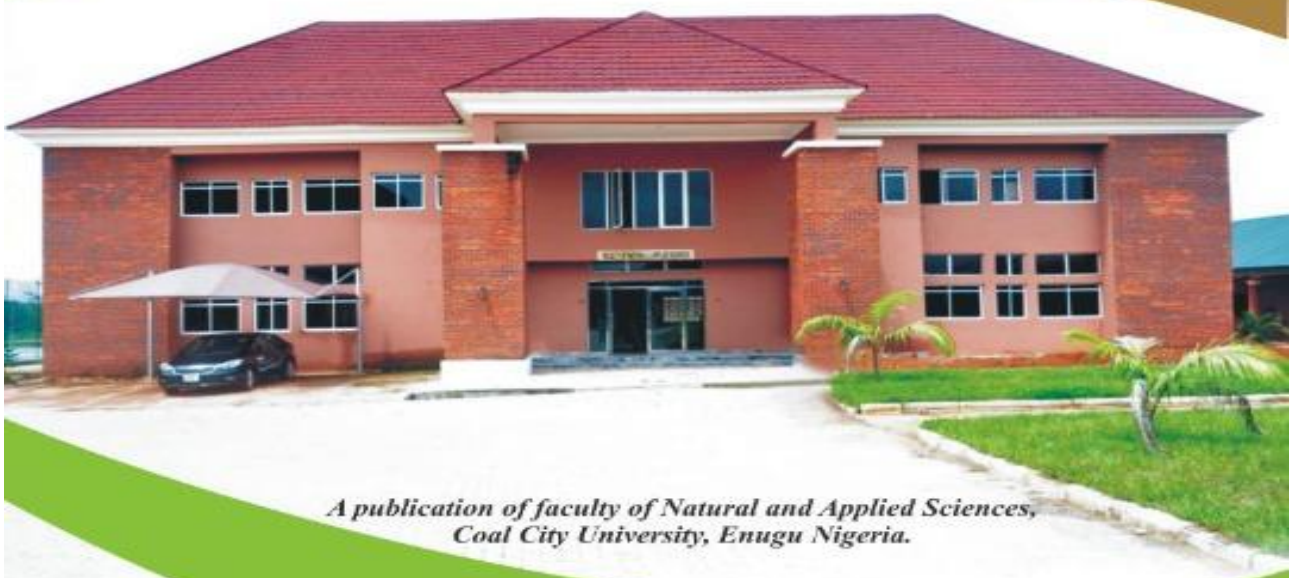
Coal City University Journal of Science

ISSN: 2734-3758(Print), 2734-3766 (Online)



A publication of faculty of Natural and Applied Sciences, Coal City University, Enugu Nigeria.

CCU Journal of Science
Vol. 3, Issue 1, July, 2023

ISSN: 2734-3758(Print), 2734-3766 (Online)
https://ccujos.com

DEEP LEARNING PREDICTING MODEL FOR COURT VERDICTS

[1]Etuk, Enefiok A. [2]Ugwu, C. and [2]**Onyejegbu L.N**

[1]Department of Computer Science, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria
etuk.enefiok@mouau.edu.ng; chidiebereugwu@uniport.edu.ng;

[2]Department of Computer Science, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria
laeticia.onyejegbu@uniport.edu.ng

Correspondence: **Etuk, Enefiok** (etuk.enefiok@mouau.edu.ng)

ABSTRACT

This paper presented deep learning prediction model for court verdicts. It leveraged on historic datasets of court cases from different countries to build the model. The datasets went through the pre-processing stages and the cleaned dataset was divided into training and testing for training and validation respectively. The bidirectional Long Short Term Memory Network (LSTM) and N gram models were used for model development. Mean square error was used as the loss function to monitor the variation between errors of target and real output, this was achieved with back propagation algorithm. ReLU activation function was used to interpret complex nonlinear function in the model, to improve on the convergence speed stability after converging to a local extreme minimum; Adam optimizer was used in the model. The results obtained show accuracy of 99.07% for training and 98.01% for validation and error loss of 0.002% and 0.003% respectively. The model was able to predict court verdicts crime types which shows that the system performed tremendously well and has the potential of assisting individuals and legal practitioners in predicting their case before approaching the court.

Keywords: Deep Learning, Loss Function, Activation Function Optimization Function

1.0 INTRODUCTION

Traditional machine Learning systems while being useful in the legal domain with promising results are still being challenged with some vital technical issues including running time (speed of processing), adaptability, misclassification and dimensionality issues(Kupiec, 1999). Most existing researches focused more on traditional machine learning approaches such as Support Vector Machine (SVM) and Logistic Regression (LR) or the now deep learning systems such as recurrent neural network, Convolution Neural Network (CNN) and Gated Recurrent Units

(GRUs).Why these approaches have given quite good accuracies, they may face optimality and run-time issues. In addition, the excessive use of pre-training and re-training with huge datasets by such techniques might not always be feasible in real world situations where timely interventions are desired, hence, the present study will design and implement an improved model for court cases prediction and sentencing exploring the use and possible modifications of more sophisticated deep learning models, (Kupiec, 2001)

