Penerbit UTHM

© Universiti Tun Hussein Onn Malaysia Publisher's Office

JSCDM

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/jscdm

Journal of Soft Computing and Data Mining

A Comparison of Sensitive Information Detection Framework using LSTM and RNN Techniques

Norfakhira Iman Mohamad Roslan[1], Cik Feresa Mohd Foozy[1]*

[1]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, Johor, MALAYSIA

*Corresponding Author

DOI: https://doi.org/10.30880/jscdm.2022.03.02.010
Received 12 July 2022; Accepted 20 Ocotber 2022; Available online 01 November 2022

Abstract: Sensitive information is meant to be stored securely to avoid any data breach. Hence this research emphasized on whether or not Long Short Term Memory and Recurrent Neural Network is suitable for sensitive information detection framework since performance analysis on this method are not common. In this research, the objectives are to design Sensitive Information Detection Framework using Deep Learning techniques, to detect sensitive information using LSTM and RNN techniques and to test and validate the sensitive information detection framework performance. This research uses raw data set given from Book Hack Enterprise company and by using Weka Explorer software tool to go through 5 phases in the proposed framework, which are Dataset, Preprocessing, Feature Extraction, Classification Algorithm and Performance Evaluation. This research evaluates the model performance based on Accuracy, Precision, Recall and F1-score.

Keywords: Deep Learning, LSTM, RNN, Sensitive Information Detection Framework

1. Introduction

Data privacy has the attention from the public and legislators as it is mandatory when storing sensitive information. Sensitive information includes biometric data, financial details and an individual identity. Ignorance in securing data is self-destructive with the advancing technology these days, because everything is stored online as well as transactions made online. Negligence in data security can result in huge losses for an organization as well as to an individual [1]. Statistics have shown in the past year 2020, there were 28,000,000 cases reported on data breaches out of more than 20 billion breaches cases. The top 3 attackers that were recorded behind the breaches are external actors, organized criminal groups and internal actors with 70%, 55% and 30% involvement respectively [2]. This research is to identify sensitive information detection framework using LSTM and RNN techniques. The problem statements related to this research include the difficulty in identifying sensitive and insensitive information [3], whether or not LSTM and RNN is suitable for a sensitive information detection framework [4] and last but not least, the lack of study in LSTM and RNN in a sensitive information detection framework [5].

The importance of this research is to be able to profile suitable features when filtering out any sensitive information that might affect the privacy and security of documents. Hence the objectives of this research are (1) to design a Sensitive Information Detection framework using Deep Learning techniques, (2) to detect sensitive and insensitive information by using LSTM and RNN techniques, and (3) to measure the detection model performance based on Accuracy, Precision, Recall and F1-Score.

The expected outcome of this research is to be able to apply a Deep Learning approach to detect sensitive and insensitive information framework and to get better results in the previously mentioned performance metrics by using LSTM and RNN techniques. The dataset used in this research is given from Book Hack Enterprise company and is later cleaned manually. In conclusion, mishandling of sensitive information can do a lot of damage to an individual and to an

*Corresponding author: feresa@uthm.edu.my

organization, therefore a sensitive information detection is useful to filter data by applying LSTM and RNN techniques where it can create new features.

2. Related Work

In this section, it will focus on the sensitive information definition, then the Deep Learning components and lastly the comparison between the existing Deep Learning techniques.

2.1 Sensitive Information Definition

Sensitive information is closely related to the Personal Data Protection Act 2010 (PDPA). It was created based on organizations involvement in digital marketing, because the purpose is to manage the confidentiality process of data secured according to PDPA requirements. Moreover, the General Data Protection Regulation (GDPR) was established to differentiate sensitive and insensitive information directly. Although GDPR encourages usage of pseudonymous information to reduce the risk of data breaches, it is also important to classify and specify data [6].

Examples of sensitive information in PDPA include political opinions and health conditions. Malaysia's PDPA comprises the following principles:

- General Principle
- Notice and Choice Principle
- Disclosure Principle
- Security Principle
- Retention Principle
- Data Integrity Principle
- Access Principle

There will be a huge sum of fine to anyone that breaches the principles mentioned above [7]. General principle is when consent is obtained for the data you collected for processing, followed by notice and choice principle, is when a written notice is provided that includes a description of data processing, purpose of data collection, the rights individuals have on their data and more. Next, disclosure principle is an express consent of what stated in earlier principles, that goes together with security principle where organizations are held accountable for keeping the data safe. Then, the retention principle highlights the period of data storing is to be within the necessary amount. Data integrity principle is like an oath to take responsibility to ensure data is well-kept and lastly, access principle is allowing individuals to access or correct with regard to the collected data.

2.2 Deep Learning Phase

There are several components in Deep Learning that need to be followed in order to get the desired result, which are Dataset, Preprocessing, Features and Classification and Performance Evaluation.

