Transactions on Engineering and Computing Sciences - Vol. 11, No. 6
Publication Date: December 25, 2023
DOI:10.14738/tecs.116.15858.
Ayankoya, F., Ngige, O., Onuiri, E., Adigun, T., & Omotunde, E. (2023). Multi-Layer Perceptron Algorithm, An Effective Tool for the Prediction of the Judgments of the Supreme Court of Nigeria. Transactions on Engineering and Computing Sciences, 11(6). 01-18.
SERVICES FOR SCIENCE AND EDUCATION

Multi-Layer Perceptron Algorithm, An Effective Tool for the Prediction of the Judgments of the Supreme Court of Nigeria

Ayankoya, F.

Ngige O.

Onuiri, E.

Adigun, T.

Omotunde, E.

ABSTRACT

This study attempted to quicken the pace of justice delivery by developing a predictive model for the classification of Supreme Court judgments in Nigeria (SCN), and improved the performance of the computational process that is required for the identification of the pattern between feature and judgment using the Pearson Correlation Coefficient to select relevant features. The study was carried out using 5585 records of precedent judgments delivered at the SCN between 1962- July 2022. Data was collected from an independently owned data repository (Primsol Law Pavilion). Data annotation and feature extraction were carried out and Pearson Correlation feature selection method was used to select the most relevant features from the initially identified features. Multi-layer perceptron with ADAM optimization function was used to develop the classification model. The result of the feature selection revealed the most relevant features. The MLP-ADAM model was evaluated and benchmarked with a related study carried out to predict judicial decisions of criminal cases from Thai Supreme Court, using both conventional and modified models. The result of MLP-ADAM showed a better performance of predicting judicial decision, showing 100% precision, 99% recall and 100% F1 Score for the as against 69.59% precision,79.87% recall and 74.38% F1score obtained by the Bi-GRU + attention model of the existing study. The study also revealed that the ADAM optimization function was significant in achieving good accuracy and generalization ability of the model. The use of structured and well-organized dataset enabled the model to train effectively.

Keywords: Prediction, Feature selection, multi-layer perceptron, ADAM optimization function, Classification, Pearson correlation coefficient.

INTRODUCTION

The role of the courts in the administration of justice is sacrosanct to the stability of the human society. The judiciary is regarded as the last hope for the oppressed considering the lofty responsibility of justice dispensation assigned to them by the society. One of the cardinal objectives of the judiciary as the third arm of the government is to defend and uphold the

constitution, and assuring that the rule of law prevails at all times [1]. The role an effective dispensation of justice can play in creating an enabling environment for the development of the country can never be underestimated. The effectiveness of the court lies not only in their ability to be fair, firm and ultimately neutral, but also for justice to be dispensed timely and fairly. Despite these crucial roles of the judiciary, the judicial process in Nigerian is often slow, and leaves much to be desired. This has made people loose confidence in using the courts as it is mostly believed that delayed justice is justice denied.

The non-transparency and lack of knowledge of the judicial proceedings has led to great disturbances and uncertainty amongst the communities confined to a certain legislature. The need for established practices and methodology is becoming more urgent with the growing availability of data [4]. In striving for transparency, many national and international courts in Europe adhere to the directive to promote accessibility and reuse of public sector information and publish their documents online [2]. Digital access to a large amount of published case law provides a unique opportunity to process this data automatically on a large scale [4].

Recent advances in data collection, systems of aggregation, algorithms, and processing power have transformed computers from machines that perform rigidly defined tasks to machines that can learn without supervision and adapt to new inputs without the need to be reprogrammed [5]. Prediction of legal judgment is a long-lasting topic in the theory and practice of law to improve judicial consistency, access to justice, and administrative efficiency [5].

Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome [6]. Intelligent prediction of judgments classifies cases with a great precision, finds their similarity, learns from them, and builds their judicial experience [7]. The purpose is to enable machines to predict the decisions of legal cases after reading the description of facts. It can not only help legal personnel to deal with cases and improve work efficiency but also help litigants to understand cases and make them have certain psychological expectations about the outcome of cases [8]. Applying predictive analytics in the justice sector has the potential of advancing equity and fairness in law practice. Some empirical works observed contradictions in judicial behavior [9]. Justices occasionally have been found to rule or pass judgment, relying on issues outside the case record. Some of these justices usually would rule in favor of litigants who they are more emotionally attached to, or to some others because they have common demographic characteristics [10]. Machine learning techniques depending on the judicial attributes or the case features can offer the much-needed reprieve by their ability to detect when the extra-legal biases are allowed to influence the decision making of judges.

The prediction of the decision of legal cases by means of machine learning algorithms has become a hot topic [11]. Legal decision systems have been established based on Artificial Intelligence (AI), by predicting verdicts automatically to support lawyers [12]. Proponents argue that big data analytics promises to decrease human prejudice in court judgments and to provide the legal system a rational, scientific approach. [13] [14]. The application of machine learning has been applied to providing solutions called algorithmic decision predictors for supporting judges in individual cases [15]. These algorithms predict the final decision of a legal

case, a decision that would otherwise be made by the judge(s) such as guilty/not guilty [16], 'violation' or 'non-violation' [2][16][17]. Algorithmic decision predictors are sometimes claimed to increase the uniformity and predictability of judicial decision-making, which is required by the equality principle. According to these claims, judges can use decision predictors in order to come to a more consistent, more informed and less biased judgments [18], [19].

Multilayer Perceptron (MLP) is a kind of artificial neural network, consisting of three distinct layers: input and output layers, and one or more hidden layers with many neurons stacked together [20]. ADAM is an algorithm that uses adaptive estimations of lower-order moments to optimize stochastic objective functions using first-order gradients.

Although these machine learning models have attained a modest level of prediction accuracy, some researches have revealed significantly lower levels of ML-based judicial outcome prediction accuracy [21]. Deep Learning (DL) algorithms have the potential to create prediction models that have higher levels of accuracy. Models built using such DL algorithms have the capability of learning the complexities of large text datasets which other ML models may not be able to do.

