

Innovative Computing Review (ICR)

Volume 2 Issue 2, Fall 2022

ISSN(P): 2791-0024 ISSN(E): 2791-0032

Homepage: <https://journals.umt.edu.pk/index.php/ICR>



Article QR



Title: Evaluating the Performance of Heterogeneous and Homogeneous Ensemble-based Models for Twitter Spam Classification

Author (s): A. O. Ameen¹, A. M. Oyelakin², I. K. Ajiboye³, I. S. Olatinwo⁴, K. Y. Obiwusi⁵, T. S. Ogundele²


Affiliation (s): ¹University of Ilorin, Ilorin, Nigeria
²Al-Hikmah University, Ilorin, Nigeria
³Abdulraheem College of Advanced Studies, Nigeria
⁴Federal Polytechnic, Offa, Nigeria
⁵Summit University, Offa, Nigeria

DOI: <https://doi.org/10.32350.icr.22.01>

History: Received: September 20, 2022, Revised: October 28, 2022, Accepted: December 5, 2022

Citation: A. O. Ameen, A. M. Oyelakin, I. K. Ajiboye, I. S. Olatinwo, K. Y. Obiwusi, and T. S. Ogundele, "Evaluating the performance of heterogeneous and homogeneous ensemble-based models for twitter spam classification," *UMT Artif. Intell. Rev.*, vol. 2, no. 2, pp. 01-16, 2022, doi: <https://doi.org/10.32350.icr.22.01>

Copyright: © The Authors

Licensing:  This article is open access and is distributed under the terms of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Conflict of Interest: Author(s) declared no conflict of interest



UMT

A publication of

School of Systems and Technology

University of Management and Technology, Lahore, Pakistan

Evaluating the Performance of Heterogeneous and Homogeneous Ensemble-based Models for Twitter Spam Classification

A. O. Ameen¹, A. M. Oyelakin², I. K. Ajiboye³, I. S. Olatinwo⁴,
K. Y. Obiwusi⁵, and T. S. Ogundele²

¹Department of Computer Science, University of Ilorin, Ilorin, Nigeria

²Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria

³Computer Science Unit, Abdulraheem College of Advanced Studies
An Affiliate of Al-Hikmah University, Ilorin, Nigeria

⁴Department of Computer Science, Federal Polytechnic, Offa, Nigeria

⁵Department of Mathematics and Computer Science, Summit University,
Offa, Nigeria

Abstract – Spam based attacks are growing in various social networks. Social network spam is a type of unwanted content that appears on social networking sites, such as Facebook, Twitter, Instagram, and others. This study used two categories of ensemble algorithms to build Twitter spam classification models. These algorithms worked by combining the strengths of individual learning algorithms and then reporting their total performances. In ensemble learning, models are formed from data based on the assumption that combining the output of multiple models is better than using a single classifier. Hence, this study used a labeled public dataset for machine learning-based Twitter spam detection. Several studies have investigated the classification of Twitter spam from the available datasets. However, there is a paucity of works that investigated how machine learning-based models, built with

homogenous and heterogeneous algorithms, behave in Twitter spam classification. ANOVA-F test was used for selecting the most promising features in the dataset. Then, homogeneous tree-based Random Forest (RF) ensemble and a heterogeneous ensemble vote classifier were employed for the classification of Twitter spam. Tree-based algorithms were used to build a homogeneous twitter spam detection model, while a combination of Support Vector Machine (SVM) and Decision Tree (DT) algorithms was used for building the heterogeneous model (using maximum voting classifier). The current study found that the performance of the Twitter spam detection models were promising. In all, the heterogeneous model recorded better performance with regards to accuracy, precision, recall, and F1-score than the model built with homogeneous base classifier.

* Corresponding Author: moyelakin80@gmail.com

Index Terms- ensemble classification, predictive accuracy, social network, Twitter spam detection

I. Introduction

Twitter is a very popular social networking platform in the internet space. It has suffered from several social spam attacks in recent years. Spam based attacks on social sites have been reported in literature in many ways [1]. Several studies reported that these attacks are carried out through bulk messages, profanity, insults, hate speech, malicious links, fraudulent reviews, fake friends, and personally identifiable information [1]. For the detection of some intrusions in the networks, Machine Learning (ML) techniques were found to be very powerful, as compared to signature based approaches [2]. Social spam is a type of spam content that appears on social media sites, such as Facebook, Twitter, and others and may include any website with user-generated content [3]. Generally, ML algorithms learn from a large set of existing data and make predictions about new data based on their learning [4].