2.2.1 Dataset

At this stage, raw datasets are obtained from several sources such as Kaggle, Datahub.io and many more. Dataset are prepared by researchers according to the research topic to ensure results with high reliability with the right dataset. Work by [9] mentioned, KDD CUP 99 dataset were used in the research that has redundant features while, work by [10] stated, Enron dataset were used despite the difficulty to characterize.

2.2.2 Preprocessing

There are many techniques that can be used in this phase since every dataset and every model has its own functionality, hence the different dataset may require different techniques. For instance, work by [8] used 3 different techniques which are Word Segmentation, Character Digitization and Vector Construction while work by [10] only used Normalization technique.

2.2.3 Features and Classification

Features extraction is to filter or create new features that are suitable for the model to receive as an input. In order for an accurate classification, features must be relevant as well. For example, work by [8], uses the LSTM algorithm with an additional feature of CRF on top of it. While work by [9] mentioned the features used such as "Prepay transactions" and "Letters of credit", in the FS-WOA-DNN algorithm research.

2.2.4 Performance Evaluation

Parameters are needed to ensure experimental results are known to the public for it to be analyzed and learned from the research. Performance evaluation can be seen through parameters used for detection values. Research by [8] uses Accuracy, Return Efficiency and F1 Index where its algorithm showed improved accuracy and efficiency rate. Work by [9] measures using Accuracy, Specificity, Sensitivity, Error and False Positive Rate which proposed algorithm has high returns.

2.3 Comparison Between Existing Deep Learning Techniques

Table 1 shows the comparison between the previous existing methods in the sensitive information detection model.

Table 1 - Comparison between existing methods

Reference	LSTM-CRF [8]	FS-WOA–DNN [9]	RNN [10]	Proposed LSTM and RNN
Dataset	Not mentioned	KDD CUP 99 [11]	Corpus Dataset [12]	Book Hack Enterprise Company
Preprocessing stage	Word Segmentation, Character Digitization, Vector Construction	Paraphrasing, Sentiment Analysis, Image Sentence Ranking,	Normalization	- Tag and label sentences - Identify sensitive and insensitive information
Feature sensitive information	Not mentioned	Example: Prepay transactions, Letters of credit	Not mentioned	Example: Tender, procurement, audit
Classification in Deep Learning	RNN	RNN	DNN	LSTM and RNN
Detection evaluation	-Accuracy -Return efficiency -F1 index	Not mentioned	-Accuracy -Specificity -Sensitivity -Error -False Positive Rate	-Accuracy -Recall -Precision -F1-score

3. Methodology

There are several phases in this research, and it includes Dataset, Preprocessing, Feature Extraction and Classification Algorithm and Performance Evaluation.

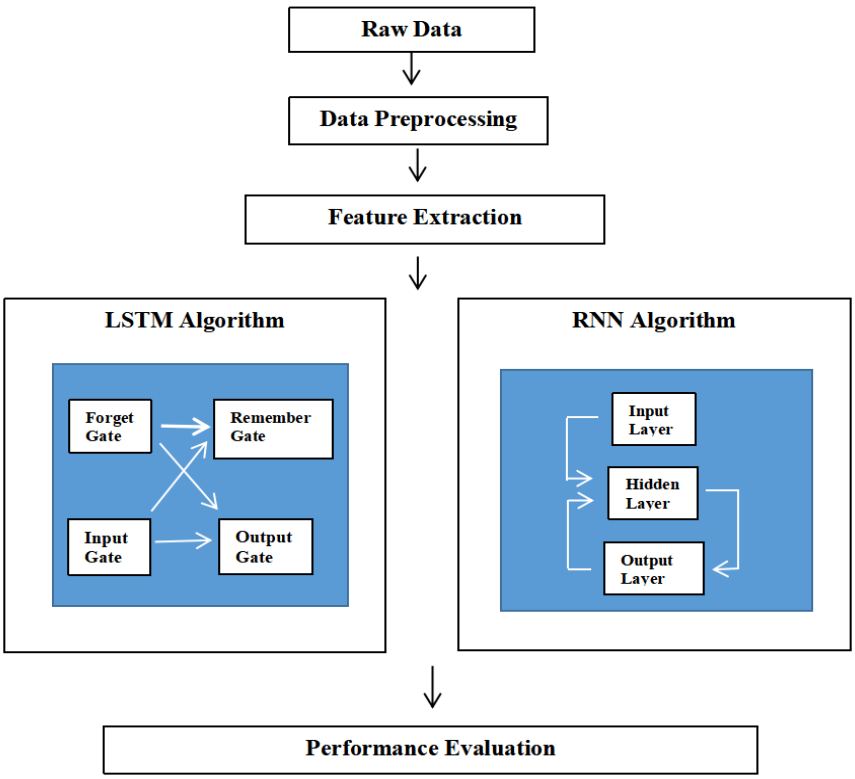

Fig. 1 - Sensitive information detection framework

3.1 Dataset

In this research, a dataset was provided by Book Hack Enterprise company. The raw dataset is then preprocessed to clean it. The dataset consists of lines with 2 columns which are the sentences and the classification of sensitive and insensitive information (refer to APPENDIX A). The dataset is based on Resume, Construction and Medical because the sentences are from project documentations, medical findings and job resumes hence the classification of sensitive and insensitive information.

