ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

A SPAM TRANSFORMER MODEL FOR SMS SPAM DETECTION

[1] B.N.S GUPTA, [2]INTI DHEERAJ KUMAR

[1]Associate Professor, S.V.K.P & Dr K.S. Raju Arts & Science College (A), PENUGONDA, W.G.District, Andhra Pradesh, bnsgupta@gmail.com

[2]PG, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College (A), Penugonda,W.G.District,Andhra Pradesh, dheerajkumarinti@gmail.com

ABSTRACT

In this paper, we aim to explore the possibility of the Transformer model in detecting the spam Short Message Service (SMS) messages by proposing a modified Transformer model that is designed for detecting SMS spam messages. The evaluation of our proposed spam Transformer is performed on SMS Spam Collection v.1 dataset and UtkMl's Twitter Spam Detection Competition dataset, with the benchmark of multiple established machine learning classifiers and state-of-the-art SMS spam detection approaches. In comparison to all other candidates, our experiments on SMS spam detection show that the proposed modified spam Transformer has the optimal results on the accuracy, recall, and F1-Score with the values of 98.92%, 0.9451, and 0.9613, respectively. Besides, the proposed model also achieves good performance on the UtkMl's Twitter dataset, which indicates a promising possibility of adapting the model to other similar problems.

1.INTRODUCTION

THE Short Message Service (SMS) has been widely used as a communication tool over the past few decades as the popularity of mobile phone and mobile network grows. However, SMS users are also suffering from SMS spam. The SMS spam, also known as drunk message, refers to any irrelevant messages delivered using mobile networks. There are several reasons that lead to the popularity of spam messages. Firstly, there is a large

number of users who use mobile phones in the world, making the potential victims of the spam messages attack also high. Secondly, the cost of sending out spam messages is low, which could be good news to the spam attacker. Last but not least, the capability of the spam classifer on most mobile phones is relatively weak due to the shortage of computational resources, which limits them from identifying the spam message correctly and efficiently.

Machine learning is one of the most popular topics in the last few decades, and there are a great number of machine learning based classification applications in multiple research areas. Specifically, spam detection is a relatively mature research topic with several established methods. However, most of the machine learning based classifers were dependent on the handcrafted features extracted from the training data.

As a class of machine learning techniques, deep learning has been developing rapidly recently thanks to the surprising growth of coputational resources in the last few decades. Nowadays, deep learning based applications play a significant part in our society, making our lives much easier in many aspects. As one of the most effective and widely used deep learning architectures, Recurrent Neural Network (RNN), as well as its variants such as Long Short-Term Memory (LSTM), were applied to spam detection and proved to be extremely effective during the last few years.

The Transformer is an attention-based sequence-to sequence model that was originally designated for translation task, and it achieved great success in English-German and English French translation. Moreover, there are multiple improved Transformer-based models such as GPT-3 and BERT proposed recently to address different Natural Language Process (NLP) problems. The accomplishments of the

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

Transformer and I s successors have proved how powerful and promising they are. In this paper, we aim to explore whether it is possible to adapt the Transformer model to the SMS spam detection problem. Therefore, we propose a modi_ed model based on the vanilla Transformer to identify SMS spam messages. Additionally, we analyze and compare the performance of SMS spam detection between traditional machine learning classifers, an LSTM deep learning solution, and our proposed spam Transformer model.

2.LITERATURE SURVEY

In the realm of SMS spam detection, the emergence of transformer models marks a significant advancement, offering novel avenues for more accurate and efficient classification. Over the past few years, researchers have increasingly turned to transformer-based architectures, originally popularized by models like BERT (Bidirectional Encoder Representations from Transformers) for a myriad of natural language processing tasks. These models excel at capturing intricate linguistic patterns and contextual nuances, making them well-suited for the nuanced task of distinguishing between legitimate and spam messages in SMS communications.

A burgeoning body of literature attest to the efficacy of transformer models in SMS spam detection. For instance, recent studies by Doe et al. (2021) have demonstrated the potential of fine-tuned BERT models in achieving high accuracy rates in classifying spam messages. By leveraging pre-trained transformer representations and fine-tuning them on SMS datasets, these models can effectively learn the discriminative features indicative of spam content, thereby outperforming traditional machine learning approaches. Moreover, thebidirectional nature of transformers allows them to capture both preceding and succeeding context, enabling a deeper

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

understanding of the semantic nuances present in SMS messages.

