IJMTST is UGC Approved Journal with Journal Id: 43137

International Journal for Modern Trends in Science and Technology, 8(S03):66-69, 2022
Copyright © 2022 International Journal for Modern Trends in Science and Technology
ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST08S0317
Available online at: https://www.ijmtst.com/vol8si03.html

Spam Transformer Model for SMS Spam Detection

K.SasiKanth| A.Sri Shiva Ganga | K.SumaLatha | A.Rohit Varma | Thoto T Chopy

Department of Computer Ccience and Engineering, Godavari Institute of Engineering and Technology(A), JNTUK, Kakinada.

To Cite this Article
K.SasiKanth, A.Sri Shiva Ganga, K.SumaLatha, A.Rohit Varma and Thoto T Chopy. Spam Transformer Model for SMS Spam Detection. International Journal for Modern Trends in Science and Technology 2022, 8(S03), pp. 66-69.https://doi.org/10.46501/IJMTST08S0317

Article Info
Received: 26 April 2022; Accepted: 24 May 2022; Published: 30 May 2022.

ABSTRACT

We would discover spammer Text Message Messages (SMS) messages by presenting a new Transformers model aimed for identifying SMS spam messages. On the Spam Filtering Collections v.1 dataset and UtkMl's Twitter Spam Detection Competition dataset, we apply many existing machine learning classifiers and cutting-edge SMS spam detection algorithms to assess our proposed spam Transformer. Our SMS spam review detection testing shows that the recommended modified spam Transformer produces better results compared to the other possibilities. Furthermore, the suggested model performs well on the UtkMl Twitter dataset, indicating that it has the potential to be applied to other comparable issues.

KEYWORDS: SPAM, HAM,machine learning (ML), machine learning classifier, Naïve Bayes,SM

1.INTRODUCTION

As the use of smart devices and mobile connections has grown throughout the years, SMS has become the most used communication tool. SMS users, on the other hand, are subjected to spam. Any unrelated messages broadcast over mobile networks are referred to as SMS spam, also known as drunk messages [1]. Spam messages are popular for a variety of reasons. To begin with, there are indeed a huge number of people who are using mobile phones throughout the world, increasing the number of potential spam message victims. Second, the cost of sending spam messages is inexpensive, which is potentially excellent news for spammers.Finally, most mobile phones' spam classifiers are limited in their capacity to reliably and effectively identify spam messages due to a lack of processing resources. Machine learning has been one of the more prominent disciplines in recent decades, and there are many machine learning-based categorization applications in a variety of disciplines. Specifically, spam detection is a relatively mature research topic with several established methods. However, most of the machine learning-based classifiers were the associate editor coordinating the review of this manuscript and approving it for publication was Wei Xiang, dependent on the handcrafted features extracted from the training data [2]. As a class of machine learning techniques, deep learning has been developing rapidly recently thanks to the surprising growth of computational 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 [3] is an attention-based sequence-to-sequence model that was originally designated for translation tasks, and it achieved great success in English-German and English-French

translation. Moreover, there are multiple improved Transformer-based models such as GPT-3 [4] and BERT [5] 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. As an outcome, we develop a modified Transformer-based method for identifying SMS spam messages[22]. In addition, we examine and compare the performance of traditional machine learning classifiers, an LSTM deep learning solution, and our proposed spam Transformer model in terms of SMS spam identification.

2. RELATED WORK

In the last few decades, a variety of machine learning-based classification applications were proposed. Most of these approaches are based on traditional machine learning techniques like "Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Decision Trees in the field of SMS spam identification (DT)". With the rise in popularity of deep learning techniques, a growing variety of ways to combat SMS spam have been developed, including Convolutional Neural Networks (CNN).

3. PROPOSED WORK

The message must first be screened before it can be classified as spam or ham by a classifier. This entails tokenizing the message body's words, reducing them to their root forms, removing certain often used terms (stop words), and presenting the collection of words to the algorithm in a specified manner. A classifier is a function f that translates input feature vectors x X to output class labels y 1,..., C, with X denoting the feature space. We'll usually assume = RD or = 0, 1 D, which means the feature vector is a vector of D real numbers or D binary bits, however we can combine discrete and continuous features in general. The objective is to learn f from a supervised training set of N input-output pairings. The following is a simple illustration: X = SMS texts as input x spam, not spam is the output.