3.2 Preprocessing

Data preprocessing is where the raw dataset is cleaned and transformed into structured data usable in an algorithm. In this research, data preprocessing has several stages which are Data Integration, Data Validation and Data Transformation as shown in Figure 2.

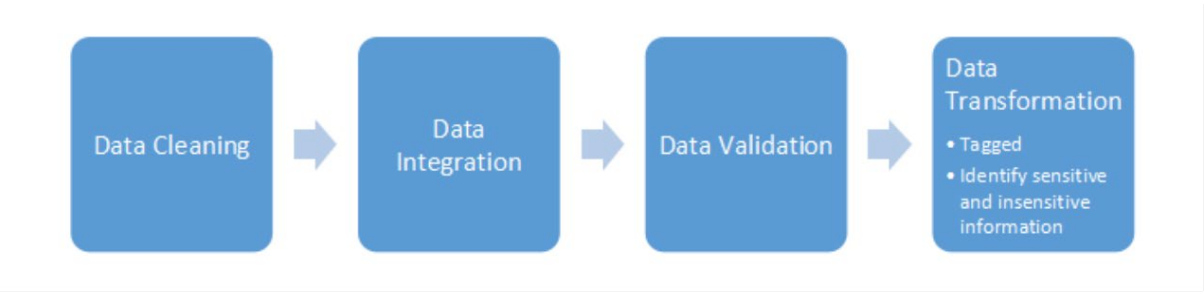

Fig. 2 - Data preprocessing flowchart

Data Integration is to combine and unify the data while, Data Validation is to ensure data is complete and accurate then, Data Transformation is to refine the dataset. The dataset is cleaned by labeling whether the sentence is sensitive or insensitive based on features of classification. Hence cleaning the dataset requires a few runs through to ensure sentences are understood leading to a better dataset, as well as better outputs when classification of sensitive and insensitive information is made clearly.

3.3 Feature Extraction

In this research, features are extracted manually because the identification of the features are extracted based on the dataset background for accurate features and classification of dataset. After manually feature extraction is performed, the sensitive keywords are then used to ensure high accuracy. One of the ways is by using the CTRL+F feature to find the sensitive keywords within the dataset such as audit and tenders are sensitive keywords because of the integrity and confidentiality of the data.

Table 2 - Samples of sensitive keywords [16], [17]

Sensitive Keywords	Description
Tenders	It is a formal documented offer involving money to a client. Most of the information contained in tender responses should be kept confidential.
Procurement	It is a highly competitive list involving great care and attention before proceeding with the project that needs to specifically and efficiently ensure the confidentiality of these documents according to.
Responsible	The party responsible for creating a contract can detail any information they wish to make confidential.
Cost	Cost is closely related with cash flow when involved with project documentation and contracts.
Quotation	A formal statement of an estimated cost for a project that is agreed upon holding such information in strict confidence.
Audit	Auditing must be conducted within a framework of complete trust and strictly confidential.
Report	There are some reports that cannot be made public due to privacy of who may be involved.
Contractual	Contractual confidentiality obligations are fundamental and necessary to help protect the parties that disclose information in these situations.
Allocation	Allocation is usually done in consultation with the borrower, who are interested in relationship banks receiving the largest allocations.

3.4 Classification Algorithm

This research uses LSTM and RNN algorithms to classify the sensitive and insensitive information in the dataset. LSTM is a classification to enhance precision of filtering by dealing with long term dependency problems. The RNN algorithm is capable of remembering the characteristics of previous input and output however, immediate output may not be accurate for the next prediction and the network depends on previous input and output [13]. Hence why, RNN is unrolled because the previous input and hidden states are used to compute the gradients to the final output of RNN. RNN can deal with sequences of variable length by defining a recurrence relation over time steps typically following the following formula:

$$S_k = f(S_{k-1} \cdot W_{rec} + X_k \cdot W_x) \qquad Eq.1 \tag{1}$$

where:

i. S_k is the state at time k.
ii. X_k is the exogenous input at time k.
iii. $Wrec = W_x = 1$ is the weight parameters.

