# Deep Learning Algorithm Models for Spam Identification on Cellular Short Message Service

Alfin Hikmaturokhman[1], Hanin Nafi'ah[1], Solichah Larasati[1], Ade Wahyudin[2], Galih Ariprawira[3], and Subuh Pramono[4]
[1]Telecommunication Engineering, Institut Teknologi Telkom Purwokerto, Indonesia
[2]Manajemen Teknik Studio Produksi, Sekolah Tinggi Multi Media MMTC, Yogyakarta, Indonesia
[3]Informatics Study Program, Faletehan University, Bandung, Indonesia.
[4]Department of Electrical Engineering, Faculty of Engineering, Universitas Sebelas Maret, Surakarta-Indonesia
Email: alfin@ittelkom-pwt.ac.id; 18101050@ittelkom-pwt.ac.id; laras@ittelkom-pwt.ac.id; gaprawira@gmail.com
adew001@kominfo.go.id; subuhpramono@staff.uns.ac.id

***Abstract***—Nowadays, the types and products of cellular telecommunications services are very diverse, especially with the upcoming of 5G technology, which makes telecommunications service products such as voice, video, and text messages rely on data packages. Even though the digital era is rapidly growing, the Short Messaging Service (SMS) is still relevant and used as a telecommunication service despite so many sophisticated instant messaging services that rely on the internet. Smartphone users especially in Indonesia are often terrorized by spam messages with pretentious content. Moreover, the SMS came from an unknown number and contained a message or link to a fraudulent site. This study develops a Deep Learning model to predict whether a short text message (SMS) is important or spam. This research domain belongs to Natural Language Processing (NLP) for text processing. The models used are Dense Network, Long Short Term Memory (LSTM), and Bi-directional Long Short Term Memory (Bi-LSTM). Based on the evaluation of the Dense Network model, it produces a loss of 14.22% and an accuracy of 95.63%. The evaluation of the LSTM model is 19.89% loss and 94.76% accuracy. Finally, the evaluation of the Bi-LSTM model is 19.88% loss and 94.75% accuracy.

***Index Terms***—SMS spam, deep learning, NLP, dense network, LSTM, Bi-LSTM

## I. Introduction

When the technology of communication and information in this digital era is widespread so fast, a person's ability to communicate and process information also developed; this makes the collected information an asset that can be analyzed and used to produce valuable knowledge or information in the future. However, the information conveyed sometimes can be either useful or not useful. Therefore, to anticipate the increasing number of spam messages on smartphones, a solution is needed to handle it.

Of the total 272.1 million population of Indonesia, internet users reached 175.4 million people. However, interestingly enough, the number of smartphones users reached 338.2 million units, almost double the number of internet users. It means the average Indonesian has more than one smartphone[1]. The development of mobile phone technology has a good impact on opening up new business opportunities. However, it also causes a problem for smartphone users and has become a concern for so long. Users have to be wary of spam messages. A common step to overcome this problem is to filter or manually delete spam messages. This application is available in many mobile phones, such as SMS Spam Manager running on Symbian OS and Spam SMS Blocker running on Google Android[2].

However, irresponsible parties still become more aggressive by including the names or personal identities of us, friends, closest family, even in some cases the messages contain threats. The perpetrators hope that by using this strategy, the smartphone user will become careless. Therefore, it will be very dangerous for the victim and profitable for the perpetrator.

Research on SMS spam detection has been carried out and developed using various classification methods, either with a machine learning approach or a combination of several methods to obtain accurate results. For example, one method to find out spam messages is the naive Bayes method that provides a measurement value in analyzing spam based on the frequency of spam and non-spam words. The conducted research uses English e-mails stored in the form of a .txt document. The data used is taken from the spam link dataset. First, the data processing technique comes from reading literature, journals, papers, and other readings related to data mining classification algorithms. Then, the collected data for research material was obtained online from the link spam dataset [3].