Predicting court verdicts and sentencing using machine learning models has been cause for concern. Many machine learning models exists, but issues of misclassification and word collocation problems are always predominant in such models (Onyema et al, 2021). Misclassification occurs in this context when the machine learning models are not able to clearly separate these verdicts to appropriate classes therefore resulting in mis-judgments and wrong sentencing. Such models give unjustified sentencing to the innocent and accused persons. Again there are issues of mismatching of keywords by the existing models for text summarization leading to misjudgements and unwarranted penalties to the innocent victims. The techniques used to generate the word vectors have greatly affected the performances of the models leading to wrong judgments in the court and wasting innocent victims' life in the prisons for offences they did not commit.(Lage-Freitas, *et al.*,2019). So this is a huge problem that needs a robust model to deal with these issues of misclassification, collocation and inefficient text summarization as a result of poor word vector generation. This study proffered a robust Long Short Term Memory Network and n-gram based application for court verdict prediction and text summarization. This model has to a great extent handled the issues of misclassification leading to misjudgement and unmerited penalties to victims and collocation problems which normally affects the semantics of sentences because of inappropriate word generation in sentence formation.

2.0 RELATED WORKS

Previous research works relying on case specific details of legal cases for outcome prediction are taken into account. Zhong and Zhong (2018), used machine learning to select which sentences in the decision are predictive of the case outcome. The summarizer computes the relative importance of sentences in a legal case document, as measured by their predictiveness and chooses a subset to generate the summary. They partitioned acceptable sentences as classified by type (i.e., Reasoning or Evidential Support sentence) and chose a set of summary sentences using maximum marginal relevance. They concluded, based on a detailed error analysis, that argument mining techniques would be required to identify more conceptual aspects of the decisions. Liu and Chen (2018), used 584 documents to compare five ML techniques k-NN, logistic regression, bagging, random forests and SVM to predict ECtHR judicial decisions. The authors used spectral clustering and N-grams to extract textual representations of topics, and they concluded that the SVM outperforms the other models. The authors also predicted law violation and non-violation using auto-sk learn to build models for 12 articles in the ECHR. The authors used n-grams, word embeddings (echr2vec) and doc2vec for the vectorization task for training Gradient Boosting (GB), Random Forest (RF), Stochastic

Gradient Descent (SGD), Decision tree (DT), and Quadratic Discriminant Analysis (QDA). The authors obtained an average of 68.83% accuracy for the different models.

Ruger et al. (2004), developed an algorithm that can predict the individual votes of the nine justices as well as the final direction of the court's decisions, that is, confirmation or revocation in the U.S. Supreme Court. This experiment was conducted with information obtained from the court for two years (2002–2003), and 628 cases were analyzed. The algorithm was based on a classification tree of the following six variables: the federal circuit in which the case originated, thematic area, type of plaintiff, type of defendant, ideological direction, and whether or not the constitutionality of a rule or practice was challenged. The results obtained were compared with the predictions made by a group of academics and lawyers. The following results were obtained: of 78 cases reviewed in 2002–2003, the algorithm could predict 75% of the court decisions and 66.7% of the individual votes, whereas the experts correctly predicted 59% of the decisions and 67.9% of the individual votes.

Katz et al. (2017), designed a model based on the Random Forest algorithm to predict the behavior of the U.S. Supreme Court using a time-evolving Random Forest classifier on a corpus of 200 documents, with an accuracy of 70.2%. It used different algorithms to predict court decisions in matters of public morality and freedom of speech using a corpus from the Turkish Constitutional court composed of 92 and 338 legal documents, respectively. They used an embedding representation to perform TF-IDF and Bag-of-Words for inputting the text to a Multi-Layer Perceptron with different architectures. The F-measure results obtained are between 60% and 98.7%.

Virtucio et al. (2016), proposed using Linear SVM and Random Forest classifiers to predict the decisions of the Supreme Court in the Philippines. They analyzed the Historical Philippine Supreme Court case decision and the Lasphil Project to gather 27, 492 cases divided into the following four categories: person, property, public order and drugs. They characterize each use case using the following: case title, case type, year, decision, classification, laws (republic, act, presidential, commonwealth, article, crime) and crime category. They used a subsampling technique to balance negative cases. They used Bag-of-Words and n-gram representations to model the documents, reaching accuracies of 55% and 59% for Linear SVM and Random Forest, respectively.

Onyema et al (2023) examined the utility and perception of mobile technology among law students in Enugu Nigeria. The study highlighted the need for the legal profession to adopt relevant emerging technologies to enhance effectiveness in legal or court proceedings. The study concluded by recommending the teaching of digital skills and use of technologies in Nigeria legal system.

Sulea et al. (2017), proposed predicting the law category, court ruling, and time of the decision of the French Supreme Court. The dataset used for these three prediction tasks was a diachronic collection of rulings from the French Supreme Court (Court de Cassation) in XML format, containing 126,865 unique documents after the cleaning phase. They use Bag-of-Words, 2-gram and 3-gram as inputs for a linear SVM classifier implemented in Sckit-learn for the different tasks. They report the results using precision, recall, accuracy and F-measure.

