and security, etc. This model is used as a baseline for many researchers to compare against their own works to highlight novelty and contributions. In the context of text classification, let $x_1, x_2, \ldots, x_l$ be training examples belonging to one class *X*, where *X* is a compact subset of RN.[37] Then we can formulate a binary classifier as follows:

$$\min \frac{1}{2} ||w||^2 + \frac{1}{\gamma l} \sum_{i=1}^{l} c_i p$$

subject to:

$$(w.\Phi(x_i)) \geq p - c_i \qquad i = 1, 2, \ldots, l \qquad c \geq 0$$

If *w* and *p* solve this problem, then the decision function is given by:

$$f(x) = sign((w.\Phi(x)) - p)$$

### e. Decision Tree

Tree-based classifiers, such as decision tree and random forest, have also been studied with respect to document categorization.[38] In recent years, graphical classifications have been considered as a classification task such as conditional random fields (CRFs).[39] However, these techniques are mostly used for document summarization and automatic keyword

---

[36] Jiang, S., Pang, G., Wu, M., & Kuang, L. (2012). An improved K-nearest-neighbor algorithm for text categorization. *Expert Systems with Applications*, *39*(1), 1503-1509.
[37] Manevitz, L. M., & Yousef, M. (2001). One-class SVMs for document classification. *Journal of machine Learning research*, *2*(Dec), 139-154.
[38] Xu, B., Guo, X., Ye, Y., & Cheng, J. (2012). An improved random forest classifier for text categorization. *J. Comput.*, *7*(12), 2913–2920
[39] Lafferty, J., McCallum, A., & Pereira, F. C. (2001). *Conditional random fields: Probabilistic models for segmenting and labeling sequence data*

extraction.[40] The structure of this technique is a hierarchical decomposition of the data space. The main idea is creating a tree based on the attribute for categorized data points, but the main challenge of a decision tree is which attribute or feature could be in the parent level and which one should be in the child level. To solve this problem, statistical modelling was introduced for feature selection in the tree. For a training set containing *p* positive and *n* negative:

$$H(\frac{p}{n + p}, \frac{n}{n + p}) = - \frac{p}{n + p} + log_2 \frac{p}{n + p}$$

Choosing attribute *A* with *k* distinct value, divides the training set *E* into subsets of $\{E_1, E_2, \ldots, E_k\}$. The entropy (EH) remains after trying attribute *A* (with branches $i = 1, 2, \ldots, k$):

$$EH(A) = \sum_{i=1}^{K} \frac{p_i + n_i}{p + n} H(\frac{p_i}{p + n}, \frac{n_i}{p + n})$$

Information gain (I) or reduction in entropy for this attribute is:

$$A(I) = H(\frac{p}{n + p}, \frac{n}{n + p}) - EH(A)$$

Choose the attribute with largest information gain as parent's node.

### *f.* *Deep Learning*

Lately, deep learning approaches have achieved better results in comparison to previous machine learning algorithms on tasks such as image classification, natural language processing, face recognition, etc. The success of these deep learning algorithms relies on their capacity to model complex and non-linear relationships within data.[41] In this paper we do not explore deep learning techniques as simpler techniques showed good performance.

## E. Evaluation Metrics

The final part of the text classification pipeline is evaluation. Understanding how a model performs is essential to the use and development of text classification methods. There are many methods available for evaluating supervised techniques. Accuracy calculation is the simplest method of evaluation but does not work for unbalanced data sets.[42] We therefore used a common metric that takes imbalance into account. The F1 score is defined as:

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

---

[40] Zhang, C. (2008). Automatic keyword extraction from documents using conditional random fields. *Journal of Computational Information Systems*, *4*(3), 1169–1180.

[41] Hinton, G., LeCun, Y., & Bengio, Y. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

[42] Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *17*(3), 299–310.