LSTM on the other hand is capable of learning long-term dependencies. LSTM has 3 step gates which are Forget gate, Input gate and Output gate. Since its main component is the memory cell, the states can maintain for a long time for regulation of information flowing in and out of the memory [14]. Forget gate controls information that can be discarded from the memory, and it works on its equation as follows:

$$f_t = \sigma\,(w_f[h_{t-1}, x_t] + b_f) \qquad Eq.2 \tag{2}$$

where:

i. ft is the forget gate.
ii. it is the input gate.
iii. ot is the output gate.
iv. σ is the sigmoid function.
v. wx is the weight for respective gate(x) neurons.
vi. ht-1 is the output of previous LSTM block (timestamp t-1).

vii. xt is the input at the current timestamp.
viii. bf is the biases for respective gates(x).

Then, input gate controls new information that is added to cell state on top of current input as equation follows:

$$i_t = \sigma\left(w_i[h_{t-1}, x_t] + b_i\right) \qquad Eq.3 \tag{3}$$

Lastly, Output gate controls what can be output from the memory such following equation:

$$o_t = \sigma\left(w_o[h_{t-1}, x_t] + b_o\right) \qquad Eq.4 \tag{4}$$

With such LSTM features, the classification of sensitive and insensitive information will have higher accuracy because of the Input gate that is receiving new information from current input on sensitive keywords.

3.5 Performance Metrics

Parameters used in this research to measure the performance value are Accuracy, Recall, Precision and F1-score [18]. In *Eq.* 5 until *Eq.* 8, *TP* means the number of accurately recognized sentences containing sensitive information, while *TN* means number of accurately recognized sentences that do not contain sensitive information. *FN* means the number of incorrectly recognized sentences containing sensitive information as negative. *FP* means the number of incorrect recognized sentences that do not contain sensitive information as positive.

Accuracy indicates the proportion of samples that is correctly classified as sensitive information.

$$\text{Accuracy} = \left(\frac{TP + TN}{TP + TN + FP + FN}\right) \qquad Eq.5 \tag{5}$$

Recall denotes the ratio of positive samples that is correctly classified as sensitive information.

$$\text{Recall} = \left(\frac{TP}{TP + FN}\right) \qquad Eq.6 \tag{6}$$

Precision indicates the mean proportion of samples that is classified as sensitive information among all samples expected to contain sensitive information.

$$\text{Precision} = \left(\frac{TP}{TP + FP}\right) \qquad Eq.7 \tag{7}$$

F1-Index is F-measure where it is the weighted average of precision and recall. It helps to evaluate the result better so the higher the F1, the better the performance of the sensitive information detection model.

$$\text{F1} = \left(\frac{2\,PR}{P + R}\right) \qquad Eq.8 \tag{8}$$

where:
i. P represents Precision.
ii. R represents Recall.

3.6 Software and Hardware Requirements

To ensure experiments can run smoothly, hardware is an important tool to carry out the experiments. The hardware specification used is stated in Table 3.

Table 3 - Hardware specifications

Hardware	Description	
Acer Swift 3	Windows Edition	System

	Windows 10 Version 20H2 for x64-based Systems	Processor: AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz Installed RAM: 4.00 GB System Type: 64-bit operating system, x64-based processor

For software requirements, this research uses Weka Explorer software tool that is provided from University of Waikato, New Zealand. It has the latest added package called Dl4jMlpClassifier consists of Deep Learning layers you can manually add to your algorithm. This software can implement LSTM and RNN algorithms and results can be shown in a simple manner such as a graph or a table.

4. Results and Discussion

This section describes the findings of this research based on the parameters mentioned previously. Experiments are done using a dataset provided by Book Hack Enterprise company and Weka Explorer software tool.

4.1 Experimental Setup

Firstly, raw dataset is preprocessed by omitting duplicates and combining into one view to make it a structured dataset. One of the functions is to omit duplicates in the dataset because of copy paste error and human error. Based on Figure 3 and Figure 4, the number of lines (instances) had changed because of the omitted duplicates. It went from 180 instances to 175 instances after omitting the duplicates.

Fig. 3 - Before omit duplicates

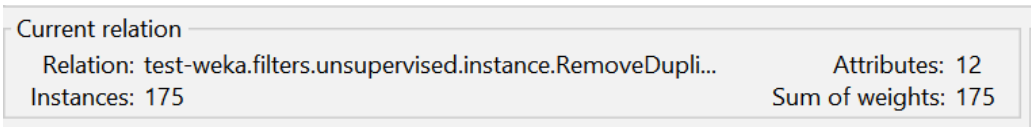

Fig. 4 - After omit duplicates

Next, the “Count” column in Figure 5 shows the number of data that has been classified into its own class of sensitive and insensitive information. Based on the column results, it shows that sensitive information has 90 data while insensitive information has 85 data.