They sampled 200 documents for the eight different classes in the law prediction task. The F-measure obtained for this task is 90.3%.For the ruling decision prediction, the SVM algorithm obtained an F-measure of 97% and 92.7% when predicting 6 and 8 classes, respectively. The authors used 1-gram and 2-gram representations for the linear SVM in the last task of temporal prediction, achieving 73.2% and 73.9% when predicting 7 and 14 classes, respectively. The works reviewed did not consider issues of collocation which has drastically affected the abilities of these models to deal with issues of semantics and also minimize the misclassification. This paper will take care of the issues as were mention above.

3.0 MATERIALS AND METHODS

The system architecture is shown in figure 1. It is made up of components of the following components; dash board, pre-processing phase, training and classification phase. The dash board provides the connection for user/system interaction. It is a flexible communication environment. The home page is the first and introductory page that opens for the user which has register page, login page and classification data page. The dataset was collected from Sherlocrepository court judgments dataset that captured convicts, acquittals and sentences from many countries with size above ten thousand. It is made up many features and labels, table 1 shows a sample of the dataset from the repository. The dataset was pre-processed to make the data clean for model building. The following are the steps;Text Normalizationwas done to transform the text in a document in order to make its contents consistent, convenient and full words for an efficient classification task. It assisted in transforming all text cases to lower case and removal of diacritics and noisy data. Irrelevant features were expunged using a function in "Pandas" Library in python called "Drop" to drop irrelevant and unnecessary columns.