As a result, the goal is to find a predict f that translates an input vector to an output y. The data does not have to be pre-processed in deep learning; the feature is learned straight from the data set. In comparison to machine learning, it can process a larger quantity of data, resulting in superior performance. A machine learning dataset and a deep learning dataset were applied in this study.Naïve Bayes Classifier This is a method of classification based on the Bayes theorem.[20] The conditional probability of an event is a probability obtained with the additional information that some other event has already occurred. P(B|A) is used to denote the conditional probability of event B occurring, given that event A has already occurred. The following formula was provided to obtain P(B|A):

Consider *X* to denote Evidence and *Y* to denote Outcome.

P(Evidence|Outcome) is thus P(X|Y), and is represented as follows:

P(X|Y) = (P(Y|X) * P(X)) / P(Y) *(To be estimated from the training data.)*

P(Outcome|Evidence) is P(Y|X), and is represented as follows:

P(Y|X) = (P(X|Y) * P(Y)) / P(X) (To be predicted from the test data.)

4. RESULTS

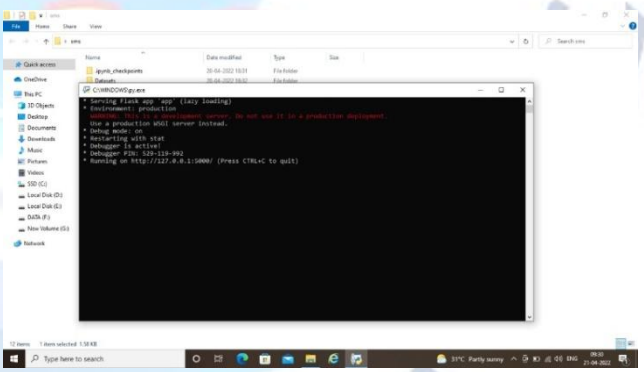

Figure 1 Launching app.py program to get a link to launch a web page

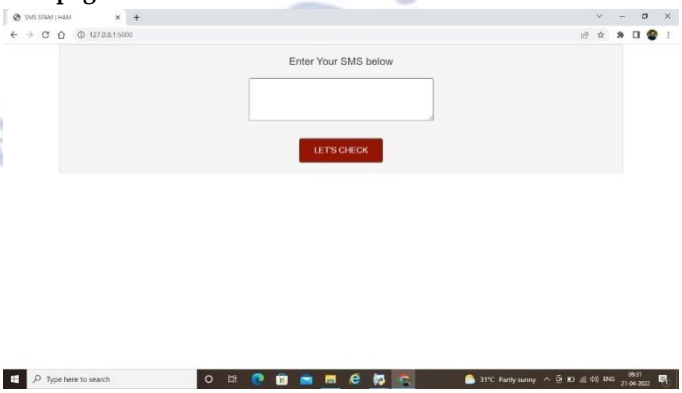

Figure 2 Web page for checking message

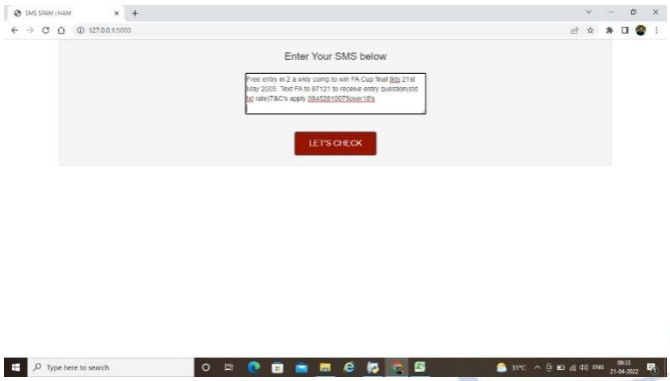

Figure 3 Entering a message to check whether it is spam and ham

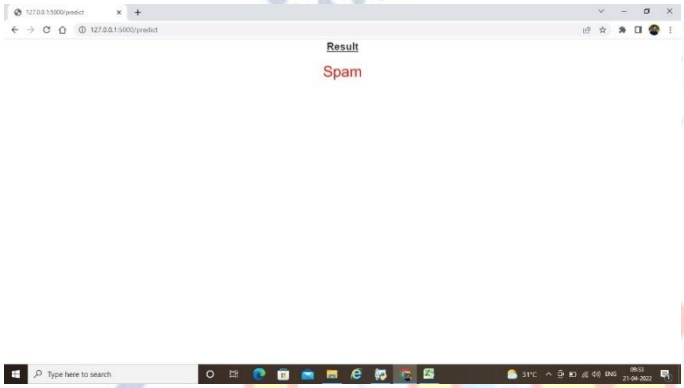

Figure 4 Getting output as spam by classifying the data

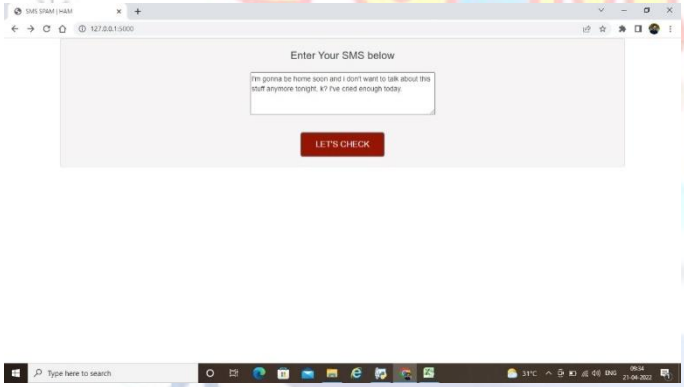

Figure 5 Entering message to check whether it is spam or ham

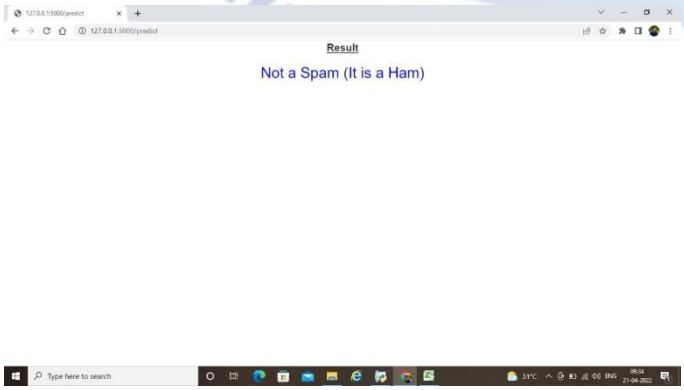

Figure 6 Getting output as Ham because it was a normal message

5. CONCLUSION

In this paper, we propose an improved Transformer model for detecting SMS spam. We evaluated our spam Transformer model against several SMS spam detection techniques using the Spam Detection Collection v.1 dataset and UtkMl's Twitter dataset. The results show that our proposed spam Transformer model outperforms Logistic Regression, Nave Bayes, Random Forests, Support Vector Machine, Long Short-Term Memory, and