Several studies investigated the classification of Twitter spam from the available datasets. However, there is a paucity of works that

investigated how machine learning-based models, built with homogenous and heterogeneous algorithms, behave in Twitter spam classification. Homogenous ensembles are ensembles of the same classifiers, while heterogeneous ensembles are built from different base learners [5-7].

This study used the homogenous Random Forest (RF) ensemble as well as two heterogeneous ensemble algorithms (Decision Trees and Support Vector Machine) based on maximum voting for the classification of Twitter spam. ANOVA-F test was used to handle the feature selection and ensemble approaches employed for the automatic classification of the evidence. This work seeks to extend the current authors' previous study in the area of Twitter spam classification. Generally, ensemble algorithms are of different types. This work focuses on investigating how two different categories of ensembles can correctly classify Twitter spam in a better way. In a heterogeneous ensemble-based model, two single learners are used for building the ensemble. The algorithms used in the current heterogeneous ensemble were Support Vector Machine

(SVM) and Decision Tree (DT) algorithms.

Ensemble algorithms work by combining the strengths of individual learning algorithms and then reporting their total performances. The current study used a dataset that is publicly available. This work seeks to extend a study by the authors in [8] which advocated the use of two separate homogeneous ensembles for Twitter spam classification. In the current study, the first ensemble was built from DTs as base learners, while the second one was built from SVM and DTs using maximum voting approaches. A Voting Classifier (VC) is an ensemble technique that combines the predictions of various models which together predict an output class based on their highest probability. DTs and SVM are all supervised learning algorithms used widely in various classification tasks.

II. Related Work

The authors in [9] proposed the use of a directed social graph model for the detection of Twitter spam. The methodology involved exploring the “follower” and “friend” relationships among users using a graph technique. Then, based on Twitter’s spam policy, novel content-based features and

graph-based features were also proposed. The authors built a prototype to analyse the data set and evaluate the performance of the detection system. Classic evaluation metrics were used to compare the performance of various traditional classification methods. Experimental results showed that the Bayesian classifier had the best overall performance in term of F-measure. The results also showed that the spam detection system can achieve 89% precision.

Furthermore, another study [10] used four Machine Learning (ML) techniques including Support Vector Machine (SVM), Neural Network (NN), Random Forest (RF), and Gradient Boosting (GB) to build four different Twitter spam detection models. The system works by using a structure which takes the client and tweet based highlights together with the tweet content to group the tweets. The study reported that Neural Network (NN) had a precision of 91.65% and outperformed the current arrangement by about 18%. Another system that focused on detecting spam more speedily through the creation of a large-scale annotated dataset for spam account detection on Twitter was proposed by [11]. The authors argued that the system is more

effective as compared to the existing approaches.

The researchers in [12] built a large dataset of over 600 million public tweets. Then, they labelled up to 6.5 million spam tweets and extracted 12 lightweight features. Moreover, they applied a ground truth mechanism through the use of Trend Micros Web Reputation Service as proposed by [13]. Experiments were conducted using six ML algorithms under various conditions. It was argued that the approach is effective for Twitter spam detection.

A similar study [14] carried out a review of spam attacks on the social media platforms. It focused on reporting the issues related to social spam detection, as well as the directions that future researches can take. The study reported that social media spam can be manifested in many ways, including bulk messages, profanity, insults, hate speech, malicious links, fraudulent reviews, fake friends, and personally identifiable information [14].

