

Development of long-short term memory neural network (LSTM) based crime prediction model

Haruna, U. D.

Department of Information Technology, Modibbo Adama University Yola, Adamawa State Nigeria

Paper History

Received: 01st May, 2024

Accepted: 21st July, 2024

Published: July, 2023

Abstract:

Data mining is used to detect patterns in the dataset within the labelled data and also predicts future trends by unraveling the latten relationship existing within data that allow organizations to make informed decisions. Vanishing gradient problems within the data mining process serve a major constrain in crime prediction. This study adopts long short-term memory neural network (LSTM) that handles the vanishing gradient problem within the data mining process to predict future crime. The model was implemented as an intelligent system using Python. Also, the model was evaluated with 140,000 (one hundred and forty thousand) crime datasets from north-wale and north-Yorkshire police stations in the United Kingdom. The model was able to predict the crime rate for the supplied predicted year as against the dataset for the index year where it shows prediction accuracy of about 92.9%. The study recommends the full adoption of the system.

Corresponding author

Haruna, U. D.

dahiruusman222@mau.edu.ng

Keywords: Data mining, LSTM, Python

1. Introduction

Data mining computerizes the detection of patterns in a dataset, using algorithms and approaches to look into labeled and unlabeled data that can be processed to predict upcoming trends. DM component predicts future activities by reading through the dataset for latten design. It allows organizations to make informed decisions and find solutions to questions that are too time-consuming to handle. DM monitors each similar design from the data collected and helps in predicting the performance [1]. Most security institutions gather data because it is required for some operational use and once it has served its purpose, it is left to languish on tape or disk or is discarded [2]. However, such data can be mined to find hidden patterns or trends that help better the organization and even the society at large. These deficiencies have shown the need to employ a systematic and intelligent technique for crime investigation.

The integration of labeled datasets into structured formats can add value to most organizations. It also provides decision-makers with better insight. The traditional databases, which have been the best means for storing data for many decades are no longer fit to the data warehouse because of the velocity and variety of data. Companies such as Google and Amazon use DW because they are schema-free. As such, there will be a profound addition in data that will be produced in and out of the establishment which will need a competent DW to manage the exponential growth. Though organizations integrate huge amounts of the dataset which has been rich and growing significantly. Organizations require tools to generate insight from the labeled data which will be used for the decision process [3]. Therefore, it is against these backgrounds that this study seeks to look into the activities involved in integrating independent data origins

that may include unstructured and structured datasets to make prediction to curtail crime and crime related issues. Yang, *et al.* [4] also proposed an artificial neural network DW-based framework to perform the identification and counting of the component using the DW data set. The model reveals an 80% accuracy level using a small dataset but fails to show a significant percentage when a larger dataset is subjected to it because of the vanishing gradient problem. Thus, the need to look at the problem of duplication, errors, and vanishing gradient that serves as constrain to providing quality data in DW for effective crime prediction. This study proposes an improved data warehouse framework that resolve the vanishing gradient problem. The framework attempt resolving vanishing gradient problems to produce efficient data mining framework for crime prediction. The aim of this study is to developed Long Short-Term Memory Artificial Neural Network (LSTM) based prediction model for crime prediction. These can be realized through development of data mining management framework for prediction with the vanishing gradients handler using a LSTM.

2. Data mining

Data mining (DM) is one of the means of hidden cognition extraction within data with the help of artificial intelligence techniques and machine learning [5]. Similarly, Prakash and Rangdale [6] defined data mining (sometimes called data or knowledge discovery) as the process of analyzing data from various perspectives and summarizing it into vital information to gain revenue, sales, cut costs, or both. However, Rajab and Masoud [7] view DM as a process of analyzing data from several perspectives and discovering patterns, imbalances, and correlations in data sets that are

insightful and helpful for foretelling results that help you make the right decision.