Again, questions have been raised about the integrity of the manual process of identification of features in a dataset [11]. This study therefore used a multi-layer perceptron algorithm to predict the decisions of the Supreme Court of Nigeria (SCN). The predictive model was built with case records collected from historical (precedent) court cases decided at the Supreme Court of Nigeria in other to support the justices in making well-informed decisions about related cases in the future.

Problem Statement

A number of related works surrounding predictive justice system depended on the use of manually identified and extracted features (both relevant and irrelevant ones) during data annotation in building the classifiers [22] [23] [24] [25] [6] [7] [26][27] [28] [29] [30] [22], while several other studies were based on the use of unstructured data set [23] [31] [32] [22]. The presence of these irrelevant and redundant features reduces the model's generalization capability, slows down the training process and may also reduce the overall accuracy of a classifier [33]. Furthermore, adding more redundant variables to a model increases the overall complexity of the model [34].

To overcome this shortcoming of previous studies, this study aimed to improve the performance of the computational process that is required for the identification of the pattern between feature and judgment by a machine learning algorithm by developing an efficient filter-based feature selection approach specifically the Pearson Correlation Coefficient to select relevant features from the initially identified and extracted features, and, subsequently, develop a predictive model using MLP classifier, improved by the ADAM search algorithm to predict Supreme Court decisions.

LITERATURE REVIEW

Predictive modeling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data [35]. It is a form of data mining that analyzes historical data with the goal of identifying trends or patterns and then using those insights to predict future outcomes.

Machine Learning (ML), a branch of artificial intelligence, enables computers to learn from prior data by employing statistical and optimization techniques. [36]. Predictive modeling is the most significant application of machine learning out of all the others. The same collection of features (attributes/independent variables) is employed to represent every instance (record/set of fields or attributes) in each dataset utilized by machine learning techniques. The features could be binary, categorical, or continuous.

Feature Selection

The goal of this procedure is to isolate and get rid of as many unnecessary and redundant features as you can [37]. The goal is to reduce the dimensions of the data to enable faster and more efficient data mining methods. Typically, features are divided into

1. **Relevant** characteristics are those that affect the target class (output). The others cannot take on their job.
2. **Irrelevant features** are those with no relationship to the target class. Their values wouldn't have any bearing on the result if they were generated at random.
3. **Redundant**: are features that can take the role of another (perhaps the simplest way to incur model redundancy).

Related Works

[38] worked on the prediction of the outcome of court cases in Brazil. The study developed a Web scraper for collecting data from Brazilian courts. The scraper first searched for the URL that contains the list of court cases from the Brazillian State higher court (appellate court). Then, the scraper extracted from these HTML files the specific case URLs and downloaded their data. Next, it extracted the metadata and the contents of legal cases and stored them in a CSV file format. The study performed a text segmentation task to identify target sentences within the legal decisions, containing information about the case description, the decision, and the decision's unanimous or non-unanimous nature. Next, the study extracted the features by assigning labels and labeling the target sentences accordingly. The study adopted the use of Guassian naïve Bayes, support vector machines, random forest, decision trees, and extreme gradient boosting for the development of the predictive model. The performance of the predictive models were evaluated based on the F1-score, precision, accuracy, and recall. The results revealed that the random forest had the best overall performance with an F1-score of 73.4%.

[2] worked on the application of machine learning for the prediction of decisions of the European courts of human rights (ECtHR). Data for the study was collected from a publicly avaialble data that was provided by the ECtHR Each judicial decisions consists of an introduction, procedure, facts containing the circumstances and relevant law, law, judgement and dissenting opinions. The reserchers sought to build a predictive model for the determinig

the violoation of fundamental human rights. The elements of the dataset that were used in the study consisted of the procedures and facts protion of each documents. Natural langauge processing techniques were used to analyze the textual components of the data based on n-grams, removal of capitalization and stop words. The SVM was used for formulation of the predictive model for violation of human rights based on the generated dataset using a 10-fold cross validation scheme. The results of the study revealed that using relatively simple and automatically obtainable information, the models were able to predict decisions correctly in about 75% of the cases. The study focused on the prediction of the violation of human rights.

[22] worked on the prediction of legal cases using machine learning algorithms. The data used in the study consisted of records collected from 86 criminal cases from the Delhi District Court in India between 2017 and 2018. The data used in the study was a structured data consistsing of 19 features including the target label which is the outcome of the case defined as a binary class as convicted or acquitted. The features were extracted from the court cases via a manual approach of reading and analyzing the judgements. A number of supervised machine learning algorithms were used to formulate the predictive model for the outcome of the criminal cases, namely: support vector machines, bagging, boosting, random forest, classification and regression trees. k-nearest neighbours, naïve bayes and linear regression methods. The results revealed that the models showed varying perfomance such that CART had the highest accuracy and F1-score, SVM had the highest recall and bagging had the highest precision. The study did not consider the identification of the most relevant legal factors among the inntially identified legal factors alongside its impact on the performance of the machine learning algorithms.

[4] worked on the prediction of the outcome of judicial cases and analysis using machine learning algorithms. The study collected data from an online repository which consists of various cases brough before the Delhi District Court in India. The features that are required for the prediction of the outcome of the court cases were extrcated from the documents. A number of supervised machine learning algorithms such as logistic regression, K-nearest neighbours, naïve Bayes, random forest and decision trees algorithms were applied. The results of the comaparative analysis revealed that the best performance was assesesed based on the precision of both classes which was achieved by the naïve Bayes classifier. The study only focused on the use of precision as a basis of evaluating the performance of the predictive model developed in the study. [8] Used convolutional neural network to extract text features from legal judgments, and applied Principal Component Analysis (PCA) algorithm to reduce the dimensionality of the data. The study built a prediction model based on process supervision was introduced to model the dependency relationship between sequential sub-datasets, and the genetic algorithm was used to optimize the parameters in other to improve the performance at the final prediction. The research achieved the best results on the four different data set used for the study (CAIL2018_Small, CAIL2018_Large, CAIL2019_Small, and CAIL2019_Large), as compared to that achieved by the benchmark.