However, F1 score assumes a positive and a negative class, and it does not make sense in our case to arbitrarily call a class positive, as it would depend on which one of the parties' perspective one looks from. We therefore used macro-averaged F1 score (MA-F1) as in the literature[43], which is the mean of per-class F1 scores that is calculated by using the alternatives.

$$\text{Macro-averaged F1 Score} = \sum_{i=1}^{n} \frac{F1\ Score\ i}{n}$$

## Methodology

*Document Collection*

The study utilized data scrapped from the Kenya law website (http://www.kenyalaw.org/) which reports on the development of Kenya's jurisprudence via Kenya Law Reports. The training dataset contained 3,087 entries with details of cases from the ELRC. The study assessed both the set of binary variables (case class, case outcome), as well as the multi-label variables (case action, nature of case, county) of this dataset. The following challenges were encountered in retrieving the above data:

i. long download times;
ii. incompleteness of data record;
iii. missing target values,
iv. and duplicate records

These issues resulted in a limited number of training records that could be used in improving the performance of the models. To avoid the problem of overfitting, which is most severe for small training sets, a *k*-fold cross-validation approach was used, in which cross validation was performed *k* different times, each time using a different partitioning of the data into training and validation (testing) sets, and the results were then averaged. The *m* available documents were partitioned into *k* disjoint subsets, each of size *m/k*. The cross-validation procedure was then run *k* times, each time using a different one of these subsets as the validation (testing) set, and combining the other subsets for the training set. Thus, each example was used in the validation set for one of the experiments and in the training set for the other $k - 1$ experiments. The larger the value of *k*, the larger the training set. A 5-fold cross validation approach was utilized.

*Text Classification*

The data collected was categorized into five groups (case class, nature of case, case action, case outcome, county), and each item of metadata was indexed, i.e., labelled as training examples and as 'true class labels' for testing samples. The study utilized five different classification methods: passive aggressive classifier; logistic regression; SVM; decision tree classifier and random forest classifier, and subspace method to the text document classification, respectively.

---

[43] K. Kowsrihawat, P. Vateekul, and P. Boonkwan, "Predicting judicial decisions of criminal cases from Thai Supreme Court using bi-directional GRU with attention mechanism," 2018 5th Asian Conference on Defense Technology (ACDT), pp. 50–55, 2018

**Findings and Analysis**

In assessing the performance of the proposed model, we consider the model's predictions on the test set which represents 30% of the entire labelled dataset. The test set is obtained using stratified sampling to ensure that the final labels within mirror the distribution of the original dataset. Finally, the top model is obtained by comparing the performance of all models on the test set and the highest test score is selected.

The results obtained in the study have implications beyond a simple comparison of algorithms. The findings indicate, not only, which algorithm performs better, but also which set of metadata fields are easier or more difficult to predict. This, in turn, offers insights into the nature of metadata fields and the structure of the court they belong to. In the 5 classification models utilized, the MA-F1 (F1 is the harmonic mean of the precision and recall) observed for predicting the metadata ranged from 82% - 98%. This means, all other factors held constant, given 100 case records, the models would correctly label 82 – 98 of the case records. Further, we note that the lowest performance was observed in the case outcome prediction. This is likely due to the fact that a large chunk of the labels were missing and a best guess of the missing data had to be input manually to aid in testing. Better performance would be expected if the actual labels were used.

Note: In the Classification Performance tables below, bold font numbers denote the average balanced accuracy in a particular classification category.

***Classification Performance***

<table>
<tr>
<td>Case Class:</td>
<td>Class: Criminal | Civil<br><br>Top Model Stats:<br>MA-F1: 85.2%<br>F1 score: Civil – 95% | Criminal – 75%<br><br>Algorithm performance:<br>
<table>
<tr><td>Logistic Regression</td><td>85.2%</td></tr>
<tr><td>Passive Aggressive</td><td>82.1%</td></tr>
<tr><td>Decision Tree</td><td>78.9%</td></tr>
</table>
</td>
</tr>
</table>

*Table 1: The table above shows the performance of the Logic Regression, Passive Aggressive, and Decision Tree classifiers in predicting text from the case class category. The highest performance was by the Logistic Regression model at 85.2%. The classifiers had wide variance in predicting text from civil cases (95%) and criminal cases (75%).*