DM is a knowledge discovery process aimed at extracting useful information from huge amounts of data gathered from various databases in the form of warehouse data. The aim of the process is decision-making [8]. Furthermore, DM can be used to decide between alternatives taking into consideration other issues relating to the decision in question [9].

2.1 Data mining process

Ali [10] looked at the Data mining process as a complicated process that has activity loops that make it an iterative procedure that includes Problem definition, Data Preparation, Data Exploration, Modeling, Evaluation, and Deployment while Rajab and Masoud [7] share the same view as relate to mining process but present the data mining process in the kind of a model even though as shown in Figure 1:

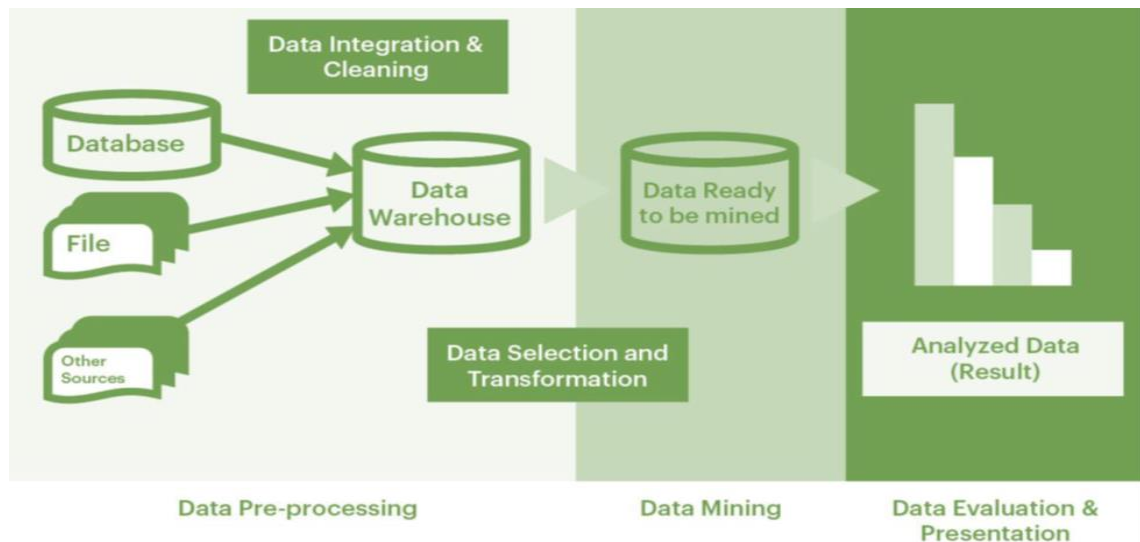


Figure1: Data mining process model. Source: [7]

2.2 Data Mining Techniques

Several data mining methods can be used to understudy the hidden design that exists on the existing repository and also distinguish the relationship between the datasets. Qayyum and Dar [5] described entity extraction, clustering techniques, association rule mining, sequential pattern mining, deviation detection, classification technique, String comparator, Social network

analysis, and artificial neural networks as data mining techniques while Tyagi and Sharma [11] are of the view that entity extraction, clustering techniques, and association rule mining are techniques for mining but Sayed, *et al.* [12] believed that various techniques are accessible but artificial neural network is the best because it analyzes textual material best as compared to other techniques. However, Reddy, *et al.* [1] are of the view that the techniques can be summarized as shown in Figure:

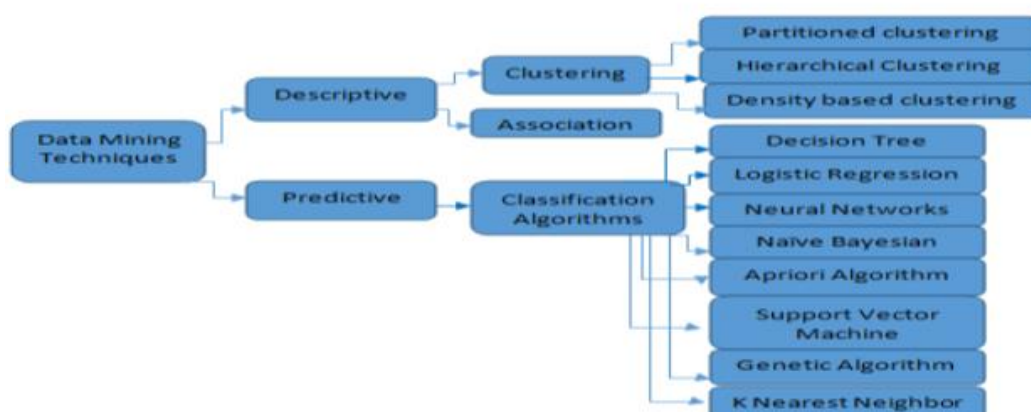


Figure 1: Data Mining Techniques. Source: [1]

2.3 Data Mining Models and Frameworks

Devakunchari, *et al.* [13] proposed a model that presumes the number of arrests that can be expected to be

made for every crime, age-group-wise, and state-wise using regression. The model was able to predict crime across the country age wise but the main weakness of the model is that

the prediction was not dependable because the data was not preprocessed enough to be utilized for framing new laws, preventing new crimes, and bringing into place new strategies to curb these actions.

Yang, *et al.* [4] build a crime-predicting framework with clustering as a data analyzing technique in unsupervised data form. The framework provides security for the dataset when outsourcing. Classification and Clustering are made on information while classifying the data, watermark subject was used for validating the classified data. The framework reveals that crime varies over time and increases continuously, it identifies crime behavior and also manages a huge amount of crime data with a definite degree of data security. The framework detects only data that have no relationship. The outcome of this framework is more dependable compared to that of [13] because a watermark was used to maintain a certain level of data integrity, though the framework will have used another mechanism to achieve a high level of data integrity.

Saltoş, and Cocco [14] come up with a framework for predicting the frequency of various types of crimes and the frequency of anti-social behavior. The model uses an instance-based learning algorithm and Linear Regression. The methodology adopted is CRISP-DM (Cross Industry Standard Process for Data Mining), which was discovered in comparison with other data mining methodologies to be well suited for predictive tasks for crime data. The results show that the decision tree was used to dependably predict crime frequency at large, as well as anti-social behavior frequency. The framework was unable to manage the time frame for prediction.

Fatlawi [15] went ahead to build a framework for crime data analysis using association, link analysis, classification, predictions, and outliers. This helps anti-crime specialists discover new patterns and hidden trends, make predictions, and possible explanations, and find relationships. Also applying criminal networks and likely identification of suspects. The framework predicts crime as relating to age. Discoveries of this work show classification rules were used to recognize time and location as suggested by Saltoş, and Cocco [14]. Pre-processing was adopted as an isolated entity to get an accurate and clean dataset which makes the result of [13] framework unreliable. Relationship between attributes was discovered, and crime types based on age taking history consideration was achieved.

Mowafy and Rezk [16] built a prediction framework using the sci-kit-learn Python Toolkit to predict crime types for improving decisions and lessen investigation efforts. It applied an unstructured text classification model which counts on adding the Multinomial Naive Bayes as a classification technique, and TF-IDF as a vector space model for text extraction for more accurate text classification results. The framework was able to predict the crime type but cannot provide a crime hotspot and cannot predict the next expected crime region. There is a significant advancement from the other research because the framework uses an unstructured dataset, but is unable to clean the data because the model has no relationship with data warehouse, which serve as the data basis for the prediction.

Soni, *et al.* [8] designed a framework that used an optimized k-means algorithm for the clustering tasks and a Support Vector Machine (SVM) to predict where the next crime will occur in the form of clusters mapped on a boxplot presented in the form of a graph. The prediction was achieved using a structured data set but time was unable to be predicted while [17] designed a mining prediction framework using a Deep Neural Network (DNN) which facilitates effective and accurate prediction of crime happenings. It works well on high dimensional and multi-model data and prediction cannot be practiced on insufficient data and it cannot provide clear information relating to a specific crime type at a given time.