[25] Carried out a comparison of the predictive performance of five machine learning models including kNN, logistic regression, bagging, random forests and SVM for judicial cases. Adopting the benchmark dataset of the European Court of Human Rights [11], the study utilized the textual converted data to train and test the five ML classifiers in other to analyze the relevant

factors that affect the predictive ability of the ML algorithms. The result showed that the SVM out-performs most of the other ML algorithms in most of the articles of the Convention due to the ability of SVM to work with small datasets, which were 250, 80 and 254 for the three articles. Furthermore, the study demonstrated that the SVM is the only one with reasonable feature selection capability among the five models. The dataset was very small.

[31] Sought to build a legal judgment prediction for UK courts using machine learning algorithms. The study tried to overcome the limited ability of Natural Language Processing (NLP) techniques to recognize complex semantic structures such as arguments, and utilized alternative text mining techniques such as word embedding features with neural network models. Using an unstructured data containing 4,959 text files, the study built a structured labeled data set for UK court documents, and applied logistic regression, SVM, random forest, and kNN for classification.in other to test alternative text mining techniques with neural network models, the study adopted the use of single layer perceptron and multi-layer perceptron. The result showed that the kNN and RF algorithms were the most consistent across the feature sets, while the choice of SLP and MLP did not provide a concise improvement over the standard machine learning models, though the MLP outperformed the SLP. Domain specialists were not involved in the study as to assist in identifying the relevant feature sets.

Working on a very simple and limited accident cases dataset, [7] attempted to use the 514 judgments delivered at the Errchidia court in Morocco between 2017 to 2019 to predict the outcome of accident cases in Morroccan courts. The study identified seven features which were manually extracted and used to train the different machine learning models including linear regression, decision trees, and random forests classifiers. The highest accuracy of 91.05% was achieved by the random forest, followed by the decision tree with 82.3% and the linear regression with 63.56%. The dataset used for the study was very small, and there may have been over fitting issues.

[39] Presented methods for the automatic classification of Polish court rulings into various departments using deep neural network and ADAM as an optimization algorithm. The study used natural language processing methods and classifiers based on convolutional and recurrent neural networks on authentic, anonymized Polish court rulings from 1998 to 2020, and analyzed various general language embedding and multiple neural networks. The results showed that such models can classify documents with very high accuracy (>99%). The study considered only accuracy as a means of performance evaluation. Feature extraction was manual and based solely on domain knowledge.

Research Gaps

Most studies were carried out using unstructured data set, which primarily were the raw text judgments. Such studies limited the work to the application of only natural language processing techniques. The snags in such systems are not only the huge financial costs, but also ambiguity and context. It can be challenging for computers to decipher the intended meaning of words and sentences because human language is frequently imprecise and context-dependent. However, this study will be making use of structured data which contains an attribute that will

be analyzed using feature selection techniques to generate a subset of features. This is computationally cheaper than adopting an NLP technique.

Most previous studies were carried out using the traditional machine learning algorithms, which although have attained a modest level of prediction accuracy, some researches have recorded significantly lower levels of ML-based judicial outcome prediction accuracy. This present study was carried out using a deep learning MLP, with ADAM optimization function which greatly improved the generalization ability of the developed model.

The dataset used in some of the previous studies was very few, mostly below 500 records. Such studies carried out with few datasets may have over-fitting issues, outliers and sampling bias. This study was however carried out using 5585 precedent case records.

Most of the previous studies were carried out with features that were manually extracted from the dataset. The consequence of this is model bias and complexity. This study was done with relevant features that were selected using Pearson Correlation technique. This increases the model's generalization capability, quickens the training process and increase the overall accuracy of the classifier.

Aim and Objective

The aim of this research is to develop a predictive model using MLP algorithm with ADAM optimization function for the classification of Supreme Court judgments in Nigeria. The research objectives are to:

1. identify the most important features that influence the classification of Supreme Court judgments in Nigeria using feature selection.
2. develop a classification model for predicting the outcomes of judicial cases based on features selected in (i) using ADAM-Multi-layer perceptron algorithm
3. evaluate the performance of the predictive models developed in (ii) using Thai Supreme Court Cases (TSCC) dataset

METHODS

To achieve objective one, secondary Data from precedent appeal cases resolved by the Nigerian Supreme Court were gathered from a web source (Primsol Law Pavillion). From the unstructured data, variables relating to court case proceedings were found in associated literature and used as the foundation for creating an electronic structured version of the dataset stored as a spreadsheet file. Filter-based Feature selection algorithm (Pearson Correlation Coefficient) was used to remove redundant and irrelevant features from the dataset generated, thus revealing information about the most relevant features required for predicting case outcomes. At this point, dataset containing information about the relevant features that were extracted from the initially identified dataset was generated. For the objective two, the predictive model required for the classification of the outcome of court cases was formulated by adopting the features selected as the input variables while the outcome of the case was adopted as the target variable using a multi-layer perceptron algorithm. The holdout approach was adopted for model simulation by using 80% of the dataset to build the classification model (the training dataset) and the remaining 20% for validating the classification model (the testing

dataset). The development of the classification model was implemented using Python within the Google CoLaboratory®, an online Python jupyter notebook.