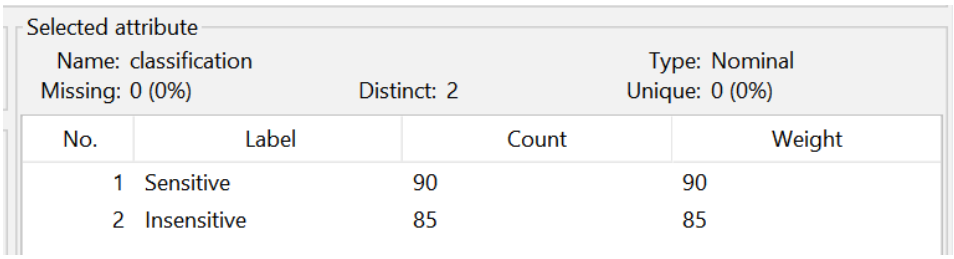

Fig. 5 - Classification of dataset

Then, the dataset is tested in LSTM and RNN algorithms. In LSTM algorithm, there are LSTM layer and Dense Layer applied to get the desired result as shown in Figure 6. So, altogether there is 3 layers in the LSTM algorithm. On the other hand, RNN algorithm had 4 nodes of hidden layer before producing the output as shown in Figure 7. There are only 4 nodes due to the small amount of dataset and for short runtime.

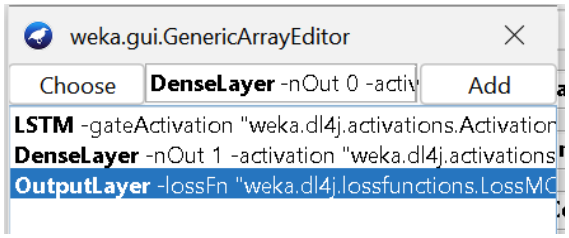

Fig. 6 - Layers implemented in LSTM

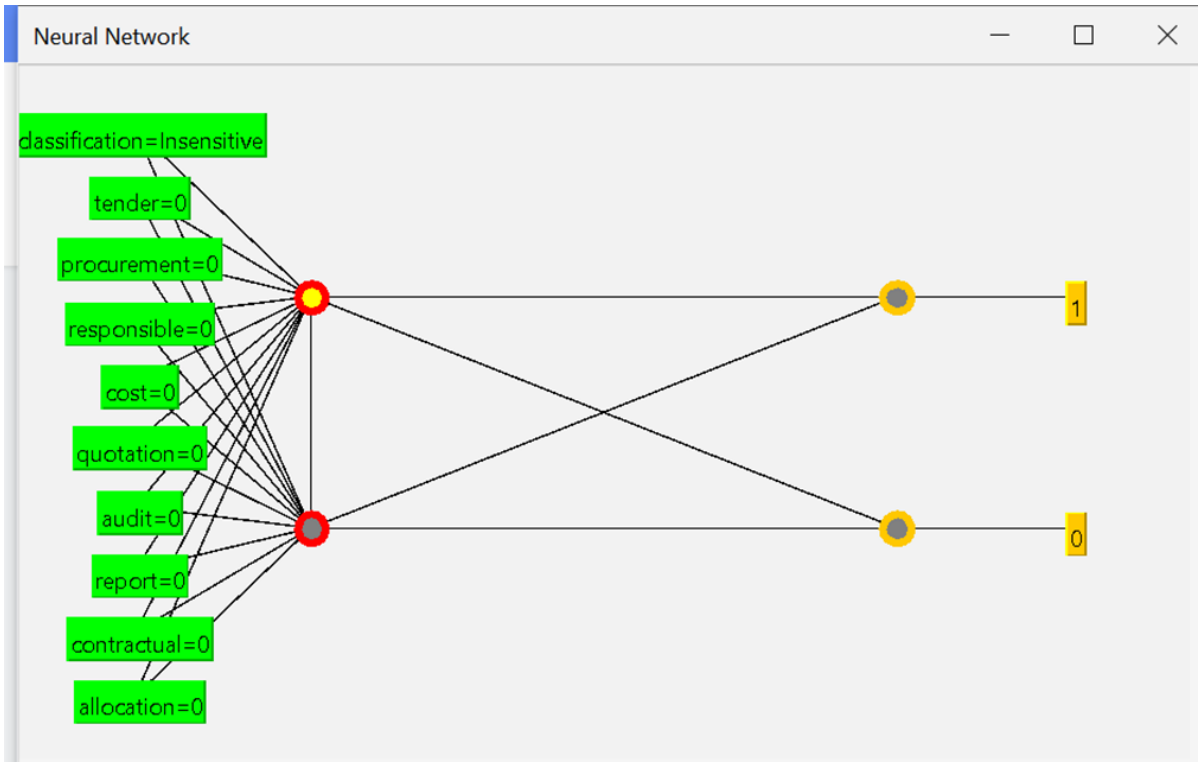

Fig. 7 - Structure of the RNN algorithm

4.2 Comparison of LSTM and RNN

In this subtopic, the results obtained from the experiments will be presented in a bar chart for better viewing of comparison between LSTM and RNN algorithms. There are 4 performance metrics result, which are Accuracy, Recall, Precision and F1-score.