III. Methodology

A. Twitter Spam Dataset Source

This study used a twitter spam dataset developed by [11]. The dataset is publicly available at

<http://nslab.org/nslab/resources/>. The files in the larger dataset were originally available in ARFF format. The first step in the methodology involved changing the files into CSV format. The feature set in the dataset is shown in Table I. Each line represents a tweet from the collection. The dataset was grouped into four and twelve light weight statistical features were generated, as shown in Table I. The last column in the dataset is the tweet class (spammer or non-spammer). Exploratory data analysis revealed that the dataset is binary in nature and it allows a machine learning-based model to classify Twitter tweets as spams or non-spams. As argued further by [11], two datasets were sampled for a continuous period of time, while the other two were randomly sampled. Despite the fact that the datasets contained a smaller feature sample space, selecting the most promising features for building the dataset is a good step. Feature subset selection is a process where the most promising features are automatically selected in the data that contribute most to the target variables. Thus, feature selection involves the process of selecting a subset of relevant features for use in machine learning-based model building.

Table I
Dataset Features and their Description

| S/N | Attribute Name | Description of Attributes |
|-----|-------------------|---|
| 1 | account_age | The age (days) of an account since its creation until the time of sending the most recent tweet |
| 2 | no_follower | The number of followers of this twitter user |
| 3 | no_following | The number of followings/friends of this twitter user |
| 4 | no_userfavourites | The number of favourites this twitter user received |
| 5 | no_lists | The number of lists this twitter user added |
| 6 | no_tweets | The number of tweets this twitter user sent |
| 7 | no_retweets | The number of retweets |
| 8 | no_hashtag | The number of hashtags included in this tweet |
| 9 | no_usermention | The number of user mentions included in this tweet |
| 10 | no_urls | The number of URLs included in this tweet |
| 11 | no_char | The number of characters in this tweet |
| 12 | no_digits | The number of digits in this tweet |

Table II
Sample Size and Featureset in the Datasets

| S/N | Derived Name for the Dataset | No. of Instances in the Dataset | Input Features in the Dataset | Binary Class (Spam or Non-spam) |
|-----|------------------------------|---------------------------------|-------------------------------|---------------------------------|
| 1 | TweetContinuous1(Dataset1) | 10,000 | 12 | YES |
| 2 | TweetRandom1(Dataset2) | 10,000 | 12 | YES |
| 3 | TweetContinuous2 (Dataset3) | 100,000 | 12 | YES |
| 4 | TweetRandom2 (Dataset4) | 100,000 | 12 | YES |

The number of features and instances in each one of the datasets are also depicted in Table II. The values were obtained from the exploratory data analysis carried out. Twitter spam datasets were grouped into four based on

the number and types of tweet patterns contained therein. The datasets were labeled as Dataset 1, Dataset 2, Dataset 3, and Dataset 4. The dataset files were converted from the ARFF format to the CSV format. Firstly, exploratory data

analysis was carried out. The essence was to understand the dataset patterns in a better way and be able to gain further insights regarding how to use the available samples and features. The characteristics of the four groups of the tweet datasets are summarised in Table II.

Exploratory data analysis also revealed that each dataset contains numeric values as input features,

B. Visualisation of the Patterns in the Dataset

with the target output as categorical. No missing values were found in the datasets and all existing values were used for making decisions about how to build spam detection models. Minimal pre-processing was carried out using the encoding of the target class. This is because the target class in a text in categorical format.