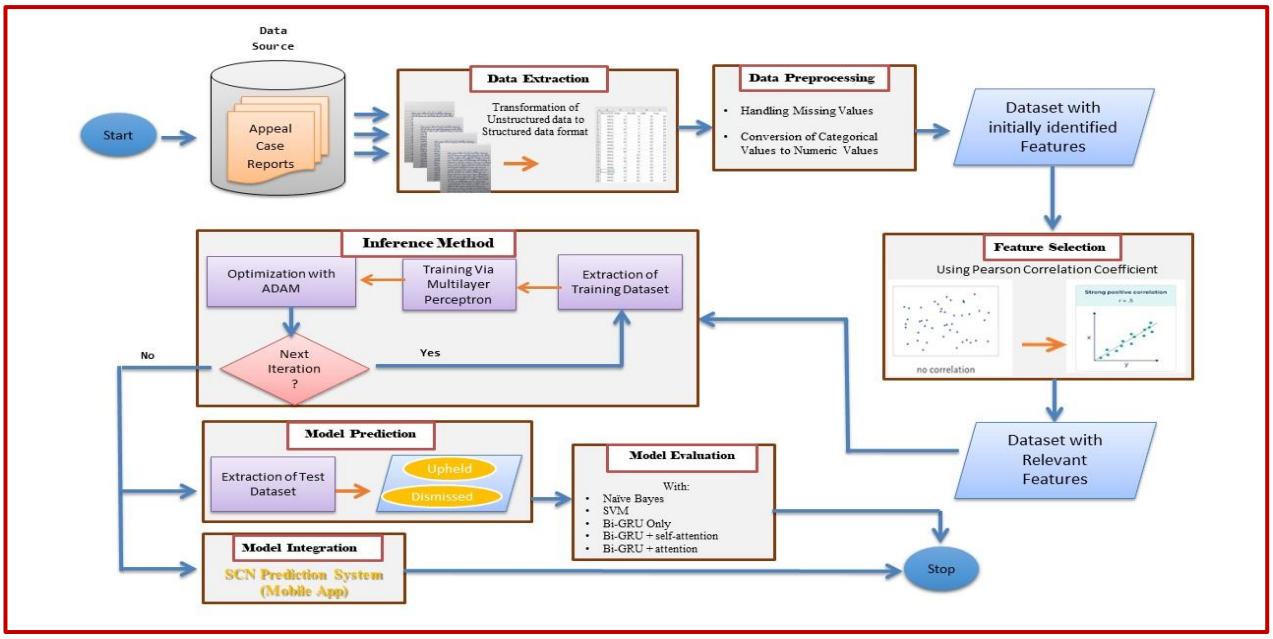


Figure 1: Conceptual Framework of Research Process

For objectives three, the results of the correct and incorrect classifications from the simulations conducted were used as a basis for the evaluation of the performance of the predictive models. The results of the correct and incorrect classifications were gotten from a confusion matrix. Performance metrics, including accuracy, recall, precision, and f1-score were estimated. The results of the performance evaluation of the predictive model based on the test dataset were compared with those of the existing benchmark studies.

Framework of the Inference Model

Training of the gene regulatory network (GRN) inference model is via an Artificial Feed Forward Neural Network called Multilayer perceptron. The architecture of multilayer perceptron is chosen because it has capability of capturing complex, non-linear and dynamic relationships among factors that contributes to the outcome of Supreme Court judgment. Handling of non-linear and dynamic relationships among the variables has initially been aided by the effective feature selection algorithm selected. Also, there is need for a learning algorithm that reliably finds a nearly globally optimal set of weights in a relatively short time. ADAM optimization algorithm is employed as the learning algorithm of this Supreme Court judgment classification modeling. This study used ADAM optimization algorithm to train MLP in achieving fast convergence rate and to avoid local minima problem. The process is to show that the computational efficiency of MLP training process is highly enhanced when coupled with the proposed updated method. The main goals of using Pearson correlation for feature selection and ADAM to calculate the weights and adjustments to the network in order to minimize the output error are to decrease the computational cost and to accelerate the learning process using a hybridization method.

Multilayer Perceptron Preliminaries

MLP is a finite and directed acyclic graph that extends the perceptron with hidden layers of processing neurons. It directly incorporates the capabilities of higher order networks but using a larger number of weights and processing elements. MLP results into a map that is very flexible and powerful, though hard to analyze, vulnerable to incremental addition of units, and exhibits slower learning. However, this model is systematically designed with a powerful feature selection algorithm and an efficient training algorithm to eliminate these limitations in building the classification network.

Design of ADAM-MLP Classifier

The architecture of the NN, the Multilayer Perceptron model consists of three layers; the input layer, hidden layer and the output layer. The input layer has 8 neurons (representing the 8 relevant features selected by the Pearson Correlation method), the number of the hidden layer neurons was set to 15 realized through trial-and-error, and the output layer has 2 neurons (representing class distribution of the Supreme Court outcome; Upheld or Dismissed). RELu activation functions were applied for the hidden and output layers respectively. The ADAM optimization algorithm was applied for training the network weights and bias. In the MLP-ADAM algorithm, the ADAM configuration parameters are set and the weight matrix and the corresponding biases for optimizing MLP are initialized. Also, the aggregate of gradients at time t (M_0) and sum of square of past gradients (V_0) are initialized to zero. In the first epoch, the initial weights and biases are passed to the MLP. The weights in MLP are calculated and compared stochastic gradient accent with ADAM. In the next cycles, ADAM will update the weights with the best possible solution and will continue searching the best weights until the last cycle/ epoch of the network is reached or either the mean square erroe is achieved. The data were partition into 80%-20% for training and testing the performance of the classifier respectively.

Training The MLP Classifier-ADAM Optimization Algorithm

Adam is a method for efficient stochastic optimization that only requires first-order gradients with little memory requirement. It derived its name from" adaptive moments". Adam combines the best properties of the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems. It is an alternative optimization algorithm that provides more efficient neural network weights by running repeated cycles of "adaptive moment estimation."

The following are the Adam Configuration Parameters;

1. alpha. Also referred to as the learning rate or step size. The proportion that weights are updated (e.g., 0.001). Larger values (e.g., 0.3) results in faster initial learning before the rate is updated. Smaller values (e.g., 1.0E-5) slow learning right down during training
2. beta1. The exponential decay rate for the first moment estimates (e.g., 0.9).
3. beta2. The exponential decay rate for the second-moment estimates (e.g., 0.999). This value should be set close to 1.0 on problems with a sparse gradient (e.g., NLP and computer vision problems).
4. epsilon. Is a very small number to prevent any division by zero in the implementation (e.g., 10E-8).