Sivarajani, *et al.* [18] came up with a framework for crime forecasting and prediction in the city of Tamilnadu that uses a clustering algorithm which is used to combine crime happenings based on some predefined instances. The result of the K-Means clustering algorithm is visualized using Google Maps. The K-nearest neighbor (KNN) classification is used for crime prediction. The performance of every clustering algorithm is analyzed using or through metrics such as recall, precision, and F-measure, and the results are compared. The drawback of the framework is that agglomerative clustering will never redo any step (once it is attached object cannot be separated) and criminals were not identified. Tyagi and Sharma [11] came up with a framework for crime data analysis. The framework uses machine learning. The data sources for the framework are telephone records, social media data, and financial transaction data for the criminals.

3. Long Short-Term Memory Neural Network

Li, *et al.* [19] observed that a typical ANN consists of 3 layers (Input, Hidden, and Output layer) but it does not handle historic data because of independency in ANN layers. As such Recurrent Neural Network (RNN) becomes an extension to ANN where its layer cannot handle long sequences with the vanishing gradient problem (a problem that occurs while moving from layer to layer) problem appearing when the size of data increases. Then an improved neural network called the long short-term memory (LSTM) neural network was introduced. It contains memory cells to solve the vanishing gradient problem in its hidden layer which consists of memory cells and three gates, (output gate, forget gate, and input gate). The forget gate controls data memorizing or removing the information from past time steps. The input gate adds information to the memory block. The output gate outputs information to the next memory cell.

Yang, *et al.* [4] proposed an ANN model to perform the identification and counting of components within a DW in real-time. The proposed model performs classification and counting. The model is tested by employing a mini dataset where the initial result shows that the proposed model has achieved an accuracy rate of 80% in classification and a 97% accuracy rate in counting. The development of the model is associated with the challenge of not being able to identify the component in a larger dataset which will be

required using an LSTM neural network that has a forget gate that can be able to tolerate a larger data set.

Li, *et al.* [19] also developed an MLM to predict the Tunnel Boring Machine (TBM) performance under uncertain geological conditions using LSTM. Its accuracy was evaluated and shows superior performance than the classical model in which only instance data can be used. In the same vein Song [20] in their study that tries to resolve a critical issue in oil explorations proposes an LSTM neural network-based model for production prediction of fractured horizontal wells in a volcanic reservoir, after evaluating the performance of the approach was compared with traditional neural networks and time-series forecasting approaches which show that the proposed LSTM model outperforms other approaches. But Zhang, *et al.* [21] in trying to improve the prediction performance of TBM developed a random forest algorithm-based model for safe and efficient tunneling. Results show the random forest algorithm generates better prediction results than the LSTM and the big operational data-based method predicts TBM performance with high accuracy as compared to the framework [19].

Ta, *et al.* [22] develop a stock prediction model that is challenged with factors such as uncertainty and instability. The LSTM network-based model was developed to predict stock movement based on historical data and results showed that the LSTM prediction model works efficiently by obtaining high accuracy results from stock prediction and also constructed portfolios based on the LSTM prediction model outperformed other constructed portfolios-based prediction models such as linear regression and support vector machine. Similarly, Gurav and Kotrappa [23] built a Stock forecast model that combines LSTM and Log Bilinear considering external factors to analyze the impact of COVID-19 on stock market performance using labeled records where performance improvement in terms of Root Mean Square Error and accuracy score in prediction have been identified after evaluation.

In addition, LSTM proved to be a more robust prediction mechanism considering a study by Chen, *et al.* [24] which uses LSTM to predict nonlinear wing speed. The model having proof to be accurate and reliable, its result was then compared with six different prediction models and got the highest prediction accuracy based on the prediction evaluation index. Similarly, Liu [25] conducted a study on slope stability prediction on the basis that others once did not provide them with the right result because they failed to consider some factors. LSTM was then introduced which takes in slope height and cohesion among other factors to make the prediction. LSTM proved to be better in slope stability prediction than Support Vector Machine (SVM) and Convolutionary Neural Network (CNN).