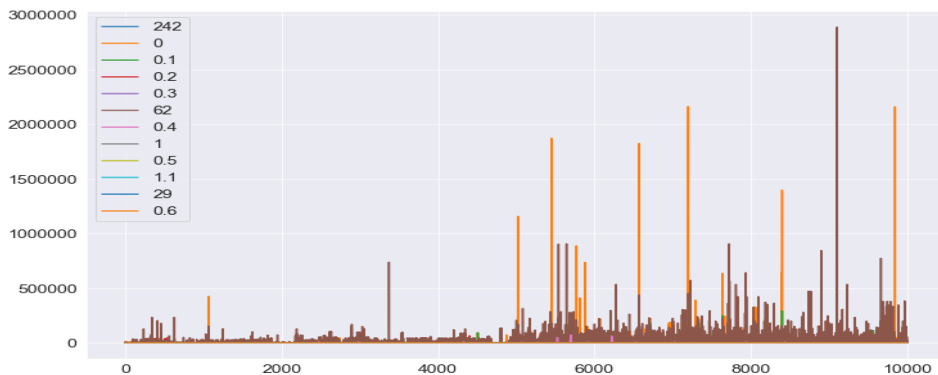


Fig. 1. Distributions chart in dataset 1

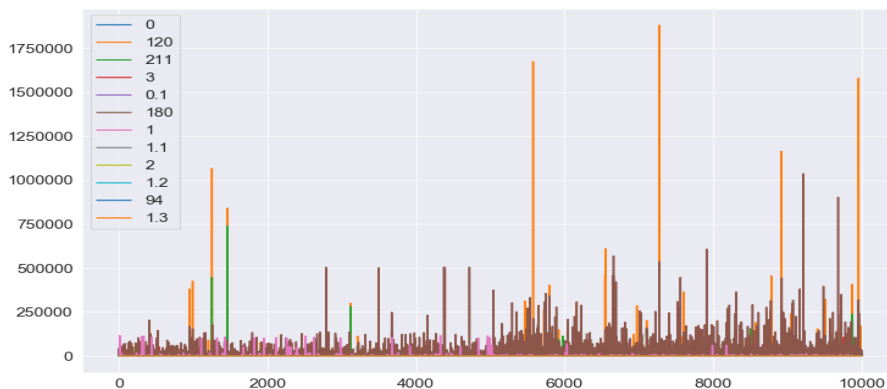


Fig. 2. Distributions chart in dataset 2

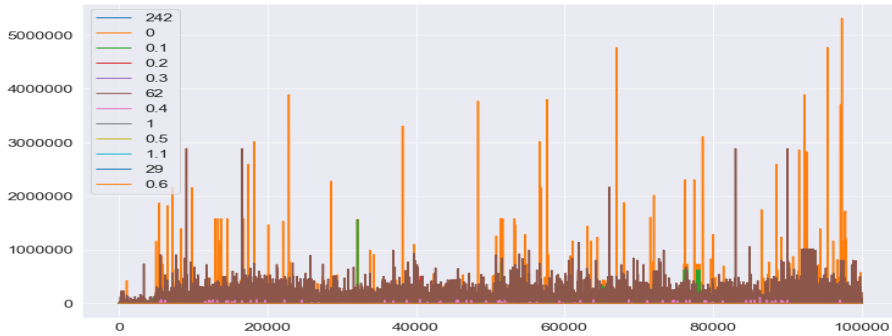


Fig. 3. Distributions chart in dataset 3

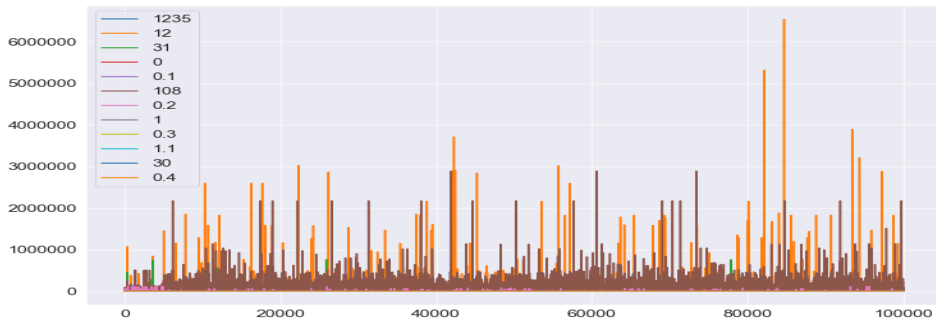


Fig. 4. Distributions chart in dataset 4

Figures 1-4 depict the distribution of data in each dataset. It is evident that data patterns differ from the first dataset to the fourth

one. Furthermore, the sample of data types in each dataset is shown below in Figure 5.