Network Performance Metrics:
The results of the simulation of the predictive model using the MLP-ADAM algorithm were interpreted using a confusion matrix which contained the correct and incorrect predictions of the model validation using the test dataset. The number of correct and incorrect prediction of the target class were used as a basis for the evaluation of the performance of the predictive model using a number of performance evaluation metrics including accuracy, recall, precision and f1-measure, such that the positive cases are the considered approved appeals and negative cases are considered as the dismissed appeals.

MLP-Based Classification Results
The use of MLP to model gene-regulatory networks (GRN) is based on the assumption that the determinant factors on the outcome of a particular Supreme Court decision is represented in form of a neural network where node represents the factors and the edge defines the decision interaction. MLP model training is a non-linear optimization problem where the model optimal parameters are to be determined by minimizing the mean square error. This is done by fitting the calculated Supreme Court binary decision with the observed Supreme Court binary decision, with the use of an objective function, fitness function or loss function to measure the quality of the solution. The objective function is given as equation 4.1:

$$fob = \sum^{N}_{i=1}(X_c - X_o)^2 \qquad \text{Equation 4.1}$$

where *N* is the number of sampling instances in the dataset representing the number of datapoints, *Xc* is the calculated binary decision value at time *t* during model training, and *Xo* is the observed binary decision value at time *t*. The feature selection algorithm incorporated at this stage is used to select the best combination of affecting case facts of the Supreme Court decision. In this work, 8 hypothetical highly ranked case facts are chosen to act as the input nodes of the MLP model and the target binary decision acts as the output node. This idea brought about great reduction in the search space, increases the accuracy of the prediction, and reduces the complexity of the system.

MLP Parameters and Initial Values
The MLP model hyper-parameters of this work include the set of parameters for its formalism, model architectural settings and the activation function. The trainable parameters for the MLP formalism are the weights w_{ij} and the bias *bi,* while the non-trainable parameters are activation function, optimization function, number and units of hidden layers, and the learning rate *r*. The process of training involves finding the optimal values of these parameters, and their initial values are set as follows:

- w_{ij} = rand [0,1]
- b_i = rand [0,1]
- r = 0.1

The settings for the architecture are the number of layers and the units of each layer. The model contains one hidden layer with the input and the output layer described below:

- Input unit, n = 9 (determined by the relevant features)
- Number of hidden layers, h = 1

- Number of units in the hidden layer = 15
- Output unit, y = 1

Relu function is used as the activation function at the different layers. The training epoch of the network is determined by a stopping criterion by the model, that is, "Training loss did not improve more than tol=0.000100 for 10 consecutive epochs", which indicates that the observed value is almost equal the calculated value.

DISCUSSION OF TSC CLASSIFICATION RESULTS

The training and prediction of the inference method was carried out using MLP classifier and ADAM optimization function. The model was implemented in Python via deep learning library Sklearn. The dataset was divided into training and test sets using 80:20.

Model Performance on Test Data Set

The model's performance is measured via F1, in both Macro- and Micro-average. F1 came from the harmonic mean of Precision and Recall. The model prediction is firstly calculated using the confusion matrix showing correctly predicted guilty cases and not guilty cases, and incorrectly predicted guilty cases and not guilty cases. Figure 2 shows the confusion matrix of the model on test data, and Figure 3 shows the performance metrics values of the model on test data.

```
In [18]: #Confusion matrix
         cm_test = confusion_matrix(target_test, pred_test)
         print("Confusion matrix on Test Dataset: " + str(cm_test))

         Confusion matrix on Test Dataset: [[120   0]
          [  1 121]]
```

Figure 2: Confusion Matrix on Test Data Set

```
In [20]: output = classification_report(target_test, pred_test)
         print(output)

                       precision    recall  f1-score   support

                    0       0.99      1.00      1.00       120
                    1       1.00      0.99      1.00       122

             accuracy                           1.00       242
            macro avg       1.00      1.00      1.00       242
         weighted avg       1.00      1.00      1.00       242
```

Figure 3: Performance Metric Values on Test Data Set

The confusion matrix shows that the developed model performed well in predicting the Thai Supreme Court judgment decisions. It presents the following:

- 20% of 1,207 cases = 242 test cases
- Number of correctly predicted guilty cases = 121
- Number of incorrectly predicted guilty cases =1
- Number of correctly predicted not guilty cases = 120
- Number of incorrectly predicted not guilty cases = 0

Also, the results represent an accurate prediction model with Accuracy = 1.0, average Recall = 1.0, average Precision = 1.0, and average F1 = 1.0.

Model Evaluation with Previous Studies

The MLP-ADAM model for predicting the outcome of Supreme Court judgment is compared with a related study [24]. The study was carried out to predict judicial decisions of criminal cases from Thai Supreme Court, it was experimented using different models both conventional and modified models (Naïve Bayes, SVM, Bi-directional GRU, and modified Bi-directional GRUs). This study used the same dataset used on all these models but in an already extracted format. The prediction results of these benchmarked models and that of MLP-ADAM are presented in Table i and Table ii in a comparative format using the performance metrics.