The performance from case class category had one of the widest variance between target class performance. Specifically, as much as the model was able to correctly identify civil cases (95%) it did not do as well in identifying criminal cases (75%). Other target class results showed minimal variance from the overall model performance. The other wide variance was seen in the county case category which used only 5 counties from the original dataset. The inadequacy in scores from between the counties is a result of the insufficiency of the size of each county sample size, where a prediction is being made with very few samples that have high dimensional features even after dimensionality reduction. Of the 5 classifiers utilized in the study, the decision tree classifier worked significantly better in predicting the outcome of texts from the 5 counties. The case action category resulted in a reduction of the original target variables due to lack of sufficient representation. Using the classes available, the models were able to distinguish between a ruling and an award with high accuracy.

***Classification Performance***

| | |
|---|---|
| *County:* | **Class:** Kisumu \| Mombasa \| Nairobi \| Nakuru \| Nyeri<br><br>**Top Model Stats:**<br>**MA-F1:** 88.4%<br>F1 score: Kisumu - 94.8% \| Mombasa - 82.1% \| Nairobi - 95.1% \| Nakuru - 75.0% \| Nyeri - 95.0%<br><br>**Algorithm performance:** |

| | |
|---|---|
| **Decision Tree** | **88.4%** |
| **Random Forest** | **61.1%** |
| **Logistic Regression** | **53.6%** |

*Table 2: The table above shows the performance of Decision Tree, Random Forest, and Logistic Regression classifiers in predicting text from ELRC documents from Kisumu, Mombasa, Nairobi, Nakuru, and Nyeri counties. The decision tree classifier had the best performance, 88.4%. There was wide variance in the performance of classifiers on documents obtained the 5 counties with the best performance found in the Nairobi ELRC, 95.1%, and the worst in Nakuru ELRC, 75%. This may be attributed to the varying size of data available in each county. Better performance corresponds to larger datasets.*

***Classification Performance***

| | |
|---|---|
| *Case Action:* | **Class:** Ruling \| Award<br><br>**Top Model Stats:**<br>**MA-F1:** 98.2%<br>F1 score: Ruling – 98.6% \| Award – 97.7%<br><br>**Algorithm performance:** |

| | |
|---|---|
| **Passive Aggressive** | **98.2%** |

| | |
|---|---|
| **Logistic Regression** | **97.7%** |
| **SVC** | **97.2%** |

*Table 3: The table above details the performance of the Passive Aggressive classifier model, the Logistic Regression model, and the SVC in determining the ruling in court documents from the ELRC. All 3 models showed a high performance in predicting the ruling and award in these documents. Using the classes available, the models were able to distinguish between a ruling and an award with high accuracy.*

The case outcome category had the lowest overall performance by the classifiers. Despite this, the observed MA-F1 score of 82% is still promising for an initial prediction model. The low performance score may be attributed to insufficient training data; more than 30% of the target variable was missing in the original dataset. That said further tuning could boost the score obtained and open up new research opportunities around case outcome prediction in future.

***Classification Performance***

*Case Outcome:*

**Class:** Allowed | Dismissed

**Top Model Stats:**
**MA-F1:** 81.8%
F1 score: Allowed – 80.2% | Dismissed – 83.5%

**Algorithm performance:**

| | |
|---|---|
| **Passive Aggressive** | **81.8%** |
| **SVC** | **78.9%** |
| **Logistic Regression** | **78.1%** |

*Table 4: The table above details the performance of the Passive Aggressive classifier model, the SVC model, and the Logistic Regression model in predicting the outcome of cases from text for the ELRC records. The case outcome category had the lowest overall performance by the classifiers. The Passive Aggressive model had the best performance. 81.8%.*

The study was unable to use the nature of case as a predictor variable, as initially planned, due to insufficient class representation that would enable proper classification experiments. This can be resolved in the future by using a larger sample dataset. The sum awarded was also a variable of interest in the case metadata. However, there were only 158 records out of 3,087 (5%) with information on case awards. Efforts to create heuristic methods of estimation for award amounts did not bear fruit. Other important factors that affect the results, other than the actual methods used in this study and detailed above, are statistics such as the lengths of case texts or number of samples in the corpora.[44] Studies that have utilized larger corpora have achieved better results in classification of legal documents across similar classes as those assessed in this project; for