| File | Edit | Format | View | Help | | | | | | | | | |
|------|------|--------|------|------|-----|-------|-----|---|-----|-----|-----|-----|-------------|
| 0 | 242 | 0 | 0.1 | 0.2 | 0.3 | 62 | ... | 1 | 0.5 | 1.1 | 29 | 0.6 | spammer |
| 1 | 834 | 1 | 5 | 0 | 0 | 64 | ... | 1 | 0 | 1 | 22 | 0 | spammer |
| 2 | 978 | 44 | 18 | 0 | 0 | 114 | ... | 1 | 0 | 1 | 35 | 0 | spammer |
| 3 | 490 | 0 | 26 | 0 | 0 | 3840 | ... | 1 | 0 | 1 | 38 | 0 | spammer |
| 4 | 248 | 0 | 0 | 0 | 0 | 72 | ... | 1 | 0 | 1 | 42 | 4 | spammer |
| ... | 123 | 2 | 611 | 0 | 0 | 319 | ... | 0 | 0 | 1 | 31 | 0 | spammer |
| 9994 | 1444 | 1327 | 273 | 751 | 53 | 20988 | ... | 0 | 0 | 1 | 70 | 0 | non-spammer |
| 9995 | 266 | 114 | 63 | 0 | 0 | 55736 | ... | 0 | 0 | 1 | 114 | 6 | non-spammer |
| 9996 | 1068 | 364 | 470 | 83 | 0 | 3669 | ... | 0 | 0 | 1 | 64 | 1 | non-spammer |
| 9997 | 1042 | 881 | 1191 | 144 | 1 | 16798 | ... | 3 | 0 | 1 | 34 | 0 | non-spammer |
| 9998 | 1549 | 244 | 198 | 79 | 1 | 16327 | ... | 0 | 0 | 1 | 60 | 0 | non-spammer |

Fig. 5. Data distribution in dataset 1

It is evident from Figure 5 that the input attributes are in a numeric form, while the target class is categorical. Thus, the target class has to be encoded as a pre-processing step prior to building the Twitter spam model from each dataset.

C. Feature Selection Technique Used

The feature selection technique used in this study is ANOVA-F test. The choice of algorithm is based on the suitability of the said technique in view of the availability of numerical input variables and a classification of target variable. The approach identified nine features as most relevant for Twitter spam classification. These features were settled for based on their ranking. The authors in [15] emphasised the essence of feature selection and feature extraction in machine learning-based studies. Despite the fact that the feature set in the selected dataset is not too large, this study considered it important to select the most promising features for building the Twitter spam detection models, so as to guide against the problem of using all features in machine learning-based model building.

D. Homogenous and Heterogeneous Twitter Spam Detection Models

As argued by Benjamin et al. [3], machine learning-based social site spam detection can be binary class based or multiclass based. The twitter spam detection models developed in this study are binary, since the datasets are binary (spammer, non-spammer) in nature. The first model (homogenous model) was built from the default base learners of Random Forest (RF) algorithm called Decision Trees (DTs), while the second model (heterogeneous model) was built using a combination of Support Vector Machine (SVM) and Logistic Regression (LR). The result in the second model is a consequence of majority voting. All the base algorithms used are supervised learning algorithms. RF is an ensemble of DTs that make use of the bagging technique [16–18]. The algorithm creates DTs on data samples for prediction and selects the best result through voting. Given a set of Twitter tweets, the goal was to identify Twitter spam based on the patterns captured by the ML algorithms from the datasets.

