머신러닝 기법을 이용한 한국어 보이스피싱 텍스트 분류 성능 분석

무사부부수구밀란두키스*, 진상윤**, 장대호**, 박동주**
*숭실대학교 컴퓨터학과
**숭실대학교 컴퓨터학부
mbmk92@soongsil.ac.kr, sangyoonjin.96@gmail.com, eogh1209@gmail.com, djpark@ssu.ac.kr

Korean Voice Phishing Text Classification Performance Analysis Using Machine Learning Techniques

Milandu Keith Moussavou Boussougou*, Sangyoon Jin**, Daeho Chang**, Dong-Joo Park**
*Dept. of Computer Science and Engineering, Soongsil University
** School of Computer Science and Engineering, Soongsil University

Abstract

Text classification is one of the popular tasks in Natural Language Processing (NLP) used to classify text or document applications such as sentiment analysis and email filtering. Nowadays, state-of-the-art (SOTA) Machine Learning (ML) and Deep Learning (DL) algorithms are the core engine used to perform these classification tasks with high accuracy, and they show satisfying results. This paper conducts a benchmarking performance's analysis of multiple SOTA algorithms on the first known labeled Korean voice phishing dataset called KorCCVi. Experimental results reveal performed on a test set of 366 samples reveal which algorithm performs the best considering the training time and metrics such as accuracy and F1 score.

1. Introduction

Text classification is a fundamental task of Natural Language Processing (NLP) which itself is defined as a field of artificial intelligence (AI) that give to machines the ability to understand human languages as it is spoken and written. Using NLP, machines are therefore able to read, understand and derive human languages in the cognitive way. For a long time now, text classification is being widely involved into various real applications [1-2] such as spam detection by scanning and analyzing text in emails, and label the email as spam or non-spam. In sentiment analysis, it used to determine the sentiment, the emotion behind a text and label this text as positive, negative or neutral, and also determine the author's intention. It is also used to build recommender systems based on topic, genre modeling. Intuitively, text classification falls into the category of supervised machine learning (ML) task since it assigns categories to text or documents based on the dataset's labels the ML model was train on. Fig. **2** shows the workflow used to perform text classification task.

In this research, considering the state-of-the-art (SOTA) ML and Deep Learning (DL) algorithms [3,4], several algorithms are selected and applied on the first known labeled Korean language voice phishing text dataset, Korean Call Content Vishing (KorCCVi) [5], for a benchmark analysis of the performance. The algorithms benchmarked include CatBoost, Gradient XGBoost, LGBM, Linear Support Vector Machine (Linear SVC) and Random Forest for the ML side, and Bidirectional Long Short-Term Memory units (BILSTM),

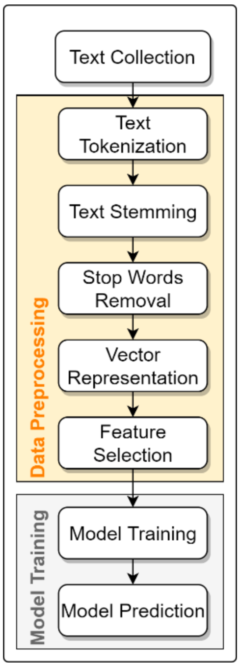

Fig. 2. Text Classification Workflow

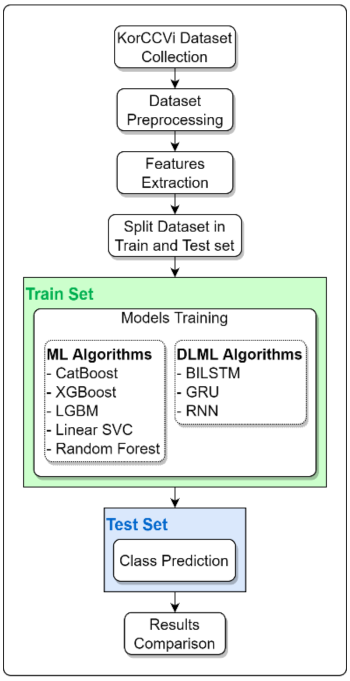

Fig. 2. Workbench Methodology

Gated Recurrent Unit (GRU) and Recurrent Neural Networks (RNN) for the DL side. Through comparative analysis of the experimental results, we provide our benchmark performance analysis result for this Korean voice phishing text classification.

2. Methodology

The methodology used conduct the benchmark performance analysis of the algorithms is presented in Fig. **2**. After collecting the KorCCVi dataset in the first step, we preprocessed it through different cleaning steps which is one of the most crucial step intext classification [6]. Once the dataset is preprocessed, features are extracted for the experimental part. The feature used include the words' term frequency-inverse document frequency (TF-IDF) and their vector representations. Then the resulting features data is split into a training and test set as shown in Table 1.The selected ML and DL text classifiers or models are thereafter train using the train set, and the final trained models are used in the prediction step to predict the class of the test set's data.

Table 1 Dataset sampling in train and test set

Algorithm Type	Training Set	Test Set	Total
Machine Learning (ML)	852	366	1218
Deep Learning (DL)	852	366	1218

In the benchmark performance analysis of the algorithms, we evaluated the model's performance using 4 metrics which are accuracy, precision, recall, F1 score, and also the modeling training time of each algorithm. Based on the metrics results of each model, we select the efficient algorithm for our Korean voice phishing text classification.