Firstly, is the Accuracy result shown in Figure 8, where LSTM has higher accuracy of 97.14% compared to RNN accuracy of 91.42%. This means that LSTM had managed to identify more sensitive keywords that were classified.

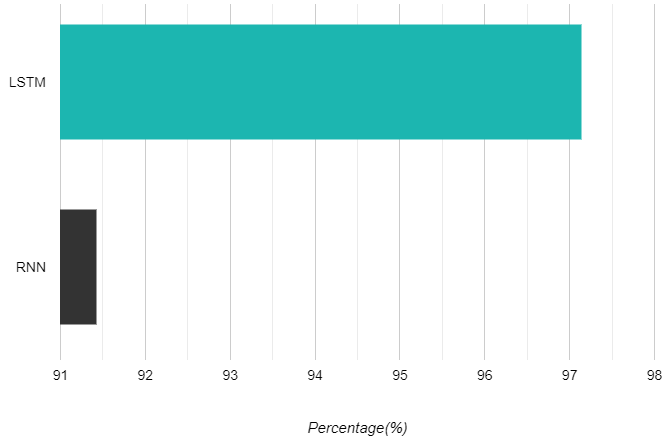

Fig. 8 - Accuracy result

Secondly, is the recall result in figure 9 which LSTM obtained a 0.971 value, while RNN obtained a 0.914 value. The perfect value for recall is 1. Hence shows that LSTM also achieved a higher result than RNN.

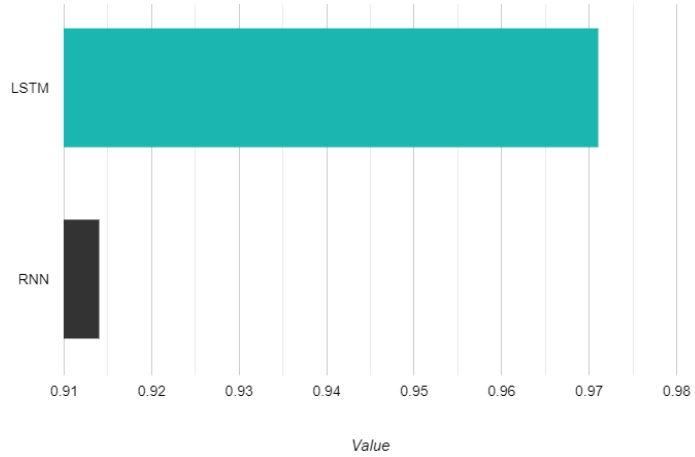

Fig. 9 - Recall result

Thirdly, is the precision result, which means the sensitive cases are correctly sorted as sensitive samples. Figure 10 illustrates that LSTM received a 0.971 and RNN received a 0.914 precision values. Precision perfect value is also 1.

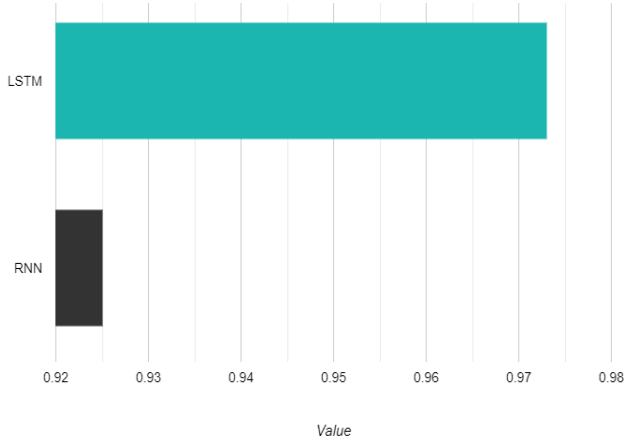

Fig. 10 - Precision result

Lastly, the F1-score which means the higher the F1-score is, the better the performance of the model. For F1-score, LSTM and RNN had reached 0.973 and 0.925 respectively as depicted in Figure 11. This means LSTM has higher F1-score as compared to RNN.