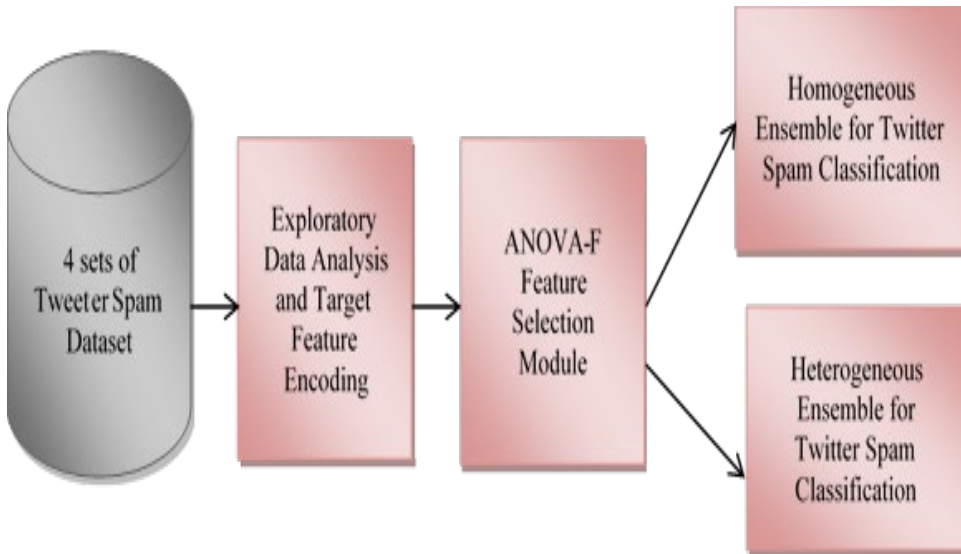


Fig. 6. Methodological flow of activities in the two Twitter spam detection models

Figure 6 is used to illustrate the different stages in the two machine learning-based Twitter spam detection models. Python was used for the implementation of various stages in the models. The basic stages in the machine learning-based model building, as argued by [19], were followed in model implementation. Since the problem at hand is of a binary type, the target was to accurately classify the Twitter spam evidence. The study used learning algorithms for automatically classifying the labeled datasets into spam and non-spam categories. The hyper parameters of the model were tuned each time until a better result was achieved.

E. Evaluation Metrics

The metrics used for evaluating the RF-based model in this study are accuracy, precision, recall, and F1-score. Brief explanation of each of the metrics is as follows:

- i. Accuracy: The ratio of the number of correctly classified cases to the total number of cases under evaluation.
- ii. Precision: The ability of a classification model to return only relevant instances.
- iii. Recall: The ability of the classifier to capture all the relevant instances.

iv. F1-score: The weighted average of the recall and precision of the respective class. The values of the metrics can be obtained by using equations (1) to (4).

$$(i) \text{ Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$(ii) \text{ Precision} = \frac{TP}{(TP+FP)} \quad (2)$$

$$(iii) \text{ Recall} = \frac{TP}{(TP+FN)} \quad (3)$$

$$(iv) \text{ F1-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4)$$

IV. Results

The results of the RF-based Twitter spam detection model were recorded and they occupied four decimal places, as shown in Table III. Similarly, the results of heterogeneous ensemble based on the voting method are shown in Table IV. The study used a repeatable train-test split approach in all scenarios for evaluating the Twitter spam classification models.

Table III
Classification Results of the Homogeneous RF-based Model

| S/N | Learning Algorithm | Metric | Model Performances |
|-----------------------------------|---------------------------------------|-----------------|--------------------|
| CASE 1 (5k continuous-Dataset 1) | | | |
| 1 | Heterogeneous Random Forest Algorithm | Accuracy | 0.9736 |
| 2 | Heterogeneous Random Forest Algorithm | Precision Score | 0.9644 |
| 3 | Heterogeneous Random Forest Algorithm | Recall | 0.9731 |
| 4 | Heterogeneous Random Forest Algorithm | F1-score | 0.9675 |
| CASE 2 (95k continuous-Dataset 2) | | | |
| 5 | Homogeneous Random Forest Algorithm | Accuracy | 0.9737 |
| 6 | Heterogeneous Random Forest Algorithm | Precision Score | 0.9696 |
| 7 | Heterogeneous Random Forest Algorithm | Recall | 0.9748 |
| 8 | Heterogeneous Random Forest Algorithm | F1-score | 0.9719 |
| CASE 3 (5k random-Dataset 3) | | | |
| 9 | Homogeneous Random Forest Algorithm | Accuracy | 0.9504 |
| 10 | Heterogeneous Random Forest Algorithm | Precision Score | 0.9426 |
| 11 | Heterogeneous Random Forest Algorithm | Recall | 0.9501 |

| S/N | Learning Algorithm | Metric | Model Performances |
|-------------------------------|---------------------------------------|-----------------|--------------------|
| 12 | Heterogeneous Random Forest Algorithm | F1-score | 0.9439 |
| CASE 4 (95k random-Dataset 4) | | | |
| 13 | Homogeneous Random Forest Algorithm | Accuracy | 0.9732 |
| 14 | Heterogeneous Random Forest Algorithm | Precision Score | 0.9676 |
| 15 | Heterogeneous Random Forest Algorithm | Recall | 0.9746 |
| 16 | Heterogeneous Random Forest Algorithm | F1-score | 0.9695 |