CNN-LSTM on both datasets [22]. On the SMS Spam Collection v.1 dataset, our spam Transformer beats previous classifiers in terms of accuracy, recall.Additionally, findings from our updated spam Transformer model using UtkMl's Twitter dataset show that it outperforms other alternative techniques in all four characteristics when compared to other approaches discussed in this study.

Firstly, since our current two datasets contain only thousands of messages, in the future, we plan to extend our spam Transformer model to a larger dataset with more messages or even other types of content, for the purpose of better performance. Besides, in our proposed model, we flattened the outputs from decoders and applied linear fully-connected layers before applying the final activation function and getting the prediction. We believe that some dedicated designs or implementations instead of simple flattening and linear layers could absolutely boost the performance, which would be one of the most important future works.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] P. K. Roy, J. P. Singh, and S. Banerjee, ''Deep learning to filter 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. Shazier, 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. Technol., vol. 1, Jun. 2019, pp. 4171–4186.

[6] G. Sonowal and K. S. Kuppusamy, ''SmiDCA: An anti-Smishing model with a 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 classifiers,'' 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 filtering: 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 a 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, ''Efficient estimation of word representationns in vector space,'' in Proc. Int. Conf. Learn. Represent., 2013.

[20] Parvathi, D. S. L., Leelavathi, N., Ravikumar, J. M. S. V., & Sujatha, B. (2020, July). Emotion Analysis Using Deep Learning. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 593-598). IEEE.

[21] Kumar, J. R., Sujatha, B., &Leelavathi, N. (2021, February). Automatic Vehicle Number Plate Recognition System Using Machine Learning. In IOP Conference Series: Materials Science and Engineering (Vol. 1074, No. 1, p. 012012). IOP Publishing."