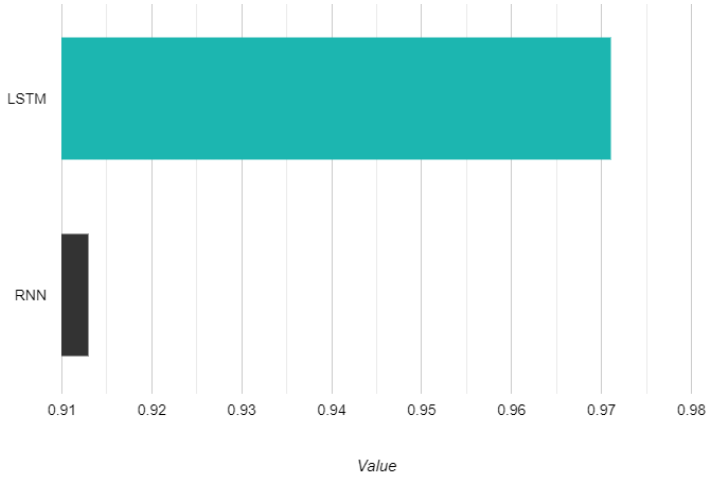

Fig. 11 - F1-score result

Ultimately, Table 4 will tabulate all the results from the experiment in comparison. The table will consist of the performance metrics that had been used to measure the effectiveness of LSTM and RNN framework in sensitive and insensitive information detection. The performance metrics are Accuracy, Recall, Precision and F1-Score.

Table 4 - Results in comparison

Performance Metrics	Accuracy	Recall	Precision	F1-Score
LSTM	97.14%	0.971	0.973	0.971
RNN	91.42%	0.914	0.925	0.913

5. Conclusion

The sensitive data in documentation are meant to keep confidential as it consists of personal information such as bank details, health conditions and ID number. The research objectives are to design a detection model for sensitive information using deep learning techniques, to detect sensitive information based on data set given by Book Hack Enterprise company using LSTM and RNN techniques and to measure the performance of the model with Accuracy, Precision, Recall and F1-Score. Based on the early findings in Weka Explorer tool, it seems that the dataset consists of duplicates or human error where the lines are now 175 from 180 lines initially after omitting duplicates function. This research managed to extract 10 features from the dataset and classify the sentences into sensitive and insensitive information based on it. The LSTM algorithm consist of LSTM layer and Dense Layer while RNN algorithm consist of 4 nodes of hidden layer before the output layer. After the dataset had been tested for both algorithms, it is found that LSTM has higher overall results as compared to RNN. However, both algorithms results are well over 90%, in every performance metrics. This research has achieved its objectives after knowing that both of the algorithms are suitable for sensitive information detection. There are also few suggestions that can be made for future works, which are, using a larger dataset to test the algorithms for higher rate of training the algorithm and higher accuracy results as well as implementing one topic background for similar features to be extracted and identified.

Acknowledgment

The authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia for its support and encouragement throughout the process of conducting this research.

Appendix A: Sample Dataset

Sentences	Classification
This case study details my involvement as a Senior Contract Researcher in preparation of a tender for the Building Affairs Department (BAD) in PWA for the Design and Build for Construction of Site 23 Al Rayyan Bus Depot Project (the Project).	Sensitive
Key issue one concerned the delay in the tender closing date due to the transfer of design liability during the tender stage, impacting the completion date of the Project.	Sensitive
Key issue two concerned the establishment of a mechanism that enables tender errors to be rectified during post-contract.	Sensitive
This case study discusses my contributions in solving these key issues during the tender stage.	Sensitive
The aim of the tender was twofold. First to provide a world-class depot function and second to develop depot facilities strategically positioned to maximize functionality and accessibility.	Sensitive
Zone 1 was tendered with the design portion (the design was to be undertaken by the Contractor) whereas Zone 2 was tendered out without the design portion (the design was undertaken by PWA).	Sensitive
As a Senior Contract Researcher with PWA in the ESD, my involvement is in preparing and reviewing the tender documents (TDs) at the tender stage and reporting to the Head Section of TSS. Simultaneously providing professional advice in terms of procurement strategy and contract form selection during tender preparation stage.	Sensitive
Liaison with other PWA Departments particularly with BAD (known as RD) is part of the tender preparation.	Sensitive
As general procedure, my department (ESD) would receive the 'Request-to-Tender? through the system from RD. Once a request is received, ESD will start preparing the tender within the timelines given. The request carries details of the project such as the scope of work, estimated value, specifications, drawings, and tender plan which contains information pertaining date of project award, project start, and project completion whereas the Tender Estimate is prepared by ESD Planning and Cost Control Section.	Sensitive
Since my tenure with PWA in 2018, my tasks have exposed me to the preparation of twenty (20) Tenders.	Sensitive
Despite guidelines to ensure tender quality (refer Appendix B), the percentage of delay in the tender closure and award phase is high.	Sensitive
Such delays impact the commencement and completion date of a project. This in turn results in delay and cost overrun.	Insensitive
My belief that the clients' objectives as a principal objective includes time, cost, and quality. Any cause of delay or increase in project cost has to be mitigated earliest possible. This part explains how key issues transpired in the Project.	Insensitive
Considering the design was proposed near the closing date, the scope of works and tender price would be amended during post-contract.	Insensitive
Higher tender price due to contractual elements in D&B (different in risk profile) i.e. design associated clauses, testing, and latent defects. This refers to the provisions stated in the D&B condition contract (as shown in Appendix E).	Sensitive
Possibility of prolonged post-contract due to request by Contractor to reprice to meet their expectation.	Insensitive
Contractor claims for design charges.	Insensitive
Based on my experience and private studies, procurement process and contract management are crucial to a successful completion of a project.	Insensitive
The initial contract must specify relevant aspects of the project work. A long chain of contractual matter or disputes may result from such as arbitrations and negotiations due to variations i.e. the design provided by the client during post-contract might not match with contractors' price.	Sensitive
This could lead to poor quality and unacceptable final results. Furthermore, this affects the payment system and contractors cash flow which eventually leads to delayed completion and cost overrun.	Sensitive