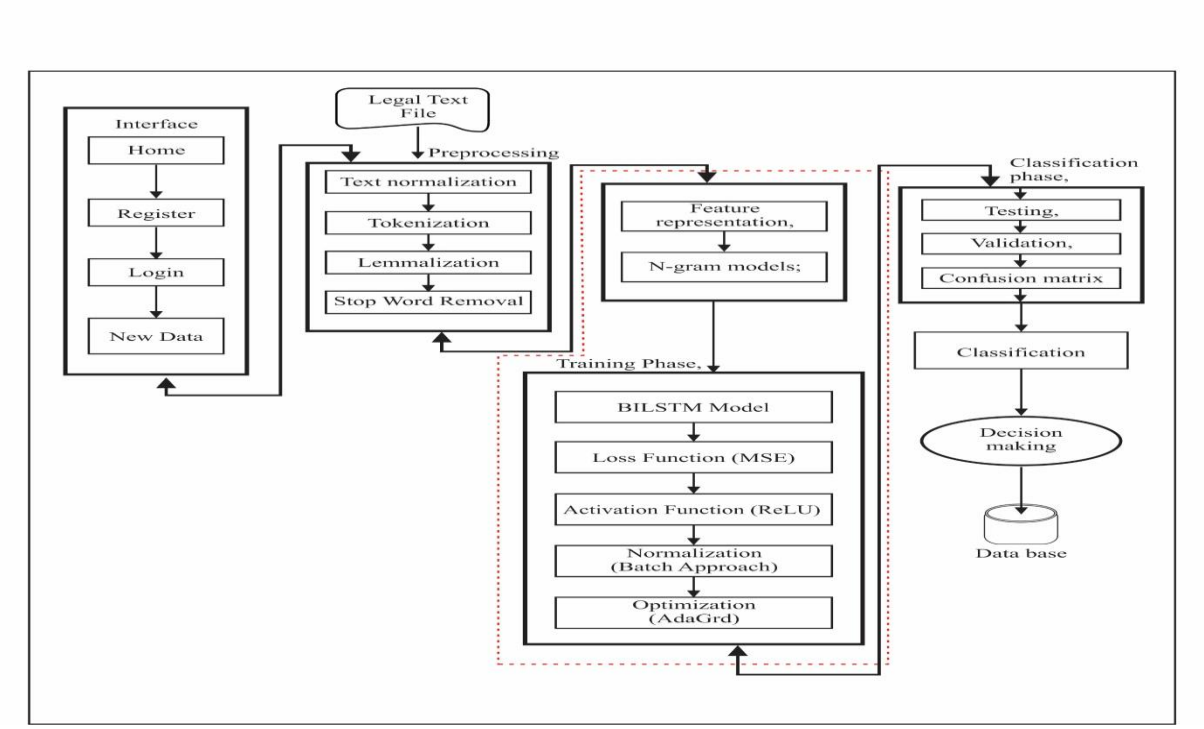


Figure 1: System Architecture

Table 1: Sample of Dataset

country_code	title	country	verdict_date	sentence_date	text	crime
NGA	JN alias GU	Nigeria	Verdict Date:2009-01-23	Sentence Date:	Cocaine has been trafficked by an organised	Drug-offences
NER	Cour suprême, chambre jud	Niger	Verdict Date:2006-04-27	Sentence Date:2006-04-27	Le 17 septembre 2004, M. A. Aa a été arrêté	Drug-offences
USA	US v Marino-Garcia	United States of Am	Verdict Date:1982-07-09	Sentence Date:	Two cases have been consolidated for the pu	Drug-offences
RWA	IKIZA RY' URUBANZA RP/P	Rwanda	Verdict Date:2021-06-30	Sentence Date:2021-06-30	This case concerns a notorious criminal grou	cybercrime
ARG	Mazza, Valeria Raquel c Ya	Argentina	Verdict Date:2021-06-24	Sentence Date:	Valeria Raquel Mazza, modelo publicitaria y e	cybercrime
NZL	Ortmann et al v the United S	New Zealand	Verdict Date:2017-02-20	Sentence Date:	In 2005, Mr Dotcom developed a business un	cybercrime
PHL	Operation Strikeback	Philippines	Verdict Date:	Sentence Date:	INTERPOL-led operation "Strikeback" target	cybercrime
RUS	Case no. 10-2501	Russian Federation	Verdict Date:2019-03-26	Sentence Date:2019-03-26	The Moscow District Court Izmaylovski found	counterfeiting
SEN	Cour suprême, chambre cri	Senegal	Verdict Date:2016-05-06	Sentence Date:2016-05-06	No text	counterfeiting
SEN	Cour suprême, chambre cri	Senegal	Verdict Date:2016-05-06	Sentence Date:2016-05-06	No text	counterfeiting
SEN	Cour suprême, 6 mai 2016,	Senegal	Verdict Date:2016-05-06	Sentence Date:2016-05-06	M. X., le prévenu, a vendu en Europe des pr	counterfeiting
ITA	Mazzarella - Operation Silk	Italy	Verdict Date:2014-01-07	Sentence Date:	The present case originates from the prelimir	counterfeiting
NGA	FEDERAL REPUBLIC OF N	Nigeria	Verdict Date:2006-01-13	Sentence Date:	No text	counterfeiting
MCO	Société Orach Placements S	Monaco	Verdict Date:2015-06-09	Sentence Date:	Les autorités sénégalaises ont émis une den	corruption
CAN	R. v. B.R.	Canada	Verdict Date:2014-03-13	Sentence Date:2014-04-04	B.R. was a police officer with the Montreal Po	corruption
PHL	Spouses PNP Director Elise	Philippines	Verdict Date:2009-02-13	Sentence Date:2009-02-13	On October 6, 2008, Gen. Dela Paz, a senior	corruption

	country	sentence-date	text	crime types
0	South Africa	2021-02-12	Mr. Solomon Sauls ran an illegal enterprise wi...	corruption
1	France	2020-06-15	Dans le cadre d'accords de coopération et d'as...	corruption
2	United States of America	2020-01-10	Baktash Akasha Abdalla and his brother, Ibrahi...	corruption
3	Antigua and Barbuda	NaN	In July 2014, Creswell Overseas S.A., a corpor...	corruption
4	International and Regional Bodies	NaN	The accused in this case, Mr. Teodoro Nguema O...	corruption
...	...	...	...	...
1853	El Salvador	NaN	No text	drug offences
1854	Brazil	NaN	No text	drug offences
1855	Italy	NaN	Italian autorithies found out the existence of...	drug offences
1856	Other	NaN	In November 2014, law enforcement and prosecut...	drug offences
1857	Andorra	2015-05-04	In the framework of an operation to combat dru...	drug offences

1858 rows × 4 columns

Figure 2: Sample of Dataset after removing redundant Features

Figure 2, shows sample of the dataset after removal of irrelevant features in table 1. This process is called dimensionality reduction. It provides datasets that can be used to build more reliable model. The clean datasets was visualized for easy identification of data trends, which would otherwise be a hassle. The pictorial representation of datasets allows one to visualize the concepts and new patterns. The graph of count of crime against crime types in figure 3shows the different distribution of the crime types which indicated that some crimes have limited number compared to others. Hence, some crimes were removed to avoid imbalance in data distribution.