4. Methodology

4.1 Data Used for the Study

This study uses secondary data. 140,000 crime-related datasets collected from the United Kingdom (UK) police station database report for north-wales and north-Yorkshire respectively. The dataset is stopped and

searched data by the police personnel. The dataset consists of data features such as longitude, latitude, crime type, officer-defined ethnicity, legislation, object of search, and outcome, among others.

4.2 Data Preparation

Data preparation is key for every neural network system. In this research, the following steps in preparing the data were followed: Removing punctuation, Lemmatization, stemming (using Porter Stemmer algorithm), Removing stop-words, Tokenization, and finally Deduplication which involves removing redundant entries from our datasets. Each dataset is partitioned into two separate groups, the splitting of the datasets was done using the Train-Test Split Procedure in Scikit-Learn, which splits the data based on a preset strategy.

4.3 Model Training and Training Algorithm

The training algorithm used in this study is the Gradient Descent optimization algorithm. The network consists of three gates, (output gate, forget gate, and input gate). The forget gate controls data memorizing or removing information from past steps. The input gate adds information to the memory block. The output gate outputs information to the next memory cell. The architecture of the model is defined such that each neuron from one layer is connected to each neuron from the following layer.

5. Result and Discussion

The objective of this study is to propose a data mining management framework for prediction with the vanishing gradients handler using a Long Short-Term Memory Artificial Neural Network. The LSTM data mining model is used to explore the hidden pattern and or insight within the dataset and result will then be visualized using any available visual tools that will suit the context as can be seen in the Figure 2.

As can be seen from Figure 2, the framework data warehouse collect data from various data sources and used the ETL mechanism in extracting the data from various sources, then transform to suit the schema of the warehouse and then loaded in to the warehouse. The data in the warehouse will then be passed in to the LSTM mechanism for prediction and the result will then be visualize using any available tool.

The framework was implemented as an intelligent system using Python programming language in a collab environment (Google Collaboratory environment) that gives ease in accessing most of the required libraries for intelligent system development without the need for a lot of resources. The implemented system was able to take in all the data sets as input, process them, and provide us with prediction results.

The result is been plotted in a bar graph showing the relationship between the index year and the predicted year. The framework was able to predict crime rate based on the index year dataset monthly and yearly with an accuracy level of 92.9% and precision level of 76%