3. Experimental Results

From the results in Table 2, we calculated the score of the different metrics selected based on the ML and DL models text classification performances on the test set. The performances between the XGBoost and LGBM models are really competitive in term of modeling time as shown in Fig. 5, whereas all the other metrics of LGBM model are far over CatBoost one. Comparison of precision and recall scores, are

Table 2 Benchmarking Analysis of the ML and DL algorithms

Model	Acc.	F1	Precision	Recall	Modeling time in sec
XGBoost	0.95	0.95	0.98	0.92	0.46
LGBM	0.99	0.99	0.99	0.99	1.28
BiLSTM	0.96	0.96	0.97	0.96	1314.22
CatBoost	0.98	0.97	0.99	0.96	7.54
GRU	0.95	0.95	0.95	0.95	260.10
LinearSVC	1.00	1.00	1.00	1.00	5.07
Random Forest	1.00	1.00	1.00	1.00	3.57
RNN	0.87	0.87	0.88	0.87	7894.89

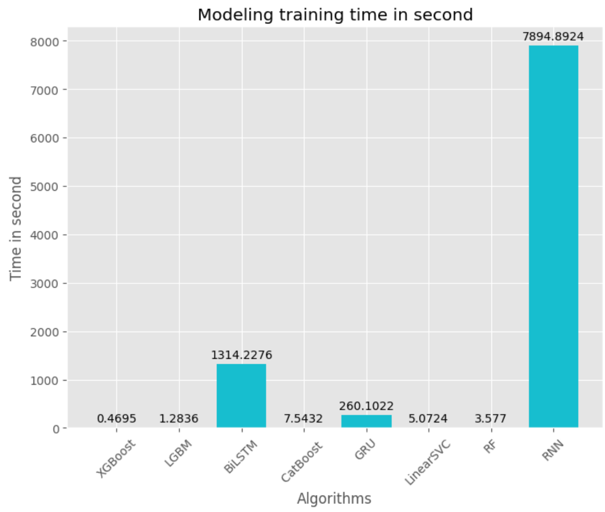

Fig. 5. Modeling Training time comparison

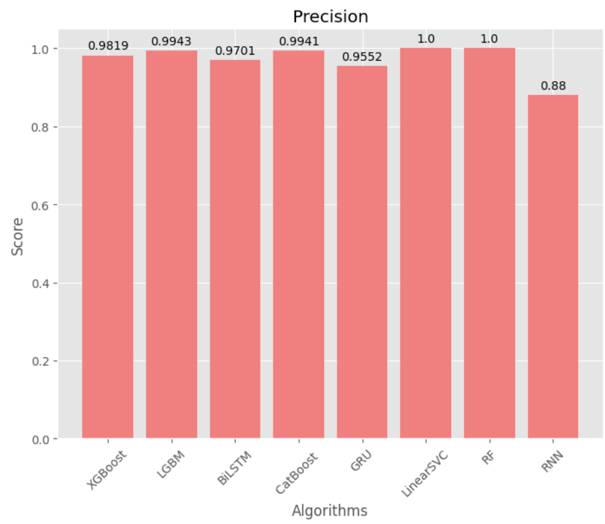

Fig. 5. Precision score comparison

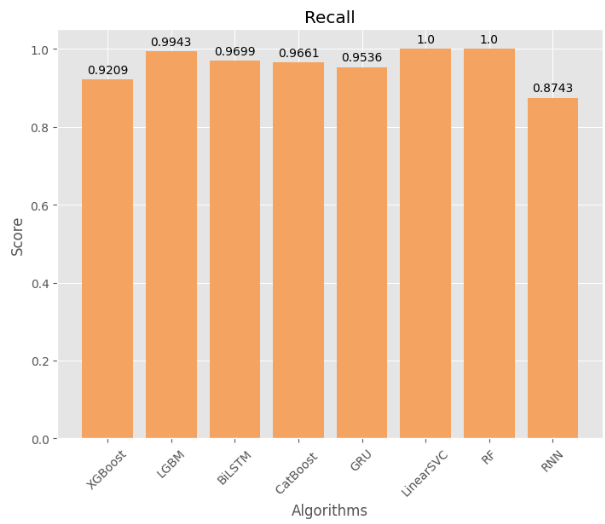

Fig. 5. Recall Score comparison

presented in Fig. 5 and Fig. 5, respectively.

4. Conclusion

Through a benchmarking performance analysis, this research has the goal of find the best text classification algorithm on the Korean voice phishing text dataset KorCCVi. From the experimental result, it has been observed that the ML algorithm LGBM outperformed all the other algorithm with the fastest modeling time of 1.28 second, and highest accuracy score and F1 score of 99% both.

5. Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW)(2018-0-00209) supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation).

6. References

[1] Kowsari, Kamran, Donald E. Brown, Mojtaba Heidarysafa, Kiana Jafari Meimandi, Matthew S. Gerber, and Laura E. Barnes. "Hdltex: Hierarchical deep learning for text classification." In 2017 16th IEEE international conference on machine learning and applications (ICMLA), pp. 364-371. IEEE, 2017.

[2] Aggarwal, Charu C., and ChengXiang Zhai. "A survey of text classification algorithms." In Mining text data, pp. 163-222. Springer, Boston, MA, 2012.

[3] Kowsari, Jafari Meimandi, Heidarysafa, Mendu, Barnes, and Brown, “Text Classification Algorithms: A Survey,” Information, vol. 10, no. 4, p. 150, Apr. 2019

[4] Ikonomakis, M., Sotiris Kotsiantis, and V. Tampakas. "Text classification using machine learning techniques." WSEAS transactions on computers 4.8 (2005): 966-974

[5] Milandu K. M. B. and, D.J. Park. "A Real-time Efficient Detection Technique of Voice Phishing with AI." 한국정보과학회 학술발표논문집 (2021): 768-770.

[6] Uysal, Alper Kursat, and Serkan Gunal. "The impact of preprocessing on text classification." Information processing & management 50, no. 1 (2014): 104-112.