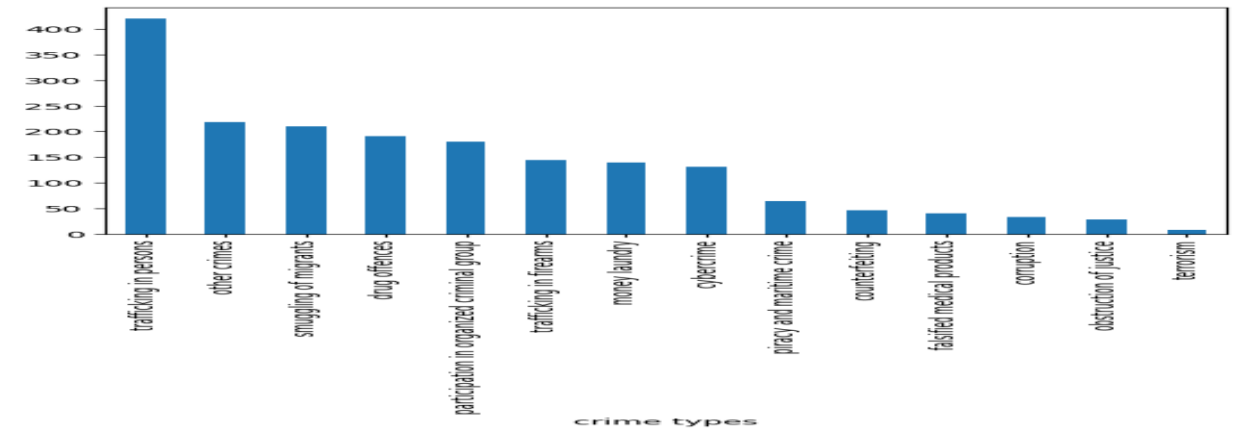


Figure 3: Graph showing the distribution of the dataset

3.1 MODEL BUILDING

The LSTM model was used in this paper because of its ability to keep the information for a longer time, this facilitated the summarization process of court verdicts. This is against other deep learning architectures that were used for text processing and mining, predominant of them is recurrent neural network which the memory cell has a limited capacity. The architecture does not have the ability to memorize all the information in the entire sequence. The bidirectional LSTM, or bi-LSTM, was used which a sequence is processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backwards direction. Bi-LSTMs effectively increase the amount of information available to the network, improving the context available to the algorithm, for example it takes account what words that will immediately follow and precede a word in a sentence for the mined and summarized text in the legal case document. Unlike standard LSTM, the input flows of Bi-LSTM is in both directions, and it's capable of utilizing information from both sides. It's also a powerful tool for modeling the sequential dependencies between words and phrases in both directions of the sequence. Therefore, the key information in the mining and summarization processes were selectively remembered using the three units of computation in bidirectional Long Short-Term Memory (LSTM). This is in variance with the recurrent neural network which also has the vanishing gradient problem. Long Short Term Memory (LSTM) was designed to overcome the problems of simple Recurrent Neural Network (RNN) by allowing the network to store data in a sort of memory that it can access at a later times. Figure 4 shows the architecture of LSTM.

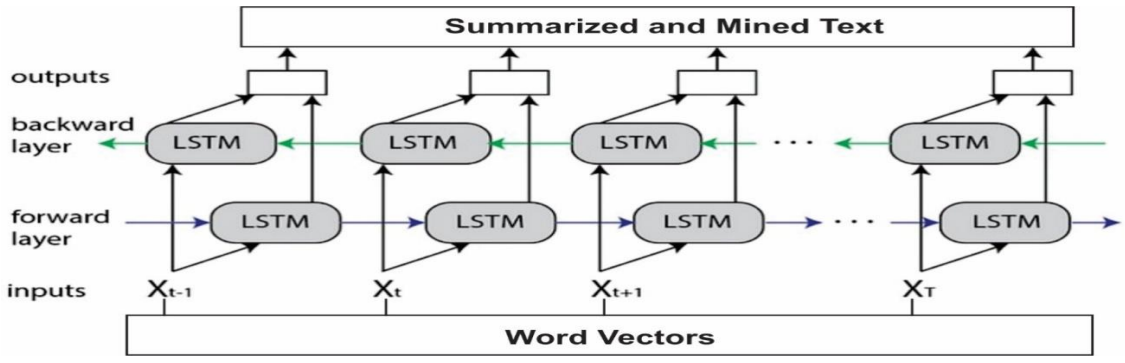


Figure 4: Architecture of LSTM

At each training step, the result of the n gram models were served as input to the LSTM model, the model processed the data using the cell state and stored the result in the hidden layer. The process was repeated for eight times with the result saved in the hidden state. Following these stages of model operations, the status of classification of court verdictwere achieved as convicted and not convicted. The core of training a machine learning model is loss function, and for this model mean square error (MSE) was the loss function that used during the training. This is the most commonly adopted loss function. The function was used to monitor the variations between the errors of targeted output and real output of classification, it can also be called a loss function (or error function).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2 \qquad \text{(eq 1)}$$

We were able to reduce the loss function with the use of back propagation algorithm which is a form of steepest-descent algorithm to facilitate training and retraining of the model where the loss function is very high. The algorithm was used to adjust the weights in the input layer and hidden layer in order to reduce the loss function. The loss function reflected the error between the target output and the actual output value of the perceptron. The algorithm is as flexible as it does not require prior knowledge about the network, very simple, fast and easy to program.

3.2: ALGORITHM OF THE SYSTEM

Algorithm 1 Back propagation Algorithm

1. Procedure TRAIN
2. $X \leftarrow$ Training Data Set of size mxn
3. $y \leftarrow$ Labels for records in X
4. $w \leftarrow$ The weights for respective layers
5. $l \leftarrow$ The number of layers in the neural network, 1…L
6. $D_{ij}^{i} \leftarrow$ The error for all l,i,j
7. $t_{ij}^{i} \leftarrow$ 0. For all I,i,j
8. For i = 1 to m
9. $a^{i} \leftarrow$ feedforward $\left(x^{(i)}, w\right)$
10. $d^{i} \leftarrow$ a(L) – y(i)
11. $t_{ij}^{l} \leftarrow t_{ij}^{l} + a_j^{l} - t_{ij}^{l+1}$
12. If j ≠ 0 then