Furthermore, the research to classify and analyze SMS spam or ham uses the Support Vector Machine, Multinomial Naïve Bayes, and Decision Tree methods. The three methods chosen are used to compare and give insight into the algorithms and determine the best solution. Support Vector Machine emphasizes word relationship correlation, and Multinomial Naïve Bayes does not have

Manuscript received March 22, 2022; revised August 17, 2022.
Corresponding author email: alfin@ittelkom-pwt.ac.id
doi:10.12720/jcm.17.9.769-776

word dependence on one another, while Decision Tree is more about dividing words into classes[4].

Various algorithms such as Naïve Bayes, Support Vector Machine (SVM), and k-Nearest Neighbor (k-NN), are used in the classification stages. For example, Support Vector Machine (SVM) is one of the best methods used in pattern classification problems. On the other hand, Naïve Bayes is a classification technique based on the popular Bayesian probability theorem and creates a well-performing and straightforward model. So the research was conducted to compare the Naïve Bayes algorithm with several other algorithms, Support Vector Machine (SVM), Neural Network (NN), and Decision Tree (DT), in classifying text documents. The results obtained are that the Naïve Bayes algorithm provides the highest accuracy with a value of 97.0%, Support Vector Machine provides accuracy with a value of 96.9%, Neural Networks of 93.0%, and Decision Tree algorithm provides the lowest accuracy with a value of 91.1%[5].

This study aims to continue the previous study to classify whether the message is important or spam, using the Deep Learning algorithm with three models: Dense Network, LSTM, and Bi-LSTM. In addition, the preprocessing data text method, based on the Natural Language Processing (NLP) domain, is expected to improve the results for better model performance and be the solution for the SMS spam problem, especially in Indonesia.

There are five parts organized in this research. Part I is an introduction related to the researched case study. Part II is the base theory to explain the research reference, and Part III is the research method explaining the process and model architecture. Part IV shows the results of the evaluation and predictions. Finally, part V is the conclusions.

## II. Basic Theory

### A. Natural Language Processing

*Natural Language Processing* (NLP) is a natural language processing that uses a *Deep Learning* algorithm method with unstructured data input in text or voice messages. NLP is a branch of computer science and linguistics that discusses the interaction between humans and computers using natural language (human language).

The algorithm of Deep Learning included the NLP because this model can generalize the work process of computer learning better than the classical machine learning approach, it frequently researched and developed. Moreover, the data will be extracted automatically from the features, which helps make the model [6].

### B. Text Preprocessing

Before making an unstructured data-based model such as text or natural human language, it is necessary to do data preprocessing techniques. This technique is the beginning of the data mining process to convert raw data collected from various sources into cleaned information and used for further processing.

Several common errors in data, such as missing values, data noise, or data inconsistency, can decrease the analysis results accuracy. That is why before processing the data, we must ensure the data we are going to use is "clean" data, meaning that we have to be sure the dataset used as material to make the model must be safe from unused characters.

The process of cleaning the data also depends on the type of problem in the data set. In this study, several preprocessing techniques were carried out, including:

a. *Text Cleaning:* the process of cleaning or deleting text or characters that are non-alphabetic (not in alphabetical order) to reduce noise. This process clears a set of numbers, symbols such as periods (.), commas (,), question marks (?), exclamation points (!), and other symbols or characters.
b. *Case Folding:* a simple and effective process that converts all letters in a document to lowercase. Only letters with a string data type from a to z can do this process.
c. *Tokenizing:* the process of dividing the text in the document into different units. These units, usually called tokens, are used to factorize numbers so that computers or machines can easily understand human language.
d. *Fit on text:* the process of association between text or sentences in the report description with tokens or numbering previously created.
e. *Text to Sequence:* The process of replacing the words in the report description sentences with their respective token numbers. This process will turn each sentence into a sequence of numbers.
f. *Pad Sequence*: the process to determine the length of the sequence of words in a sentence (message) and turn it into an array of 2D Numpy
g. *Splitting dataset:* the process to divide the dataset into training data and test data, regardless of what type of data this process use, the process is important. The goal is to train and test the performance of the model used.