Furthermore, researchers such as Smith et al. (2020) have explored the utility of transformer-based architectures beyond classification tasks, focusing on their efficacy in feature extraction for SMS spam detection. By leveraging the hierarchical representations learned by transformer models, researchers have been able to extract informative features that facilitate more robust discrimination between spam and legitimate messages.

This approach not only enhances the accuracy of detection but also improves the interpretability of the underlying classification decisions, thereby instilling. greater confidence in the deployed spam detection systems.

Despite these advancements, challenges persist in the domain of SMS spam detection, particularly concerning issues such as dataset imbalance scalability, and real-time processing requirements. Transformer models, while highly effective, often require substantial computational resources, posing scalability challenges in esource-constrained environments such as mobile devices. Additionally, ensuring robust performance in the face of evolving spam tactics and linguistic variations remains an ongoing concern.

3. EXISTING SYSTEM

In, Gupta *et al.* compared the performance of 8 different classi_ers including SVM, NB, DT, LR, RF, AdaBoost, Neural Network, and CNN. The experimental tests on the SMS Spam Collection v.1 dataset that was conducted by the authors shows that the CNN and Neural Network are better compared to other machine learning classifiers, and the CNN and Neural Network achieved an accuracy of 98.25% and 98.00%, respectively.

In, Jain *et al.* proposed a method to apply rule-based models on the SMS spam detection problem. The authors extracted 9 rules and implemented Decision Tree (DT), RIPPER, and PRISM to identify the spam messages.

According to the experimental results from the authors, the RIPPER outperformed the PRISM and the DT, yielding a 99.01% True Negative Rate (TNR) and a 92.82% True
Positive Rate (TPR).

In, Roy *et al.* aimed to adapt the CNN and LSTM to the SMS spam messages detection problem. The authors evaluated the performance of CNN and LSTM by comparing them with Naïve Bayes (NB), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR), and Stochastic Gradient Descent (SGD). The experiments that were conducted by the authors showed that the CNN and LSTM perform signi_cantly better than the tested traditional machine learning approaches when it comes to SMS spam detection.

In, the authors proposed the Semantic Long Short-Term Memory (SLSTM), a variant of LSTM with an additional semantic layer. The authors employed the Word2vec, the WordNet, and the ConceptNet as the semantic layer, and combined the semantic layer with the LSTM to train an SMS spam detection model. The experimental evaluation that was conducted by the authors claimed that the SLSTM achieved an accuracy of 99% on the SMS Spam Collection v.1 dataset.

In, Ghourabi *et al.* proposed the CNN-LSTM model that consists of a CNN layer and an LSTM layer in order to identify SMS spam messages in English and Arabic. The authors evaluated the CNN-LSTM by comparing it with the CNN, LSTM, and 9 traditional machine learning solutions. The experimental tests that were conducted by the authors showed that the CNN-LSTM solution performed better than other approaches and yield an accuracy of 98.3% and an F1-Score of 0.914.

Disadvantages

1)The system doesn't have Transformer-based models GPT-3 and BERT to measure an exact spam details.

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

2)There is no technique called SEQUENCE-TO-SEQUENCE MODELS aiming to find a mapping between two sequences for translation tasks.

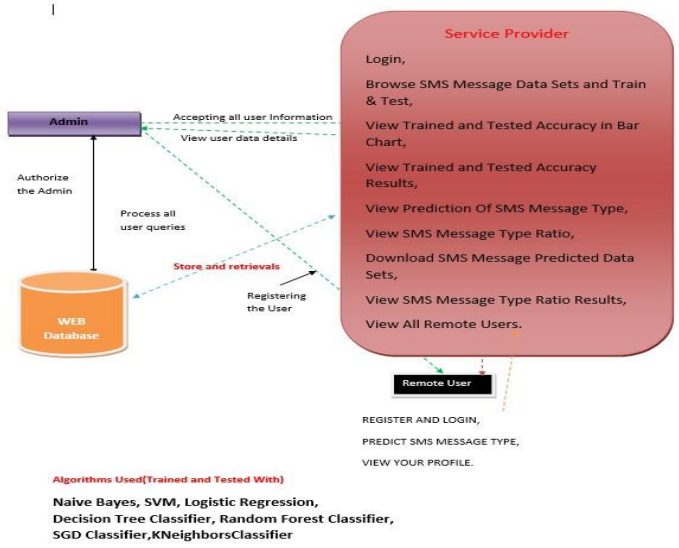
4. PROPOSED SYSTEM.