Table IV
Classification Results of the Heterogeneous Model Based on Maximum Voting

| S/N | Learning Algorithm | Metric | Model Performance |
|----------------------------------|----------------------------------|-----------------|-------------------|
| CASE 1 (5k continuous-Dataset1) | | | |
| 1 | Heterogeneous Ensemble Algorithm | Accuracy | 0.9922 |
| 2 | Heterogeneous Ensemble Algorithm | Precision Score | 0.9918 |
| 3 | Heterogeneous Ensemble Algorithm | Recall | 0.9922 |
| 4 | Heterogeneous Ensemble Algorithm | F1-score | 0.9917 |
| CASE 2 (95k continuous-Dataset2) | | | |
| 5 | Heterogeneous Ensemble Algorithm | Accuracy | 0.9831 |
| 6 | Heterogeneous Ensemble Algorithm | Precision Score | 0.9833 |
| 7 | Heterogeneous Ensemble Algorithm | Recall | 0.9831 |
| 8 | Heterogeneous Ensemble Algorithm | F1-score | 0.9817 |
| CASE 3 (5k random-Dataset3) | | | |
| 9 | Heterogeneous Ensemble Algorithm | Accuracy | 0.9970 |
| 10 | Heterogeneous Ensemble Algorithm | Precision Score | 0.9971 |
| 11 | Heterogeneous Ensemble Algorithm | Recall | 0.9970 |
| 12 | Heterogeneous Ensemble Algorithm | F1-score | 0.9970 |
| CASE 4 (95k random-Dataset4) | | | |
| 13 | Heterogeneous Ensemble Algorithm | Accuracy | 0.9987 |
| 14 | Heterogeneous Ensemble Algorithm | Precision Score | 0.9986 |

| S/N | Learning Algorithm | Metric | Model Performance |
|-----|----------------------------------|----------|-------------------|
| 15 | Heterogeneous Ensemble Algorithm | Recall | 0.9987 |
| 16 | Heterogeneous Ensemble Algorithm | F1-score | 0.9986 |

V. Discussion

The study carried out exploratory data analysis which provided useful information regarding how to use the datasets to build Twitter spam detection models. Having gained better insights into four different datasets as released by [11] through experimental analysis, two ensemble learning algorithms were used for building Twitter spam detection models. Then, ANOVA-F technique was used for selecting the most promising features. The selected attributes were used to build the two models. The algorithms used to build these models were based on homogeneous and heterogeneous approaches. The results obtained from the two Twitter spam classification models are described in Table III and Table IV, respectively.

During experimentation, the current study used varying training and test-split ratios to achieve the validation of the models. Good results were recorded at the split ratios of 75:25 for training and testing sets, respectively. The performance of the models was checked based on

the four selected metrics of accuracy, precision, recall, and F1-score. It was observed that the feature selection and ensemble classification methods used contributed largely to the good performance of the two models. The results obtained for different cases remain promising for both homogeneous and heterogeneous ensembles. However, it was observed that the model built with heterogeneous machine-learning algorithms (Support Vector Machine and Decision Trees) outperformed the ones built with homogeneous algorithms (Tree-based).

A. Conclusion

A general introduction to the Twitter spam classification problem as well as the promises of using machine learning-based techniques for the identification of Twitter spam attacks was made. Data pre-processing and feature selection approaches were used to feed the two ensemble algorithms with the data available in a good form. This study focused on building two different ensemble-based models in four different cases using four groups of Twitter

spam datasets. The datasets used in this study were binary in nature. Several experiments were carried out which involved using the same type of base classifiers (Decision Trees) to build a RF-based model. Furthermore, SVM-based and DT-based classifiers were used to build a heterogeneous model. During the experiments, varying random split ratios were used to achieve model validation. Good results were recorded at the split ratios of 75:25 for training and testing sets, respectively. Good performance of the models was judged based on the four selected metrics of accuracy, precision, recall, and F1-score. It was observed that the model built with heterogeneous machine learning-based algorithms outperformed the ones built with homogeneous (Tree-based) algorithms.