### C. Dense Network

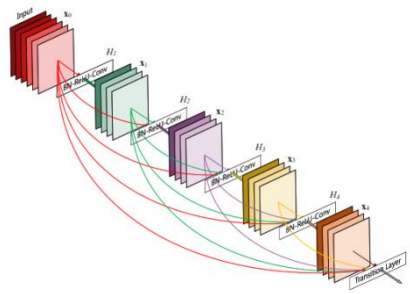


Fig. 1. *Dense Network* (DensNet) architecture[7]

In Dense Network (DensNet) architecture, each layer connects to another layer, which happens for each layer. Thus, the feature maps of all previous layers function as input, and the feature maps themselves work as an input

for each subsequent layer. DenseNet is a discovery in neural networks for visual object recognition [7]. Fig. 1 shows the DensNet architecture.

The DenseNet develops specifically to improve the decreased accuracy caused by a vanished gradient in high-level neural networks. In simple terms, the information disappears before reaching its destination due to the longer path between the input and output layers[8].

### D. Long Short Term Memory

Long Short Term Memory (LSTM) was first mentioned in 1997 by Hochreiter and Schmidhuber. LSTM is also a neural network with an adaptable architecture, so its shape can be adjusted depending on the application. Long Short Term Memory is a derivative of the RNN (Recurrent Neural Network) method. Shown in Fig. 2 is LSTM architecture.

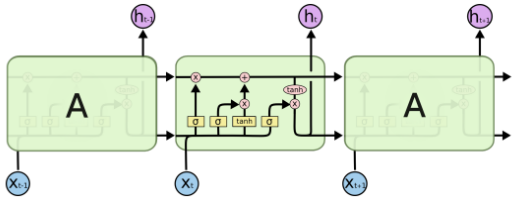


Fig. 2. *Long Short Term Memory* (LSTM) Architecture [9]

A Recurrent Neural Network is a repetitive neural network specifically designed to handle sequential data. However, RNN has the problem of vanishing and exploding gradient, which changes the range of values from one layer to the next in the architecture. The LSTM was built and designed to overcome the gradient vanishing of RNNs when dealing with vanishing and exploding gradients. The LSTM architecture consists of an input layer, an output layer, and an adaptable hidden layer [10].

### E. Bi-directional Long Short Term Memory (Bi-LSTM)

Another version of the LSTM with a bidirectional RNN (Bi-RNN) structure is called the Bidirectional LSTM (Bi-LSTM). In this version, the performance of the LSTM model can be improved, especially in the classification process. In contrast to the standard LSTM structure, two different LSTM networks train for sequential input on the BiLSTM architecture. Fig. 3 shows the basic structure of BLSTM running on sequential input

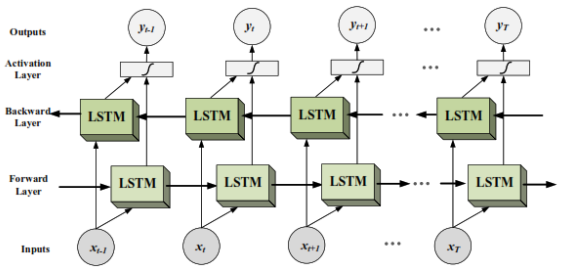


Fig. 3. *Bidirectional Long Short Term Memory* (Bi-LSTM) Architecture [11]

The Bi-LSTM structure can give better results than other network structures, depending on the problem area. For example, Bi-LSTM gives good results in voice processing tasks with an important message[11].

## III. Research Methods

### A. SMS Spam Dataset

Datasets are an important ingredient or component in research case studies and a machine learning model trainer. Various methods can do the data collection or data acquisition process. Some of those methods include observation, surveys, web scrapping, using tools (cameras, sensors, recorders, et cetera.), or accessing online websites that provide open datasets for the public, such as Kaggle and Tensorflow Dataset, et cetera.

Short Messaging Service (SMS) is a telecommunication service using text messages that are still widely used. SMS is still proven to be faster to read than messages through other chat applications. Especially now that SMS integrates with a cloud service system that is cheaper when used by certain companies or start-ups. With many cases of hacking social media accounts, such as e-mail, Facebook, e-commerce, and mobile banking or credit card services, SMS also functions as an authorization system and as double security for the account verification process.