Table i: Performance Comparison of MLP-ADAM with benchmarked models

Models	Guilty			Not Guilty			%Macro F1	%Micro F1 (Acc.)
	%P	%R	%F1	%P	%R	%F1		
Naïve Bayes	71.15	49.66	58.50	47.18	69.07	56.07	57.28	57.32
SVM	67.63	78.52	72.67	56.16	42.27	48.24	60.45	64.23
Bi-GRU only	68.80	57.72	62.77	47.93	59.79	53.21	57.99	58.54
Bi-GRU + self-attention	69.53	59.73	64.26	49.15	59.79	53.95	59.11	59.76
Bi-GRU + attention	69.59	79.87	74.38	60.00	46.39	52.33	63.35	66.67
MLP-ADAM	100.00	99.00	100.0	99.00	100.0	100.0	100.00	100.00

Table ii: Performance Statistics of Predicted Output by Each Model

Models	# Guilty (isguilty=1)	# Not Guilty (isguilty=0)
Naïve Bayes	104	142
SVM	173	73
Bi-GRU Only	125	121
Bi-GRU + self-attention	128	118
Bi-GRU + attention	171	75
MLP-ADAM	121	120

The result of MLP-ADAM shows a better performance of predicting judicial decision. Besides, this method shows great accuracy above both traditional Machine Learning techniques and modified techniques meaning that the model has better advantage in being accurate and a less complex method. Figures 4. 5 and 6 describe the performance comparison of the proposed model with the benchmarked model.

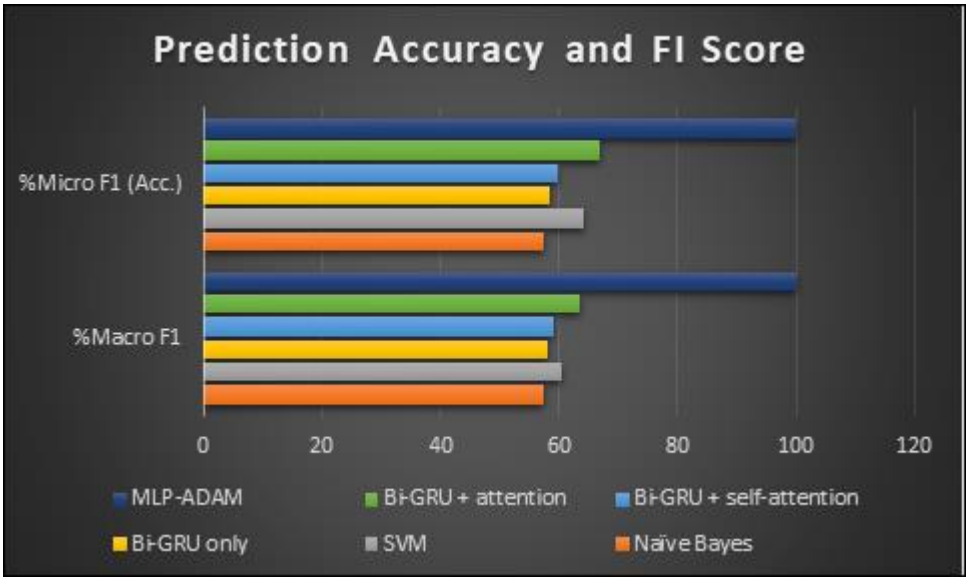


Figure 4: Model Comparison based on Prediction Accuracy and F1 Scores

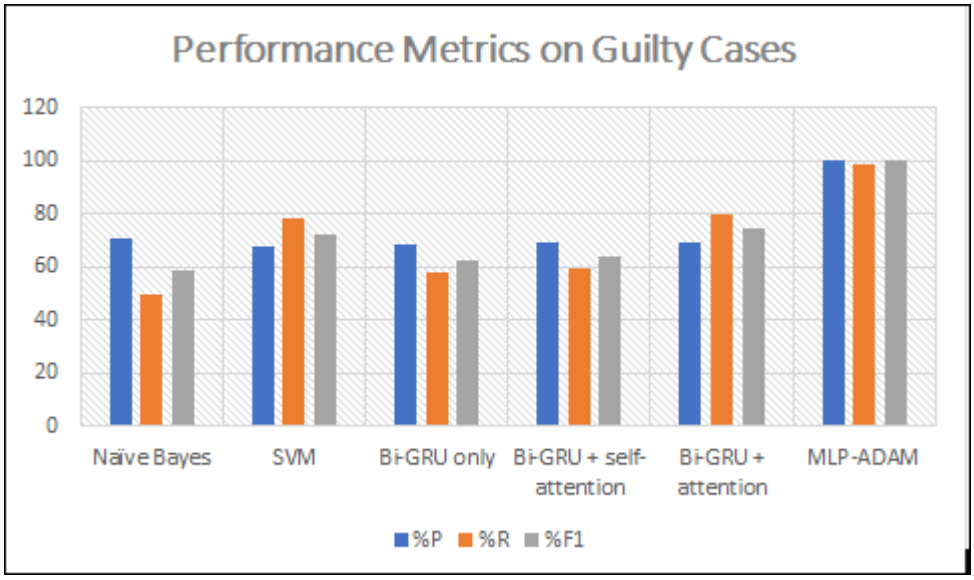


Figure 5: Model Comparison based of Precision, Recall and F1 in Predicting Guilty Cases

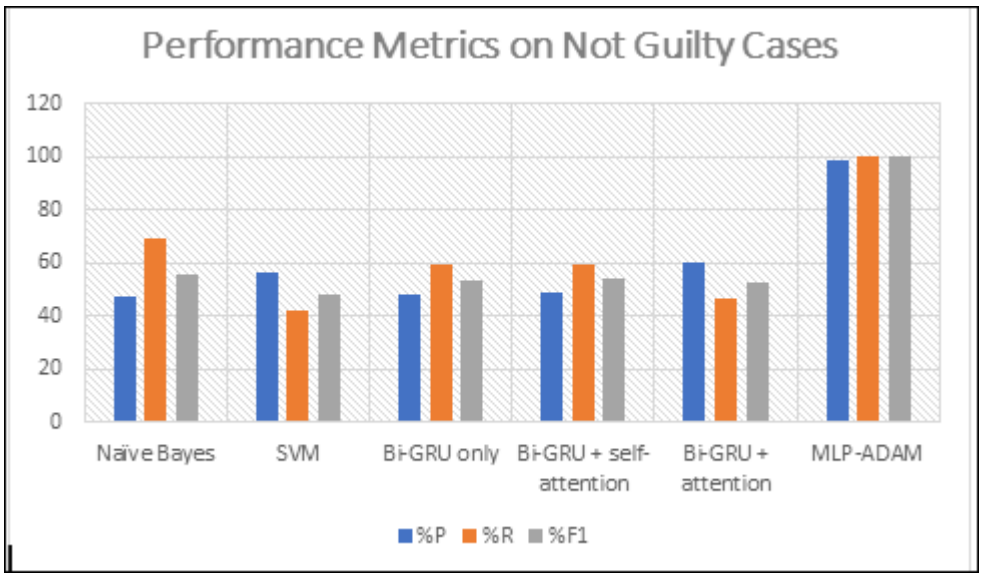


Figure 6: Model Comparison based of Precision, Recall and F1 in Predicting Not Guilty Cases