Acknowledgement

Authors wish to acknowledge the constructive comments of anonymous reviewers who made it possible for them to achieve the improved manuscript.

References

- [1] S. Rao, A. K. Verma, and T. Bhatia, "A review on social spam detection: Challenges, open issues, and future directions," *Expert Sys. Appl.*, vol. 186, Art. no. 115742, Dec. 2021, doi: <https://doi.org/10.1016/j.eswa.2021.115742>
- [2] A. Pektaş and T. Acarman, "Botnet detection based on network flow summary and deep learning," *Int. J. Netw. Manag.*, vol. 28, no. 6, pp. 1–15, July 2018, doi: <https://doi.org/10.1002/nem.2039>
- [3] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," *Proc. 5th Int. Workshop Advers. Inform. Retriev. Web – AIRWeb*, 2009.
- [4] S. Penchikala, "Big data processing with apache spark-part 4," *Spark Mach. Lear.*, 2016
- [5] D. Opitz, and R. Maclin, "Popular ensemble methods: An empirical study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, 1999.
- [6] R. Polikar. (2006). Ensemble based systems in decision making, *IEEE Circuits and Systems Magazine*. 6 (3): 21–45. doi:10.1109/MCAS.2006.1688199. S2CID 18032543.
- [7] Rokach, L., "Ensemble-based classifiers". *Artificial Intelligence*

- Review 2010. **33** (1–2): 1–39, 2010, doi: <https://doi.org/10.1007/s10462-009-9124-7>
- [8] R. G. Jimoh et al., “Experimental evaluation of ensemble learning-based models for twitter spam classification,” 5th Information Technology for Education and Development (ITED), 2022.
- [9] A. H. Wang, “Machine learning for the detection of spam in twitter networks,” paper presented at 7th International Joint Conference, ICETE, Athens, Greece, July 26–28, 2010 .
- [10] D. Thilagavathy, A. Muthumanickam, S. Naveenkumar, and A. U. Kumar, “Spam detection in twitter using light weight detectors,” *Int. J. Sci. Res. Comput. Sci. Appl. Manag. Stud.*, vol. 8, no. 2, 2019.
- [11] F. Concone, G. Lo Re, M. Morana, C. Ruocco, “Twitter spam account detection by effective labeling,” in Proc. 3th Italian Conf. Cyber. Secu., Pisa, Italy, Feb. 13–15, 2019.
- [12] C. Chen, J. Zhang, X. Chen, Y. Xiang, and W. Zhou, “6 Million spam tweets: A large ground truth for timely twitter spam detection,” *IEEE Int. Conf. Commun. Info. Syst. Secur. Symp.* London, UK, 2015, pp. 7065–7070, doi: <https://doi.org/10.1109/ICC.2015.7249453>
- [13] J. Oliver, P. Pajares, C. Ke, C. Chen, and Y. Xiang, “An indepth analysis of abuse on twitter,” *Trend Micro Res. Paper*, 2014
- [14] S. Rao, A. K. Verma, and T. Bhatia, “A review on social spam detection: Challenges, open issues, and future directions,” *Expert Syst. Appl.*, 2021, vol. 186, Art. no. 115742, doi: <https://doi.org/10.1016/j.eswa.2021.115742>
- [15] A. M. Oyelakin and R.G. Jimoh, “A survey of feature extraction and feature selection techniques used in machine learning-based botnet detection schemes,” *VAWKUM Transac. Comput. Sci.*, vol. 9, no. 1, pp. 1-7, 2021.
- [16] L. Breiman, “Bagging predictors,” *Mach. Learn.*, vol. 26, no. 2, pp. 123–140, 1996. <https://doi.org/10.1007/BF00058655>
- [17] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

- <https://doi.org/10.1023/A:1010933404324>
- [18] G. Brown, “Ensemble learning,” in *Encyclopedia of Machine Learning*. Springer. Boston, MA, USA, 2010.
- [19] M. Swamynathan, *Mastering machine learning with Python in six steps, A practical implementation guide to predictive data analytics using Python*. Apress, Berkeley, CA.