Table I: Description of Dataset Information

| | Number of Messages Record | Number of Missing Value |
|---|---|---|
| Text | 1143 | 0 |
| Label | 1143 | 0 |

Table I shows the information on the dataset used as the research material: methods and the type for the acquiring research datasets acquired from online open datasets and the collection of messages obtained from surveys and the results of previous research observations[12]. After carrying out the dataset exploration, the information was obtained, including the dataset consisting of 2 columns, "Text" and "Label," a total of 1,143 messages consisting of 569 important SMS, 335 messages of fraud and 239 promotional messages.


Fig. 4. Wordcloud words on important SMS

Fig. 4 shows a visualization of words that fall into the SMS category of important messages. A Wordcloud that represents word frequency can be displayed in an attractive but informative form. The more often a word is used, the larger the size of the word that will be displayed

in the wordcloud. While in Fig. 5 shows a visualization of words that fall into the SMS category of spam messages.


Fig. 5. Wordcloud words on spam SMS

Even though the dataset used is free from missing values or empty data, the data preprocessing stage is still needed. One is to label important messages and then combine fraudulent, spam, and promo messages into one label. Then clean up the data by eliminating numbers and characters that are not needed in the modeling process, converting all letters in the message to lowercase, and dividing the dataset into 80% training and 20% testing data.

### B. Hyperparameter

The Hyperparameter has to be thought of when building a model because it could affect the model performance, increasing its accuracy. Here are some hyperparameters in Deep Learning, which is:

a. The number of hidden layers
b. The number of hidden neurons in each hidden layers
c. Batch Size (represents the amount of data trained in each epoch)
d. Iteration, for example, the number of 10,000 datasets and Batch size = 200, then one epoch consists of 50 iterations (10.000 divided by 200)
e. Epoch (represents one set of iteration)
f. Learning Rate
g. Regularization Parameter

Hyperparameters are configuration variables outside the model, and their values are hard to estimate from the data. That means hyperparameters cannot be learned directly from the data in standard model training. Instead, hyperparameters need to be set by machine learning engineers before training and carried out by experimentation (trial and error) until the best predictive score is obtained[13].

### C. Parameter Training

Before training using the dataset to create a model, the next stage determines the model training parameters or rules. The following are some of the important components to consider when creating a model training set, including:

a. *Activation functions* are used in artificial neural networks to activate or deactivate neurons.
b. *Loss functions* are used in artificial neural networks to calculate model errors during the optimization process.
c. *Optimizers* are used in model optimization by iteratively updating the weighted network based on training data.

The role of each parameter in the training of the Deep Learning neural network model is very important to consider. For example, how to choose the correct activation, loss, and optimizer function for our predictive modeling problem. To obtain model training results with the smallest error evaluation and the shape of the output function approaches the desired target.

### D. Supervised Learning

The proposed model is a type of classification included in the Supervised Learning algorithm or machine learning with certain guidelines using labels. The more the model processes the data, the higher the level of accuracy will be.

There are two variables in the Supervised Learning algorithm: the input variable, usually called the X variable, and the output variable, usually called the Y variable. The Supervised Learning algorithm aims to study the mapping function from the X variable to the Y variable. The general formula for mapping the X and Y variables is Y. = f(X). The ultimate goal of the Supervised Learning algorithm is to estimate the mapping function (f) so that we can predict the Y variable when we have new studied input data (X variable). Fig. 6 shows the Supervised Learning work process.

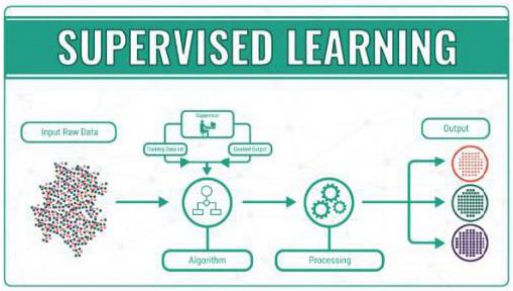


Fig. 6. *Supervised Learning* work process[14]

Supervised learning is predictive so that from the results of the many approaches that have been carried out, a prediction will be obtained. This algorithm can solve data problems linearly, multilinearly, or polynomially[14]. Supervised learning algorithms learn from labeled data. The algorithm then selects which label to apply to new data after comprehending it by associating patterns to the unlabeled new data. The evaluation of performance measurement for the machine learning classification problems using supervised learning using combinations of predicted and actual values. The following are the values of the parameters used to generate the classification model's evaluation[15].