13. $D_{ij}^{(l)} \leftarrow \frac{1}{m} t_{ij}^{(l)} + \lambda w_{ij}^{(l)}$

14. Else

15. $D_{ij}^{(l)} \leftarrow \frac{1}{m} t_{ij}^{(l)}$

16. where $\frac{\partial}{\partial w_{ij}^{(l)}} J(w) = D_{ij}^{(l)}$

The rectified linear function (ReLU) activation was used in the work which currently is the most widely applied activation function. The ReLU was used to interpret complex nonlinear functions in the LSTM model. Without the activation function, a neural network can only represent one linear function, no matter how many layers it has. ReLU is not bounded by upper limit, so neurons will never reach saturation, which effectively alleviates the vanishing gradient and can converge much faster in gradient descent. Other activation functions such as sigmoid, tanh ,Softsign etc. all require exponential equations, which is quite computing intensive .The data expansion was used to prevent overfitting by increasing the size of training set as the probability of overfitting will decrease with the increase of training set. Also, embedding layer and spatial dropout was used in the model to reduce overfitting in the training of Long Term-Short Memory model.To improve on the convergence speed, stability after converging to a local extremum, and the efficiency of adjusting hyper parameters, Adam optimizer was used. The Adam (Adaptive Moment Estimation) optimizer is a popular optimization algorithm in machine learning, particularly in deep learning applications. It combines the benefits of two other optimization techniques; Momentum and Adaptive Gradient Algorithm (AdaGrad) to provide an efficient and adaptive update of model parameters. By computing both first-order momentum (moving average of gradients) and second-order moment (moving average of squared gradients) of the loss function, Adam adjusts the learning rate for each parameter individually, ensuring a smooth and fast convergence. This optimization technique has gained popularity because of its adaptive learning rates, robustness to noise, and suitability for handling sparse gradients, making it an acceptable choice for training various machine learning models, including neural networks.

$$w_{t+1} = (1 - \lambda) w_t - \eta \nabla f_t(w_t) \qquad \text{(eq 2)}$$

4.0 RESULTS AND DISCUSSION

This section captured in details the setup used for the experiments, and presentation of some very important results obtained from the developed model. In developing and implementing the model, bootstrap framework, flask framework, python programming language and MySql Database were used. The development tools and environment used were Jupyter Notebook, Spyder and Anaconda (Python Distribution).

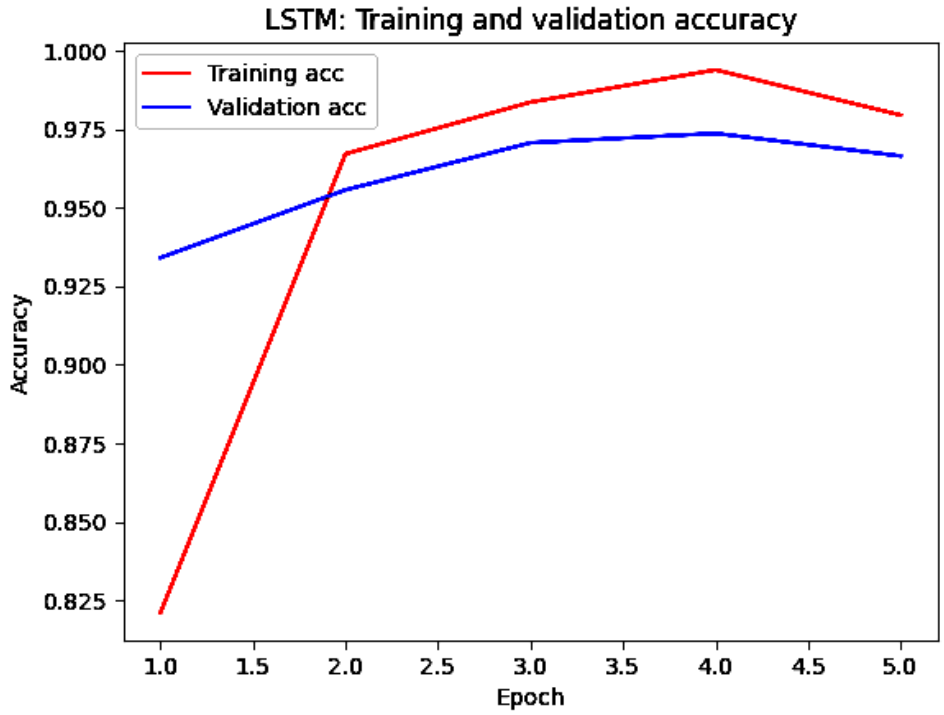


Figure 5: Accuracy of the trained Model

From the graph on figure 5, the model performed during training was depicted. The model achieved an accuracy of 99.07% for the training data and about 98.01% for the validation or testing data. The blue line represents the model training accuracy, whereas the orange line represents the validation test accuracy. To validate model performance, the testing dataset was used. The line graph shows the performance of the model at each training step. The line graph shows the performance of the model at each training epoch. At epoch 1, the training performance of the model was 94.11% and the validation score was 99.41%, at epoch5, the training performance of the model was 99.67% and the validation data was 98.11%, at epoch 10, the model had training performance of 99.07% and the validation score was 98.01%.