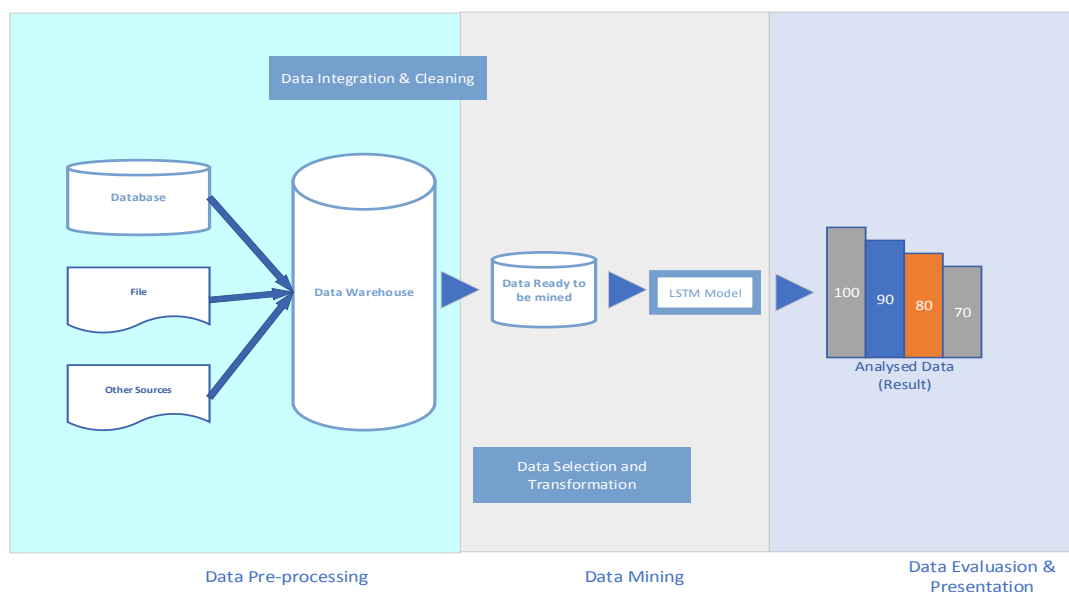


Figure 2: Data Mining Framework using LSTM based [7]

6. Conclusions

The introduction of LSTM into the data mining process gives a better and improved prediction result using crime datasets because of the forget gate in the LSTM structure unlike other mining algorithms. The study was able to evaluate the model performance in terms of accuracy, precision, and scalability.

Recommendations

The following recommendation were identified:

- The study recommends full implementation of the framework as an intelligent system in any crime related organization to make prediction and the result will serve as an early warning indicator to prevent crime occurrence
- Considering the speed, the study recommends the full implementation of the framework in a full-scale Python environment to be able to evaluate the speed because Google Colab has pre-defined functions or libraries that serve as services to the user.

References

- Reddy, R. P., Mandakini, C. and Radhika, C. A., (2020). A review on data mining techniques and challenges in medical field. *International Journal of Engineering Research and Technology*, 9(08), 329–333. <https://doi.org/10.17577/ijertv9is080143>
- Mahmud, N., Zinnah, K. I., Rahman, Y. A. and Ahmed, N., (2017). CRIMECAST: A crime prediction and strategy direction service. *19th International Conference on Computer and Information Technology, ICCIT 2016*, 414–418. <https://doi.org/10.1109/ICCITECHN.2016.7860234>
- Santoso, L. and Yulia, W., (2017). Data warehouse with big data technology for higher education. *Procedia Computer Science*, 124, 93–99. <https://doi.org/10.1016/j.procs.2017.12.134>
- Yang, J. X., Li, D. and Rasul, M., (2021). A conceptual artificial neural network model in warehouse receiving management. *International Journal of Machine Learning and Computing*, 11(2), 130–136. <https://doi.org/10.18178/ijmlc.2021.11.2.1025>
- Qayyum, S. and Dar, H., (2018). A survey of data mining techniques for analyzing crime patterns. *University of Sindh Journal of Information and Communication Technology (USJICT)*, 2(1), 1–6. <https://doi.org/10.1109/ACDT.2016.7437655>
- Prakash, G. H. and Rangdale, S. P., (2017). ETL data conversion: extraction, transformation and loading data conversion. *International Journal Of Engineering And Computer Science*, 6(10), 2319–7242. <https://doi.org/10.18535/ijecs/v6i10.03>
- Rajab, A. R. and Masoud, A. R., (2021). The concept of data mining and knowledge extraction techniques. *Qubahan Academic Journal*, 1(2), 17–20. <https://doi.org/10.48161/qaj.v1n2a43>
- Soni, A., Kumar, V., Kaur, R. and Hemavathi, D., (2018). Predicting student performance using data mining techniques. *International Journal of Pure and Applied Mathematics*, 119(12), 221–227.
- Anusha, C. and Jyothi, B. S., (2020). Design of a data warehouse for medical information system using data mining techniques. *Journal of Engineering Sciences*, 11(6), 308–313.
- Ali, A. D., (2013). Security and Economic Development in Nigeria Since 1960. *Kuwait Chapter of Arabian Journal of Business and Management Review*, 2(6), 1–7. <https://doi.org/10.12816/0001202>
- Tyagi, D. and Sharma, D. S., (2020). An approach to crime data analysis: a systematic review. *International Journal of Engineering Technologies and Management Research*, 5(2), 67–74. <https://doi.org/10.29121/ijetmr.v5.i2.2018.615>