References

[1] Carr, R. (2020, March 30). What is sensitive information? Retrieved December 24, 2021, from https://www.zettaset.com/blog/what-is-sensitive-information/

[2] Chester, D. (2021, September 09). Data breaches 2021 - statistics, Tips, guides. Retrieved December 24, 2021, from https://cooltechzone.com/threats/malware-removal/biggest-data-breaches

[3] Spirion, LLC. (2020, August). Data Classification (Data Management): A complete overview. Retrieved December 24, 2021, from https://www.spirion.com/data-classification/#phase-2

[4] Kumar, R. (2021, July 31). Recurrent neural network - an overview. Retrieved December 24, 2021, from https://medium.com/nerd-for-tech/recurrent-neural-network-an-overview-1128ffc34ce7

[5] Y. Lin, G. Xu, G. Xu, Y. Chen and D. Sun, "Sensitive Information Detection Based on Convolution Neural Network and Bi-Directional LSTM," 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2020, pp. 1614-1621, doi: 10.1109/TrustCom50675.2020.00223.

[6] Tunggal, A. T. (2021, September 14). What is sensitive data? Retrieved December 24, 2021, from https://www.upguard.com/blog/sensitive-data

[7] Shahari, F. (2018, May 31). Personal Data Protection act PDPA 2010: Malaysia. Retrieved December 24, 2021, from https://cloudrock.asia/blog/pdpa-2010-malaysia/

[8] Wu, L., & Pan, M. (2021). English grammar detection based on LSTM-CRF Machine Learning Model. Computational Intelligence and Neuroscience, 2021, 1-10. doi:10.1155/2021/8545686

[9] Agarwal, A., Khari, M., & Singh, R. (2021). Detection of DDOS attack using deep learning model in cloud storage application. Wireless Personal Communications. doi:10.1007/s11277-021-08271-z

[10] Neerbeky, J., Assentz, I., & Dolog, P. (2017). Taboo: Detecting unstructured sensitive information using recursive neural networks. 2017 IEEE 33rd International Conference on Data Engineering (ICDE). doi:10.1109/icde.2017.195

[11] UCI, I. (1999, October 28). KDD Cup 1999 Dataset. Retrieved December 24, 2021, from http://kdd.ics.uci.edu/databases

[12] Cohen, W. W. (2015, May 08). Enron Email Dataset. Retrieved October 26, 2022, from https://www.cs.cmu.edu/~enron/

[13] Chicho, B. T., & Sallow, A. B. (2021). A Comprehensive Survey of Deep Learning Models Based on Keras Framework. *Journal of Soft Computing and Data Mining*, *2*(2), 49-62.

[14] Özlü, A. (2020, June 13). Long short term memory (LSTM) networks in a Nutshell. Retrieved December 24, 2021, from https://ahmetozlu.medium.com/long-short-term-memory-lstm-networks-in-a-nutshell-363cd470ccac

[15] Thakur, D. (2018, July 06). LSTM and its equations. Retrieved December 24, 2021, from https://medium.com/@divyanshu132/lstm-and-its-equations-5ee9246d04af

[16] State Government of Victoria. (2018, July 01). Maintaining confidentiality of tender participants' confidential information. Retrieved October 26, 2022, from https://www.buyingfor.vic.gov.au/probity-maintain-confidentiality-tender-participants-confidential-information-construction-guidance

[17] Popa, A. (2018, July 17). The power of a stamp or on the confidentiality of documents submitted in Public Procurement: Article: Chambers and partners. Article | Chambers and Partners. Retrieved December 24, 2021, from https://chambers.com/articles/the-power-of-a-stamp-or-on-the-confidentiality-of-documents-submitted-in-public-procurement

[18] Bakar, F. A., & Nawi, N. M. (2021). Predicting depression using social media posts. Journal of Soft Computing and Data Mining, 2(2), 39-48.