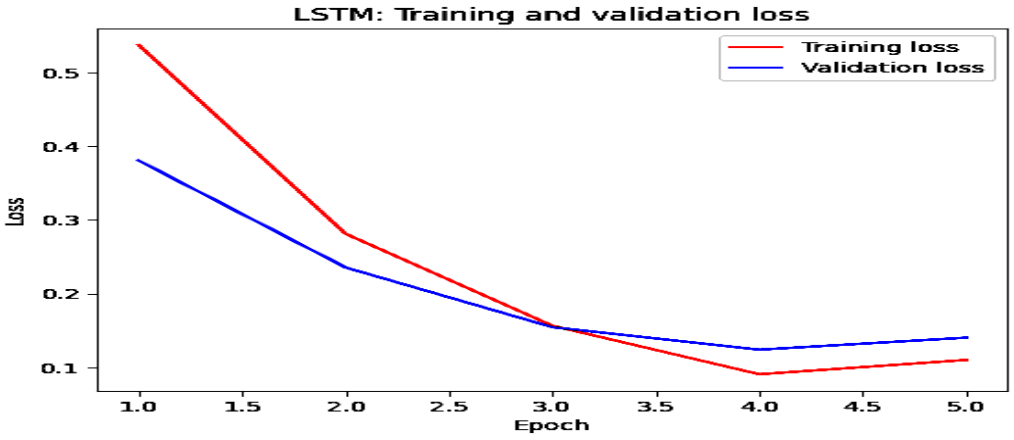


Figure 6: Model loss

Figure 6,depicts the losses witnessed by the model during training and testingphases in the model building. The green line indicates the loss witnessed by the model during training, and the orange line indicates the loss acquired by the model during testing. The loss values are acquired at each training steps, starting from epoch 1 to epoch 20. This depicts that model achieved a loss value of about 0.002% for the training data and 0.007% for the validation or testing data.

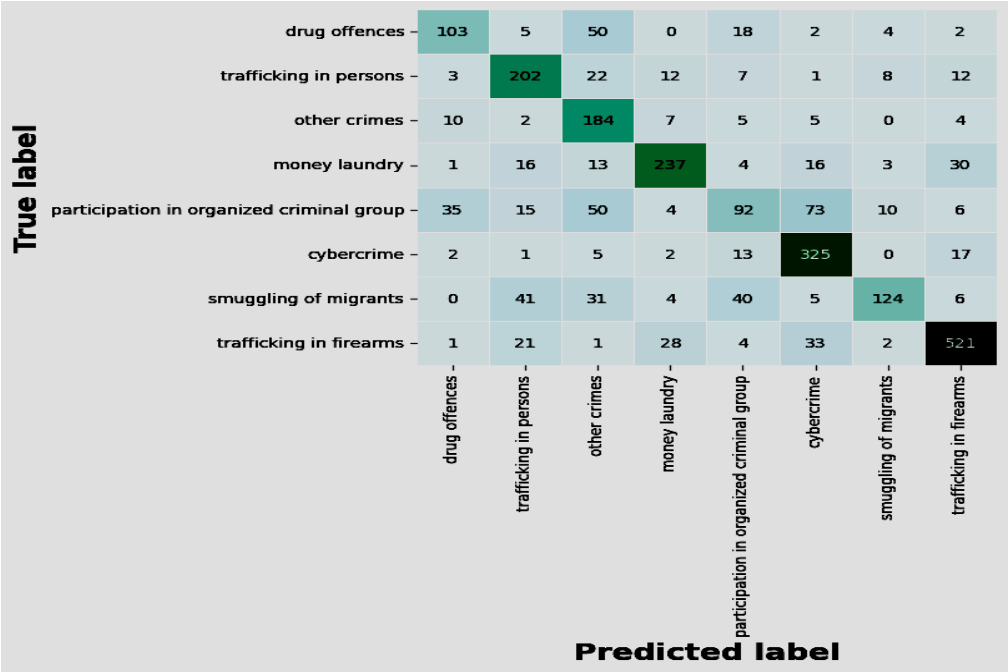


Figure 7: Confusion Matrix for Classification

Figure7, is the confusion matrix for the true labels and predicted labels of the diffèrent crime catégories by the model such as Drug offenses, Money Laundery, Cyber crime, Smuggling of Migrants, Trafficking in Firearms, Organised crimes and other crimes. Figure 8 shows the predictions of categories of crimes.

```
Test: drug offences -------- Predicted: (drug offences)
Test: trafficking in persons -------- Predicted: (other crimes)
Test: trafficking in persons -------- Predicted: (trafficking in persons)
Test: drug offences -------- Predicted: (drug offences)
Test: other crimes -------- Predicted: (trafficking in persons)
Test: other crimes -------- Predicted: (smuggling of migrants)
Test: money laundry -------- Predicted: (money laundry)
Test: participation in organized criminal group -------- Predicted: (money laundry)
Test: cybercrime -------- Predicted: (cybercrime)
Test: cybercrime -------- Predicted: (cybercrime)
Test: smuggling of migrants -------- Predicted: (smuggling of migrants)
Test: cybercrime -------- Predicted: (money laundry)
Test: drug offences -------- Predicted: (drug offences)
Test: drug offences -------- Predicted: (drug offences)
Test: other crimes -------- Predicted: (smuggling of migrants)
Test: trafficking in persons -------- Predicted: (trafficking in persons)
Test: participation in organized criminal group -------- Predicted: (cybercrime)
Test: money laundry -------- Predicted: (money laundry)
Test: participation in organized criminal group -------- Predicted: (smuggling of migrants)
Test: trafficking in persons -------- Predicted: (drug offences)
Test: smuggling of migrants -------- Predicted: (smuggling of migrants)
Test: trafficking in persons -------- Predicted: (trafficking in persons)
Test: trafficking in firearms -------- Predicted: (participation in organized criminal group)
```