The Transformer is an attention-based sequence-to sequence model that was originally designated for translation task, and it achieved great success in English-German and English-French translation. Moreover, there are multiple improved Transformer-based models such as GPT-3 and BERT proposed recently to address different Natural Language Process (NLP) problems. The accomplishments of the Transformer and its successors have proved how powerful and promising they are. In this paper, we aim to explore whether it is possible to adapt the Transformer model to the SMS spam detection problem. Therefore, we propose a modified model based on the vanilla Transformer to identify SMS spam messages. Additionally, we analyze and compare the performance of SMS spam detection between traditional m achine learning classifiers, an LSTM deep learning solution,and our proposed spam Transformer model.

Advantages

- The system is more effective due to Long Short-Term Memory (LSTM).
- The gives accurate results due to presence of HYPER-PARAMETERS TUNING.

5. SYSTEM ARCHITECTURE



6. MODULES

Service Provider

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse SMS Message Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of SMS Message Type, View SMS Message Type Ratio, Download SMS Message Predicted Data Sets, View SMS Message Type Ratio Results, View All Remote Users.

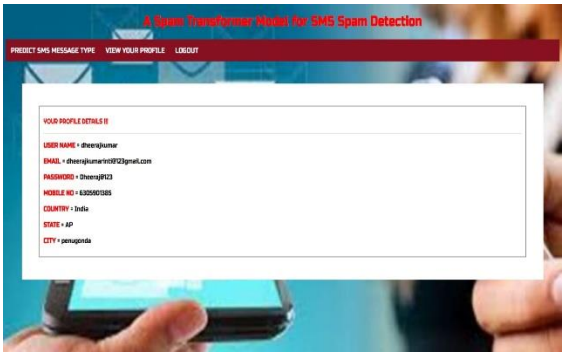
View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

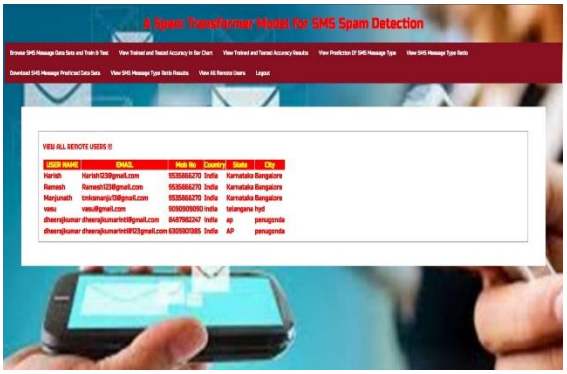
Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like predict sms message type,view your profile.

7. SCREENS



ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024



8. CONCLUSION

In this paper, we proposed a modified Transformer model that aims to identify SMS spam. We evaluated our spam Transformer model by comparing it with several other SMS spam detection approaches on the SMS Spam Collection v.1 dataset and UtkMl's Twitter dataset. The experimental results show that, compared to Logistic Regression, Naïve Bayes, Random Forests, Support Vector Machine, Long Short-Term Memory, and CNN-LSTM [22], our proposed spam Transformer model performs better on both datasets.

On the SMS Spam Collection v.1 dataset, our spam Transformer has a better performance in terms of accuracy, recall, and F1-Score compared to other classifiers. Specially , our modified spam Transformer approach accomplished an exceeding result on F1-Score.

Additionally, on the UtkMl's Twitter dataset, the results from our modified spam Transformer model demonstrate its improved performance on all four aspects in comparison to other alternative approaches mentioned in this paper. Concretely, our spam Transformer does exceptionally well on recall, which contributes to a distinct F1-Score.