[44] Mumcuoğlu, E., Öztürk, C. E., Ozaktas, H. M., & Koç, A. (2021). Natural language processing in law: Prediction of outcomes in the higher courts of Turkey. *Information Processing & Management*, *58*(5), 102684.

example, studies by Long et al. and Şulea et al. which did training on around 130,000 and 100,000 case documents, respectively.[45, 46]

## Discussion and Conclusion

Case metadata prediction is a machine learning and natural language processing application which has not received due consideration and attention in the Kenyan legal system. This paper presents findings from the study of several text classifier models applied to digitized records in Kenya's Employment and Labor Relations court. The aims of the study was to determine the viability of using AI tools to ensure data completion and accuracy in the document digitization process undertaken by the Kenyan judiciary. The research used open source data on cases from the ELRC. The findings from this study showed that text classifier models can be used to help resolve the challenge of missing and inaccurate case meta data. Overall performance accuracy across the class categories for the 5 models (decision tree, passive aggressive, random forest, SVM, and logic regression) utilized ranged from 82% to 95%, indicating that use of these classification techniques is a viable method to complete missing data in digitized records. The study also found that the amount of available digitized records varied across counties with the largest number of records available in the ELRC in Nairobi and the lowest in Nakuru. Additionally, findings from the study indicate that the current approach of documenting award amounts in cases is not viable. Less than 5% of the cases analyzed in this study contained information on case awards. In general, though, the results obtained from the study show that text classification can be automated successfully using machine learning techniques to generate case metadata. Despite the data limitations faced in this study, the good results recorded help increase confidence that advanced NLP techniques have matured enough to be applicable to legal text in the Kenyan Judiciary. In addition, given that our original dataset of 3,087 records had missing values ranging from 4% - 55% (within the target variables) and still achieved good MA-F1 scores (as high as 98.2%), the proposed model shows practical application value in addressing issues of missing data and offers a viable alternative to the manual review and entry of this case meta data. We acknowledge that studies across different courts are needed to conclusively affirm this claim.

Our results have implications beyond comparing algorithms and demonstrating their predictive power. There is a variation in results obtained for different metadata fields, which has interesting potential interpretations. More work is needed to uncover the meaning of this difference, but we hypothesize that it is related to the different content of the cases and different structure of the different types of metadata fields. One possibility is that certain fields have more predictable results because of the nature of the data or bookkeeping that is not substantially related to the content of the proceedings or the structure of the law. This suggests that the success rates of the

---

[45] Long, S., Tu, C., Liu, Z., & Sun, M. (2019, October). Automatic judgment prediction via legal reading comprehension. In *China National Conference on Chinese Computational Linguistics* (pp. 558-572). Springer, Cham.

[46] Sulea, O. M., Zampieri, M., Malmasi, S., Vela, M., Dinu, L. P., & Van Genabith, J. (2017). Exploring the use of text classification in the legal domain. *arXiv preprint arXiv:1710.09306*.

algorithms are not merely dependent on text content, but the context of this content is also a determining factor - the nature of the cases and the structure of the legal system is important.

This work should contribute to setting general baselines in this field of research on the use of AI tools to aid the digitization of records in the Kenyan judiciary. There are several technical issues that may be taken up in future work, such as more detailed feature extraction and sentence-level supervision for systems that are not end-to-end. Further experiments and research on the retrieval of case metadata can also be addressed. It is our belief that the scope and systematic nature of this study provides a framework that can be applied to the study of other legal systems, especially those in the Global South, where the process of digitizing court records is in its infancy. The method outlined in this paper may be applied where (i) the legal system corpus is systematically separated to sub-corpora according to the different types and levels of courts within the hierarchy of that particular legal system, (ii) a reproducible method of pre-processing data that makes it suitable for further higher-level processing is provided, and (iii) experiments to characterize the performances of baseline using classical machine learning approaches like SVMs and random forests and several contemporary deep learning based methods with and without attention mechanisms can be performed.