Figure 8: Prediction of Crime Categories

5.0: CONCLUSION

The paper developed a model for the prediction of court verdicts using a historic dataset of court cases from different countries assembled in a repository as mentioned in the section of materials and method. One of the foremost deep learning architectures; Long Short Term Memory Network (LSTM) was used in the model building. This model has cleared to some extent,certainties in knowing what the outcome of a case will be and the actions that will be taken after during the judgement in the court. The issues of lack of memorisation as experienced by other models in the literature were addressed. Also addressed were problems of misclassification and collocation issues. This will properly guide the individuals and legal teams on whether to pursue cases or not.

REFERENCES

Katz, D. (2014). Predicting the behaviour of the Supreme Court of the United States: A general approach. *International Journal of Computer Science and Mobile Computing (IJCSMC)*, 2(4), 102 – 208

Katz, D.M., Bommarito, M.J.,Blackman, J. (2017) A generalapproach for predicting the behavior of the Supreme Court of the UnitedStates. PLoS ONE 2017, 12, e0174698.

Kupiec J., (1999) "Robust part-of-speech tagging using a hidden Markov model", *Journal of the American Society for Information Science*, 50(2): 151-161.

Kupiec K. M., Pedersen J. and Chen F. (2001) “A Trainable Document Summarizer”, *Languages, and Computation*, Addison Wesley.

Lage-Freitas A., Allende-Cid H., Santana O., and de Oliveira-Lage L., (2019). Predicting Brazilian court decisions. arXiv 2019, arXiv:1905.10348.

Liu, Y.H., Chen, Y.L. (2018) A two-phase sentiment analysisapproach for judgementprediction. *J. Inf. Sci.* 2018, 44, 594–607.

Onyema, EM; Chinecherem, DE; Ugboaja SG; Ezemoyih, CM; Madubuezi, CO; and Richard-Nnabu, NE (2023). Utility and Perception of Mobile Technology among Law Students in Enugu Nigeria. Babcock University Journal of Education, 9 (3), 56-71. https://journal.babcock.edu.ng/article/07101026-49e3-4165-89b1-399ec6a2b524

Onyema EM., Ugorji, C. C., Nduanya, U. I., Onyewuchi, C., Ohwo , S. O., & Ikedilo, O. E. (2021). Prospects and Limitations of Machine Learning in Computer Science Education. Benin Journal of Educational Studies, 27(1), 48–62. Retrieved from http://beninjes.com/index.php/bjes/article/view/70

Ronde K. de, (2021). Classifying Dutch Fiscal Case-Law Articles Using Natural Language Processing. Master’s Thesis, Erasmus University Rotterdam, Rotterdam, The Netherlands.

Ruger, T. W., Kim, P.T., Martin, A. D., & Quinn, K. M. (2004). The Supreme Court forecasting project: Legal and political science approaches to predicting Supreme Court decisionmaking. Columbia Law Review, 1150-1210.

Sivaranjani N., Jayabharathy J., and Teja P. (2021), Predicting the supreme court decision on appeal cases using hierarchical convolutional neural network. *Int.J. Speech Technol*. 2021(24) 643–650.

Strickson B., De La Iglesia B., (2020). Legal Judgment Prediction for UK Courts. In Proceedings of the 2020 the 3rd International Conference on Information Science and System, Cambridge, UK, 19–22 March 2020; 204–209.

Sulea, O. (2017). Predicting the Law Area and Decisions of French Supreme Court Cases.*International Journal of Computer Application (IJCA)*, 114(4), 16 – 18

Sulea, O.M., Zampieri, M., Vela, M., van Genabith, J. (2017) Predicting the law area and decisions of French supreme court cases. arXiv 2017, arXiv:1708.01681.

Visser P.R.S., (1995). *Knowledge Specification for Multiple Legal Tasks; A Case Study of the Interaction Problem in the Legal Domain*, Computer/Law Series, No. 17, Kluwer Law International, The Hague, The Netherlands.

Virtucio, M.L., (2016), Predicting Decision of the Philippine Supreme Court using Natural Language Processing and Machine Learning.

Zhong H., Guo Z., Tu C., Xiao Liu C., Z., Sun M., (2018). Legal judgment prediction via topological learning. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; 3540–3549.