DISCUSSION OF SCN CLASSIFICATION RESULTS

The Supreme Court of Nigeria dataset was subjected to the built, trained and tested MLP-ADAM model in order to predict the classification of Supreme Court judgment in Nigeria. The model prediction efficiency is calculated using the confusion matrix showing correctly predicted dismissed cases and upheld cases, and incorrectly predicted dismissed cases and upheld cases. With this local data, the ratio of the train to test data sets is set at 70:30 due to large dataset. The confusion matrix shows that the developed model performed well in predicting the Supreme Court decisions. It presents the following:

- 30% of 4,698 cases = 1,409 test cases
- Number of correctly predicted dismissed cases = 898
- Number of incorrectly predicted dismissed cases = 23
- Number of correctly predicted upheld cases = 479
- Number of incorrectly predicted not upheld cases = 9

The result above shows a better prediction but not as accurate as when the model is used on TSCC dataset. This is believed to be because of the largeness of the SCN dataset and probably the high level of missing values and imbalance in the dataset (i.e more of dismissed cases than upheld cases).

The accuracy of the model using the local dataset is shown in Figure 7. The accuracy was also low because of the reasons stated.

```
In [70]: # Make prediction on train and test dataset
         pred_train = classifier.predict(dependent_train)
         score_train = np.round(accuracy_score(pred_train, target_train),2)
         print("Accuracy on Train Dataset: "+  str(score_train))
         pred_test = classifier.predict(dependent_test)
         score_test = np.round(accuracy_score(pred_test, target_test),2)
         print("Accuracy on Test Dataset: "+  str(score_test))

         Accuracy on Train Dataset: 0.67
         Accuracy on Test Dataset: 0.64
```

Figure 7: Accuracy Score of MLP-ADAM Model on SCN Train and Test Datasets

SUMMARY

This study was focused on the application of neural network algorithms to the development of predictive model required for the classification of the outcome of appeals presented to the SCN. The study adopted the use of data containing information about the various appeals that had been presented at the SCN for about 60 years which were stored in electronic text files. The features that were associated with the determination of the outcome of the decision of appeals were identified from related works following which information about the features were taken from the dataset. The dataset containing the initially extracted features was subjected to feature selection based on the Pearson's correlation metrics. Datasets containing information about the features selected by the metrics were generated. The MLP-ADAM algorithms were adopted for the development of the predictive models using the datasets that were generated. The experiment was conducted using 80% of the dataset for training and 20% of the dataset for testing. The model was later trained using the dataset from the existing benchmark studies, and the performance evaluation metrics were taken and compared using metrics, namely: accuracy, precision, recall and f1-score.

CONCLUSION

The study concluded that the outcome of appeal cases that are brought before the SCN can be predicted in advance via the development of a predictive model which is based on information about associated features. It was shown that feature selection can influence the performance of predictive models however the metrics provided may provide differing results. The study revealed that using correlation-based methods for feature selection can provide better. It was determined that using feature selection can provide a smaller yet relevant set of features from the original feature space which are bound to provide performance better than the initially identified feature space. The study demonstrated that the evaluation of the performance of predictive models using machine learning algorithms can be justified through the use of various metrics rather than limiting to one metric.

References

[1] H. Ahmad, M. U Asghar, A. Khan, A.H Mosavi, "A Hybrid Deep Learning Technique for Personality Trait Classification from Text". IEEE Access, 9, 146214–146232, 2021.

[2] M. Medvedeva, M. Wieling, and M. Vols, "Rethinking the Field of Automatic Prediction of Court Decisions". Artificial Intelligence and Law, 1-19. Retrieved from https://doi.org/10.1007/s10506-021-09306-3.2022

[3] M. Marković, and S. Gostojić "Open Judicial Data: A Comprehensive Analysis" Social Science Computer Review, vol 38, pp. 295-314. 2018

[4] P. Bhilare, N. Parab, N. Soni, B. Thakur, B, "Predicting outcome of judicial cases and analysis using machine learning". International Research Journal in Engineering Technology, 6, 326-330, 2019

[5] K. Ashley, and S. Brüninghaus, "Automatically classifying case texts and predicting outcomes". Artificial Intelligence in Law, 17(2), 125-165, 2009

[6] B. Waltl, G. Bonczek, E. Scepankova, J. Landthaler, & F. Matthes, Predicting the outcome of appeal decisions in germany's tax law. In Electronic Participation: 9th IFIP WG 8.5 International Conference, ePart 2017, St. Petersburg, Russia, September 4-7, 2017, Proceedings 9 (pp. 89-99). Springer International Publishing.2017

[7] A. Haidar, T. Ahajjam, I. Zeroua & Y. Farhaoui. Application of machine learning algorithms for predicting outcomes of accident cases in Moroccan courts. Indonesian Journal of Electrical Engineering and Computer Science, 26(2), 1103-1108, 2022.

[8] X. Shang. A computational intelligence model for legal prediction and decision support. Computational Intelligence and Neuroscience, 2022.