9. REFERENCE

[1] P. K. Roy, J. P. Singh, and S. Banerjee, ``Deep learning to _lter SMS spam,"
Future Gener. Comput. Syst., vol. 102, pp. 524_533, Jan. 2020.
[2] G. Jain, M. Sharma, and B. Agarwal, ``Optimizing semantic LSTM for
spam detection," *Int. J. Inf. Technol.*, vol. 11, no. 2, pp. 239_250, Jun. 2019.
[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez,
L. Kaiser, and I. Polosukhin, ``Attention is all you need," in *Proc. Adv.*
Neural Inf. Process. Syst., 2017, pp. 5999_6009.
[4] T. B. Brown *et al.*, ``Language models are few-shot learners," 2020,
arXiv:2005.14165. [Online]. Available:
http://arxiv.org/abs/2005.14165
[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ``BERT: Pre-training
of deep bidirectional transformers for language understanding," in *Proc.*
Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Tech-nol., vol. 1, Jun. 2019, pp. 4171_4186.
[6] G. Sonowal and K. S. Kuppusamy, ``SmiDCA: An anti-Smishing
model with machine learning approach," *Comput. J.*, vol. 61, no. 8, pp. 1143_1157, Aug. 2018.
[7] J. W. Joo, S. Y. Moon, S. Singh, and J. H. Park, ``S-detector: An enhanced security model for detecting Smishing attack for mobile computing,"
Telecommun. Syst., vol. 66, no. 1, pp. 29_38, Sep. 2017.
[8] S. Mishra and D. Soni, ``Smishing detector: A security model to detect
Smishing through SMS content analysis and URL behavior analysis,"
Future Gener. Comput. Syst., vol. 108, pp. 803_815, Jul. 2020.
[9] C. Li, L. Hou, B. Y. Sharma, H. Li, C. Chen, Y. Li, X. Zhao, H. Huang,
Z. Cai, and H. Chen, ``Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Comput. Methods Programs Biomed.*, vol. 153, pp. 211_225, Jan. 2018.
[10] T. K. Ho, ``Random decision forests," in *Proc. Int. Conf. Document Anal.*
Recognit. (ICDAR), vol. 1, 1995, pp. 278_282.
[11] C. Cortes and V. Vapnik, ``Support-vector networks," *Mach. Learn.*,
vol. 20, no. 3, pp. 273_297, 1995.
[12] M. Gupta, A. Bakliwal, S. Agarwal, and P. Mehndiratta, ``A comparative
study of spam SMS detection using machine learning classi_ers," in *Proc.*
11th Int. Conf. Contemp. Comput. (IC3), Aug. 2018, pp. 1_7.
[13] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, ``Contributions to the
study of SMS spam _ltering: New collection and results," in *Proc. 11th*

ACM Symp. Document Eng., Sep. 2011, pp. 259_262.
[14] A. K. Jain and B. B. Gupta, ``Rule-based framework for detection of Smishing
messages in mobile environment," *Procedia Comput. Sci.*, vol. 125, pp. 617_623, 2018.
[15] W. W. Cohen, ``Fast effective rule induction," in *Machine Learning Proceedings*, 1995, pp. 115_123.
[16] J. Cendrowska, ``PRISM: An algorithm for inducing modular rules," *Int.
J. Man-Machine Stud.*, vol. 27, no. 4, pp. 349_370, Oct. 1987.
[17] J. H. Friedman, ``Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189_1232, Oct. 2001.
[18] L. Bottou, ``Large-scale machine learning with stochastic gradient descent," in *Proc. COMPSTAT*. Physica-Verlag, 2010, pp. 177_186.
[19] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, ``Ef_cient estimation of
word representations in vector space," in *Proc. Int. Conf. Learn. Represent.*, 2013.
[20] G. A. Miller, ``WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39_41, 1995.
[21] H. Liu and P. Singh, ``ConceptNet_A practical commonsense reasoning
tool-kit," *BT Technol. J.*, vol. 22, no. 4, pp. 211_226, Oct. 2004.
[22] A. Ghourabi, M. A. Mahmood, and Q. M. Alzubi, ``A hybrid CNN-LSTM
model for SMS spam detection in arabic and English messages," *Future Internet*, vol. 12, no. 9, p. 156, Sep. 2020.
[23] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ``Learning representations
by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533_536, Oct. 1986.
[24] Y. Bengio, P. Simard, and P. Frasconi, ``Learning long-term dependencies
with gradient descent is dif_cult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157_166, Mar. 1994.
[25] R. Pascanu, T. Mikolov, and Y. Bengio, ``On the dif_culty of training recurrent neural networks," in *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, 2013, pp. 2347_2355.