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

where:

TP : True Positive
FP : False Positive

Precision expresses the proportion of units our model says are Positive and they actually Positive. In other words, Precision tells us how much we can trust the model when it predicts an individual as Positive[15].

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

where:

TP : True Positive

FN : False Negative

The Recall measures the model's predictive accuracy for the positive class: intuitively, it measures the ability of the model to find all the Positive units in the dataset[15].

$$F1 - score = 2\ x\ \frac{(precision\ x\ recall)}{(precision+recall)} \quad (3)$$

The formula of F1-score can be interpreted as a weighted average between Precision and Recall, where F1-score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall are equal onto the F1-score and the harmonic mean is useful to find the best trade-off between the two quantities[15].

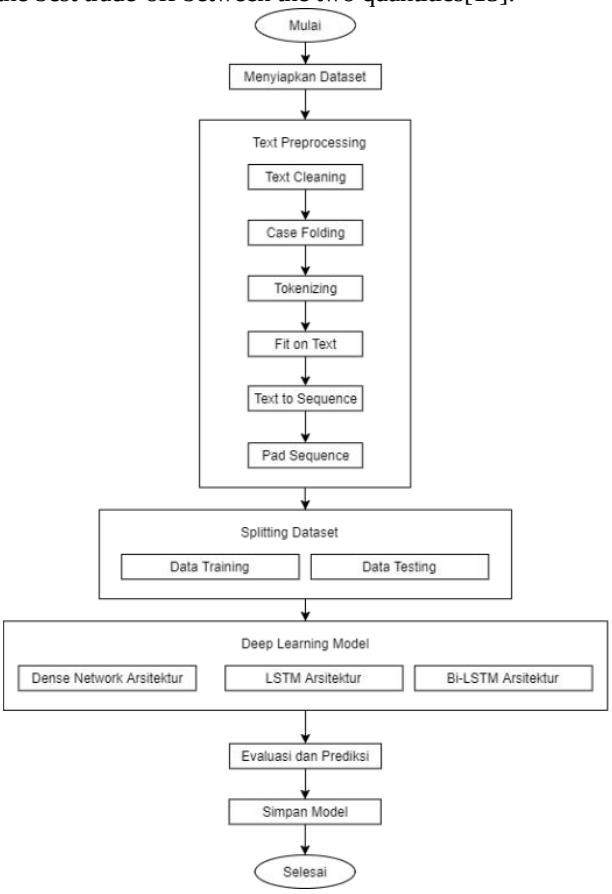


Fig. 7. *Deep Learning* process diagram for SMS identification

Fig. 7 shows a series of processes or research work steps. Starting with preparing a labeled dataset, the preprocessing stage of text messages consists of Text Cleaning, Case Folding, Tokenizing, Fit on Text, Text to Sequence, Pad Sequence. It is then going to Splitting the Dataset into training data and testing data. Next, Modeling consists of building a model architecture, predicting and evaluating the model performance results. Finally, we obtained a model for determining the output classification, whether the included message is important or spam.

### *E. Deep Learning Neural Network*

Neural Network is an information processing process inspired by the working system of human biological neural networks. A neural network consists of an input layer, hidden layer, and output layer, where each layer has several units that are interconnected between layers and have weights. There are no definite provisions in determining the architecture of each model. However, an architectural size that is too small will result in the model not learning well. Conversely, an architecture size that is too large will be weak in generalization and take up a lot of training time[16].