[9] D. L Chen, Machine learning and the rule of law. Revista Forumul Judecatorilor, 19, 2019

[10] J. J Rachlinski, & A.J Wistrich. Judging the judiciary by the numbers: Empirical research on judges. Annu. Rev. Law Soc. Sci, 13, 17-32, 2017

[11] N. Aletras, D. Tsarapatsanis, D. Preotuic-Pietro, and V. Lampos, "Predicting Judicial decisions of the European Court of Human Rights." Peer Journal of Computer Science, 2, 93-99, 2016

[12] F. Kelbert, F. Shirazi, H. Simo, T. Wüchner, J. Buchmann, A. Pretschner, and M Waidner, "State of online privacy: A technical perspective." In Internet Privacy (pp. 189-279). 2012 Berlin/Heidelberg: Springer

[13] A. Završnik, "Algorithmic justice: Algorithms and big data in criminal justice settings" European Journal of Criminology Vol. 18(5) pp. 623–642, 2018

[14]. R. Simmons, "Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System". U.C. Davis Law Review, 52(2), pp. 1067-1118, 2018

[15] K. Ashley, "A Brief History of the changing roles of case prediction in AI and Law: Law in context" Socio-Legal Journal, 36(1), 93-112, (2019).

[16]. M. Medvedeva, M. Vols, and M. Wieling, "Using Machine Learning to predict decisions of the European Court of Human Rights". Artificial Intelliegence and Law, 28(2), pp. 237-266, 2020

[17] A. Kaur, & B. Bozic. Convolutional Neural Network-based Automatic Prediction of Judgments of the European Court of Human Rights. In AICS (pp. 458-469), 2019, July.

[18] C. O'Neil, "Weapons of Math Destruction: How Big Data increases inequality and threatens democrac". Crown, 2016.

[19]. I. Chalkidis, I. Androutsopoulos, and N. Aletras, "Neural Legal Judgement Prediction in English". 57th Annual Meeting of the Association for Computational Linguistics, pp. 4317-4323, 2019.

[20] H. Alla, L. Moumoun, &Y. Balouki. A multilayer perceptron neural network with selective-data training for flight arrival delay prediction. Scientific Programming, 2021, 1-12.

[21] M.B.L. Virtucio, J.A. Aborot, J.K.C Abonita, R.S. Avinante, R.J.B Copino, M. P. Neverida, & G.B.A Tan. Predicting decisions of the philippine supreme court using natural language processing and machine learning. In 2018 IEEE 42nd annual computer software and applications conference (COMPSAC) (Vol. 2, pp. 130-135). IEEE. July 2018.

[22] R.A. Shaikh, T.P. Sahu, and V. Anand, V, "Predicting outcomes of legal cases based on legal factors using classifiers". Procedia Computer Science, vol.167, pp.2393-2402, 2020

[23] M.F. Sert, E. Yıldırım, & I. Haşlak. Using artificial intelligence to predict decisions of the turkish constitutional court. Social Science Computer Review, 40(6), 1416-1435, 2022

[24] K. Kowsrihawat, P. Vateekul, & P. Boonkwan, predicting judicial decisions of criminal cases from Thai Supreme Court using bi-directional GRU with attention mechanism. In 2018 5th Asian Conference on Defense Technology (ACDT) (pp. 50-55). IEEE. October, 2018.

[25] Z. Liu, & H. Chen. A predictive performance comparison of machine learning models for judicial cases. In 2017 IEEE Symposium series on computational intelligence (SSCI) (pp. 1-6). IEEE.

[26] A. Sharma, A.S. Chudhey & M. Singh. Divorce case prediction using machine learning algorithms. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 214-219). IEEE. March, 2021.

[27] H. Zhong, Z. Guo, C. Tu, C. Xiao, Z. Liu, & M. Sun, M. Legal judgment prediction via topological learning. In Proceedings of the 2018 conference on empirical methods in natural language processing (pp. 3540-3549). (2018.

[28] S. Sengupta, & V. Dave, predicting applicable law sections from judicial case reports using legislative text analysis with machine learning. Journal of Computational Social Science, 1-14, 2021

[29] J.S. Howe, L.H. Khang, & L.E. Chai. Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments. arXiv preprint arXiv:1904.06470, 2019.

[30] S. Agrawal, E. Ash, D. Chen, S.S Gill, A. Singh, A; & K. Venkatesan. Affirm or Reverse? Using Machine Learning to Help Judges Write Opinions. In: (2017). [13] S. Bird, E. Klein, and E. Looper.

[31] B. Strickson, & B. De La Iglesia. Legal judgement prediction for uk courts. In Proceedings of the 3rd International Conference on Information Science and Systems (pp. 204-209). March, 2020.

[32] N. Sivaranjani, J. Jayabharathy, & P.C Teja, Predicting the supreme court decision on appeal cases using hierarchical convolutional neural network. International Journal of Speech Technology, 24(3), 643-650, 2021

[33] C. Chu, A. L Hsu, K.H. Chou, P. Bandettini, C. Lin, & Alzheimer's Disease Neuroimaging InitiativeDoes feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage, 60(1), 59-70, 2012

[34] M. Afshar, & H. Usefi. Optimizing feature selection methods by removing irrelevant features using sparse least squares. Expert Systems with Applications, 200, 116928, 2022

[35] Ali, R. (2020). Predictive Modeling: Types, Benefits, and Algorithms. Retrieved from Oracle Netsuite: https://www. netsuite. com/portal/resource/articles/financial-management/predictive-modeling. shtml.

[36] R.R. Quinlan, “Induction of Decision Trees”. Machine Learning 1: 81 – 106, 1996.

[37] L. Yu, and H. Liu, “Efficient feature selection via analysis of relevance and redundancy”. JMLR 5: pp. 1205 – 1224, 2004.

[38] A. Lage-Freitas, H. Allende-Cid, O. Santana, and O. Lıvia, “Predicting Brazilian court decisions” Found at: https://arxiv.org/abs/1905.10348

[39] L. Kostrzewa, & R, Nowak. Polish Court Ruling Classification Using Deep Neural Networks. Sensors, 22(6), 2137, 2022