- [12]. Sayed, E., Abdelsamee, A. A. and Ghazaly, N. M., (2020). Classification model using neural network for centrifugal pump fault detection. *International Journal of Systems Signal Control and Engineering Application*, 13(4), 120–126.
- [13]. Devakunchari, R., Bhowmick, S., Bhutada, S. P. and Shishodia, Y., (2019). Analysis of crimes against women in India using regression. *International Journal of Engineering and Advanced Technology*, 8(2), 124–127.
- [14]. Saltos, G. and Cocea, M., (2017). An exploration of crime prediction using data mining on open data. *International Journal of Information Technology and Decision Making*, 16(5), 1155–1181. <https://doi.org/10.1142/S0219622017500250>
- [15]. Fatlawi, H. K., (2016). *Crime data analysis using data mining techniques to improve crimes prevention procedures*. Iraqi Conference for Information Technology (ICIT) First, January 2016, Kufa Iraq, 122-131.
- [16]. Mowafy, M., and Rezk, A. (2018). Building unstructured crime data prediction model (practical approach). *International Journal of Computer Application*, 4(8), 1–7.
- [17]. Kang, H. W., and Kang, H. B. (2017). Prediction of crime occurrence from multimodal data using deep learning. *PLoS ONE*, 12(4), 1–19. <https://doi.org/10.1371/journal.pone.0176244>
- [18]. Sivarajah, U., Kamal, M. M., Irani, Z. and Weerakkody, V., (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- [19]. Li, J., Li, P., Guo, D., Li, X. and Chen, Z., (2021). Advanced prediction of tunnel boring machine performance based on big data. *Geoscience Frontiers*, 12(1), 331–338. <https://doi.org/10.1016/j.gsf.2020.02.011>
- [20]. Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L. and Cheng, Z., (2020). Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *Journal of Petroleum Science and Engineering*, 186, 106682. <https://doi.org/10.1016/j.petrol.2019.106682>
- [21]. Zhang, Q., Liu, Z., and Tan, J. (2020). *Predicting the performance of tunnel boring machines using big operational data*. Proceedings - 2020 IEEE 6th International Conference on Big Data Computing Service and Applications, BigDataService 2020, Hangzhou, China 179–182. <https://doi.org/10.1109/BigDataService49289.2020.00035>
- [22]. Ta, V. D., Liu, C. M., and Tadesse, D. A. (2020). Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. *Applied Sciences (Switzerland)*, 10(2), 1-20. <https://doi.org/10.3390/app10020437>
- [23]. Gurav, U. and Kotrappa, D. S., (2020). Impact of COVID 19 on stock market performance using efficient and Predictive LBL-LSTM based mathematical model, *International Journal on Emerging Technologies*, 11(4), 108–115.
- [24]. Chen, G., Tang, B., Zeng, X., Zhou, P., Kang, P. and Long, H., (2022). Short-term wind speed forecasting based on long short-term memory and improved BP neural network. *International Journal of Electrical Power and Energy Systems*, 134(March 2022), 107365. <https://doi.org/10.1016/j.ijepes.2021.107365>
- [25]. Liu, J., Huang, Z., and Ma, Z. (2023). *Guiding Mathematical Reasoning via Mastering Commonsense Formula Knowledge*. Proceedings of Make Sure to Enter the Correct Conference Title from Your Rights Confirmation Email (Conference Acronym 'XX'), Hefei, China, 1(1), 1477–1488. <https://doi.org/10.1145/3580305.3599375>
- [26]. Vishnupriya, V. and Valarmathi, M. (2017). An effective data mining techniques for analyzing crime patterns. *Journal of Computer Engineering*, 1, 26–30.