Deep Learning is an artificial neural network with many layers (multi-layer) or composed of more than one layer between the input and output layers, often called the hidden layer. Networks with multiple layers can solve more complex problems than single ones, of course, with more complicated learning. Deep Learning is different from traditional machine learning techniques. Deep Learning algorithms perform representations of data such as images, videos, or text automatically without introducing code rules or human domain knowledge[16]. The following are some Deep Learning algorithms used to create three types of SMS spam classification models:

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 50, 16)            8000
_________________________________________________________________
global_average_pooling1d (Gl (None, 16)                0
_________________________________________________________________
dense (Dense)                (None, 24)                408
_________________________________________________________________
dropout (Dropout)            (None, 24)                0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 25
=================================================================
Total params: 8,433
Trainable params: 8,433
Non-trainable params: 0
_________________________________________________________________
```

Fig. 8. *Dense Network Model Summary*

Fig. 8 shows the first model architecture based on Dense Network, which has five layers: 1 Embedding Layer, 1 Global Average Pooling 1D Layer, 2 Dense Layer, and 1 Dropout Layer. The total training parameters are 8,433.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 50, 16)            8000
_________________________________________________________________
lstm (LSTM)                  (None, 50, 20)            2960
_________________________________________________________________
lstm_1 (LSTM)                (None, 20)                3280
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 21
=================================================================
Total params: 14,261
Trainable params: 14,261
Non-trainable params: 0
_________________________________________________________________
```

Fig. 9. LSTM model summary

Fig. 9 shows the second model architecture based on Long Short Term Memory (LSTM) which has four layers:

1 Embedding Layer, 2 LSTM and 1 Dense Layer. The total training parameters are 14,261.

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_2 (Embedding)      (None, 50, 16)            8000
_________________________________________________________________
bidirectional (Bidirectional (None, 50, 40)            5920
_________________________________________________________________
lstm_3 (LSTM)                (None, 20)                4880
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 21
=================================================================
Total params: 18,821
Trainable params: 18,821
Non-trainable params: 0
_________________________________________________________________
```

Fig. 10. Bi-LSTM model summary

Fig. 10 shows the third model architecture based on Bi-directional Long Short Term Memory (Bi-LSTM), which has four layers: 1 Embedding Layer, 1 Bi-LSTM, 1 LSTM Layer, and 1 Dense Layer. The total training parameters are 18,821.

## IV. Results and Discussion

This study aims to understand how to build an SMS spam detection model using TensorFlow2. Specifically, to build a binary classification model to detect whether the short message (SMS) is spam or not spam (important). In addition, we have also learned how to apply Deep Learning models to Dense Network (DensNet), Long Short Term Memory (LSTM), and Bidirectional-LSTM (Bi-LSTM) using the TensorFlow2 Keras API layer.

### A. Performance Model

Evaluating the model's performance uses the same type of configured parameters, namely Binary_Crossentropy as the loss function, Adam as the optimizer, and accuracy as the metrics. Next, the epoch Hyperparameter was assigned 30 times. The results of accuracy and loss obtained can be seen in Table II as follows.

Table II: Information Dataset Description

| | *Epoch* | *Accuracy* | *Loss* |
|---|---|---|---|
| *Dense Network* | 30 times | 0,9563 | 0,1442 |
| LSTM | 15 times | 0,9476 | 0,1989 |
| Bi-LSTM | 10 times | 0,9475 | 0,1988 |

The Dense Network (DensNet) architecture model produces loss: 0.1442 and accuracy: 0.9563 within 30 epochs. The Long Short Term Memory (LSTM) architecture model produces loss: 0.1899 and accuracy: 0.9476 within 15 epochs. The model of the Bi-directional Long Short Term Memory (Bi-LSTM) architecture resulted in a loss of 0.188 and an accuracy of 0.9475 within 10 epochs. When observed, the epoch or iteration process in each model has a different number, and this is due to the Early Stopping method so that if the model starts to show signs of Overfitting, it will stop the training process.

### B. Prediction Results

Testing the system on the model uses a test dataset of 299 data taken at random based on a previously divided dataset. Confusion Matrix becomes a reference to represent predictions and actual conditions of the resulting data. Confusion Matrix can provide comparative information on classification results such as calculating accuracy, Precision, recall, and F-score values. The following is a comparison of the test results of the three trained models:

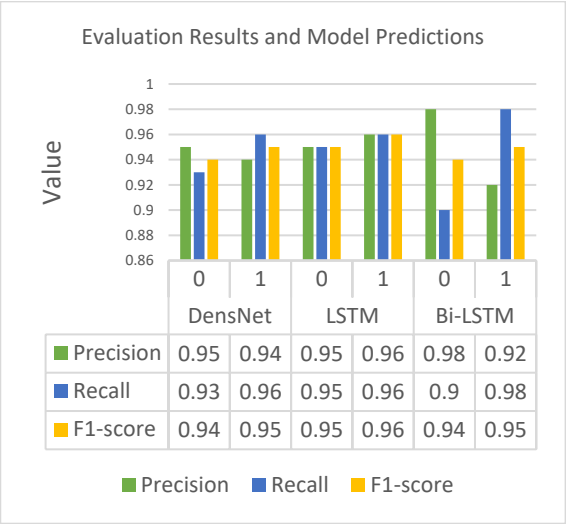


Fig. 11. Model evaluation and prediction chart

When viewed in general, in Fig. 11, the results of the evaluation and the predictions of the three models using data testing get a good value. Recall, Precision, and F-Score values are calculation values widely used to measure the performance of the trained model. Recall can measure the success rate of the system in retrieving information. The Precision determines the level of accuracy between the information provided and the answer predicted by the system. Finally, the combination of Precision and Recall produces an F-Score value.

The average value obtained by the DensNet model is a precision value of 0.945, a recall value of 0.945, and an F-score of 0.945. The average value obtained by the LSTM model is a precision value of 0.955, a recall value of 0.955, and an F-score of 0.955. The average value obtained by the Bi-LSTM model is a precision value of 0.95, a recall value of 0.94, and an F-score of 0.945. In categorical classification, Precision is useful as a positive predictive value. For example, in the case study of short message classification (SMS), the predicted value considered is Precision. Following the precision calculation, a True Positive condition occurs if they detect a new message as important. Because if we consider the calculation value with a False Positive condition, all of the incoming messages would be considered spam, and important information may not appear because the messages identified as spam messages.

## V. Conclusion

In the machine learning process, to create a model, many algorithms can be used. However, the model architecture selection needs to adjust for each type of case study currently studied. Furthermore, developing a model

does not have definite rules, so experiments are needed, for example, when processing raw datasets into datasets that are ready to use, determining hyperparameter configurations and training and research parameters so that the model accuracy results are better.

The testing process uses the three types of Deep Learning models: Dense Network, LSTM, and Bi-LSTM in the Natural Language Processing (NLP) domain. The result is a solution for smartphone users, so they do not have to spend much time deleting spam SMS in their inboxes. Furthermore, the Deep Learning algorithm method is proven to have good evaluation results. Therefore, it can support classifying SMS spam, especially for the Indonesian language in smartphones.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Hanin Nafi'ah and Ade Wahyudin carried out the desain of the study and performed the statistical analysis. Galih Ariprawira and Subuh Pramono participated in the sequence alignment and drafted the manuscript. Solichah Larasati participated in the sequence alignment and coordination and helped to draft the manuscript. Alfin Hikmaturokhman have been lead and give any recommendation about the paper and checked the paper. All authors read and approved the final manusript.

## References


[1] KOMINFO. Kadis Kominfo Sulsel Hadiri Pembukaan Sertifikasi Kompetensi Bidang Teknisi Ponsel. *20 Juni*. (2021). [Online]. Available: https://kominfo.sulselprov.go.id/post/kadis-kominfo-sulsel-hadiri-pembukaan-sertifikasi-kompetensi-bidang-teknisi-ponsel

[2] R. Mahardika, "Identifikasi konten pornografi berbahasa indonesia menggunakan algoritma Support Vector Machine (SVM)," 2018.

[3] D. Juang, "Analisis spam dengan menggunakan naïve bayes," *J. Teknovasi*, vol. 3, no. 2, pp. 51–57, 2016.

[4] A. Setiyono and H. F. Pardede, "Klasifikasi sms spam menggunakan support vector machine," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 275–280, 2019.

[5] Widyawati and Sutanto, "Perbandingan algoritma naïve bayes dan Support Vector Machine ( Svm )," *J. Ilm. Sains Dan Teknol.*, vol. 3, no. 2, pp. 178–194, 2019.

[6] S. Sharma, Natural Language Processing and its application in Human Resourcing.

[7] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017, pp. 2261–2269, 2017.

[8] Gaurav Singhal. Introduction to DenseNet with TensorFlow. *6 Mei*. (2020). [Online]. Available: https://www.pluralsight.com/guides/introduction-to-densenet-with-tensorflow

[9] C. Olah, "Understanding LSTM Networks [Blog]," *Web Page*, pp. 1–13, 2015,

[10] L. Wiranda and M. Sadikin, "Penerapan long short term memory pada data time series untuk memprediksi penjualan produk pt. Metiska farma," *J. Nas. Pendidik. Tek. Inform.*, vol. 8, no. 3, pp. 184–196, 2019.

[11] Ö. Yildirim, "A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, no. March, pp. 189–202, 2018.

[12] Y. Wibisono, "Dataset klasifikasi bahasa Indonesia (SMS Spam) & klasifikasi teks dengan scikit-learn," *5 Agustus* (2018). [Online]. Available: http://bit.ly/yw_sms_spam_indonesia

[13] Team Coach OFA, "Natural language processing (Introduction)," *Orbit Future Academy*, Jakarta, 2021.

[14] H. Abijono, P. Santoso, and N. L. Anggreini, "Algoritma supervised learning dan unsupervised learning dalam pengolahan data," *J. Teknol. Terap. G-Tech*, vol. 4, no. 2, pp. 315–318, 2021.

[15] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: An overview," arXiv preprint arXiv:2008.05756, pp. 1–17, 2020.

[16] S. R. Asriningtias, H. S. Dachlan, and E. Yudaningtyas, "Optimasi training neural network menggunakan hybrid adaptive mutation," *Eeecis*, vol. 9, no. 1, pp. 79–84, 2015.




**Alfin Hikmaturokhman** received the Bachelor's degree in Electrical Engineering from the University of Gadjah Mada (UGM), Indonesia, in 2002, and the Master's degree in Electrical Eng from Telkom University (Tel-U), Indonesia, in 2011. He has published many journal papers and conference proceedings, currently a lecturer at the Faculty of Eng, Institut Teknologi Telkom Purwokerto-Indonesia. He is a Vice Chair of IEEE Comsoc Indonesia. His research interests are mobile and wireless communication technology, both concerning technical research and regulatory policy management.

**Hanin Nafi'ah** was born in Kebumen Regency, Central Java Province, Indonesia, in 2001. She is fresh graduate student at Institut Teknologi Telkom Purwokerto, major in Telecommunication Engineering Bachelor degree. She is participating in Study Independent of Kampus Merdeka Program and learning about Artificial Intelligence. Her research

interests include wireless communication technology and Artificial Intelligence.

**Ade Wahyudin** was born in Karawang Regency, West Java Province, Indonesia, in 1985. He received the Bachelor of Engineering degree from Telkom Institute of Technology, 2010, and Master of Engineering Degree from Telkom Institute of Technology, Bandung, 2013, both in Electrical and Telecommunication Engineering. He is currently working as a lecturer in Sekolah Tinggi Multi-Media MMTC of Ministry of Communication and Information Technology, Yogyakarta. His research interest include ICT regulatory and wireless communication.

**Galih Ariprawira** obtained his B.Eng degree from the Indonesian College of Science and Technology (ST-INTEN), Indonesia, in 2007. He obtained his M.Eng degree in telecom engi from Telkom Indonesia, in 2011. He currently serves as the lecturer and Head of the Informatics Study Program, faculty of sains and technologi at the Faletehan University. His research interests include communication technology, spatial, data science and regulation of ICT.

**Subuh Pramono** was born in Indonesia. He received the B. Eng degree in Electrical Engineering from Telkom University, Indonesia in 2003, and the M. Eng degree in Electrical Engineering from Institut Teknologi Bandung (ITB), Indonesia, in 2009. Now, he is currently pursuing his Ph.D at Graduate School of Science and Engineering, Chiba University, Japan. He has published many journal papers and conference proceedings. He is currently a lecturer at Depart of Electrical Eng, Faculty of Eng, UNS, Indonesia.

**Solichah Larasati** received the S.T,M.T. degrees in telecommunication eng from Institut Teknologi Telkom Purwokerto and Telkom University Indonesia in 2016 and 2018, respectively. Her current research interests include channel coding, information theory, and wireless communication.