[Future Internet] Manuscript ID: futureinternet-1797785 - Accepted for Publication

External

Inbox

F

Future Internet Editorial Office <futureinternet@mdpi.com> Sat, Jul 23, 1:25 PM

to me, Jimmy, Fawaz, Mohammed, Future, Joseph

Dear Dr. Setyanto,

Congratulations on the acceptance of your manuscript, and thank you for submitting your work to Future Internet:

Manuscript ID: futureinternet-1797785
Type of manuscript: Article
Title: CCrFS: Combine Correlation Features Selection for Detecting Phishing Websites Using Machine Learning
Authors: Jimmy Moedjahedy *, Arief Setyanto *, Fawaz Khaled Alarfaj *, Mohammed Alreshoodi
Received: 16 June 2022
E-mails: jimmy@unklab.ac.id, arief_s@amikom.ac.id, fkarfaj@imamu.edu.sa, mo.alreshoodi@qu.edu.sa
Submitted to section: Big Data and Augmented Intelligence, https://www.mdpi.com/journal/futureinternet/sections/augmented_intelligence
https://susy.mdpi.com/user/manuscripts/review_info/5c76d596d01b0c8b9b7c5e9ccbad2381

We will now edit and finalize your paper, which will then be returned to you for your approval. Within the next couple of days, an invoice concerning the article processing charge (APC) for publication in this open access journal will be sent by email from the Editorial Office in Basel, Switzerland.

If, however, extensive English edits are required to your manuscript, we will need to return the paper requesting improvements throughout.

We encourage you to set up your profile at SciProfiles.com, MDPI's researcher network platform. Articles you publish with MDPI will be linked to your SciProfiles page, where colleagues and peers will be able to see all of your publications, citations, as well as other academic contributions.

We also invite you to contribute to Encyclopedia (https://encyclopedia.pub), a scholarly platform providing accurate information about the latest research results. You can adapt parts of your paper to provide valuable reference information, via Encyclopedia, for others both within the field and beyond.

Kind regards,
Ms. Eioau Li/MDPI
Section Managing Editor
E-Mail: eioau.li@mdpi.com
Skype: live:.cid.79367cda9a286be0

EMAIL – 5
[Future Internet] Manuscript ID: futureinternet-1797785 - Manuscript Resubmitted

External

Inbox

F

Future Internet Editorial Office <futureinternet@mdpi.com> Fri, Jul 22, 6:07 PM

to me, Jimmy, Fawaz, Mohammed

Dear Dr. Setyanto,

Thank you very much for resubmitting the modified version of the following manuscript:

Manuscript ID: futureinternet-1797785
Type of manuscript: Article
Title: CCrFS: Combine Correlation Features Selection for Detecting Phishing Websites Using Machine Learning
Authors: Jimmy Moedjahedy *, Arief Setyanto *, Fawaz Khaled Alarfaj *, Mohammed Alreshoodi
Received: 16 June 2022
E-mails: jimmy@unklab.ac.id, arief_s@amikom.ac.id, fkarfaj@imamu.edu.sa, mo.alreshoodi@qu.edu.sa

https://susy.mdpi.com/user/manuscripts/review_info/5c76d596d01b0c8b9b7c5e9ccbad2381

A member of the editorial office will be in touch with you soon regarding progress of the manuscript.

Kind regards,
Future Internet Editorial Office
Postfach, CH-4020 Basel, Switzerland
Office: St. Alban-Anlage 66, CH-4052 Basel
Tel. +41 61 683 77 34 (office)
E-mail: futureinternet@mdpi.com
https://www.mdpi.com/journal/futureinternet/

EMAIL - 4
[Future Internet] Manuscript ID: futureinternet-1797785 - Revision Reminder

External

Inbox

F

Future Internet Editorial Office <futureinternet@mdpi.com> Wed, Jul 20, 3:49 PM

to me, Jimmy, Fawaz, Mohammed, Future

Dear Dr. Setyanto,

We sent a revision request for the following manuscript on 12 July 2022.

Manuscript ID: futureinternet-1797785
Type of manuscript: Article
Title: CCrFS: Combine Correlation Features Selection for Detecting Phishing Websites Using Machine Learning
Authors: Jimmy Moedjahedy *, Arief Setyanto *, Fawaz Khaled Alarfaj *, Mohammed Alreshoodi
Received: 16 June 2022
E-mails: jimmy@unklab.ac.id, arief_s@amikom.ac.id, fkarfaj@imamu.edu.sa, mo.alreshoodi@qu.edu.sa

May we kindly ask you to update us on the progress of your revisions? If you have finished your revisions, please upload the revised version together with your responses to the reviewers as soon as possible.

You can find your manuscript and review reports at this link:
https://susy.mdpi.com/user/manuscripts/resubmit/5c76d596d01b0c8b9b7c5e9ccbad2381

Thank you in advance for your kind cooperation and we look forward to hearing from you soon.

Kind regards,
Mr. Joseph Wang
E-Mail: joseph.wang@mdpi.com

--
MDPI Wuhan Office No.6 Jingan Road, 5.5 Creative Industry Park, 25th Floor, Hubei Province, China

MDPI Future Internet Editorial Office

St. Alban-Anlage 66, 4052 Basel, Switzerland
E-Mail: futureinternet@mdpi.com
http://www.mdpi.com/journal/futureinternet

Arief Setyanto <arief_s@amikom.ac.id> Thu, Jul 21, 2:28 PM

to joseph.wang, Jimmy, Fawaz, Mohammed, Future

Dear Editor in chief,
Currently we are progressing with the revision of the manuscript, I would expect tomorrow, 22 of July to submit the final revision.

Regards

Arief Setyanto

Joseph Wang <joseph.wang@mdpi.com> Thu, Jul 21, 4:26 PM

to me, Jimmy, Fawaz, Mohammed, Future

Dear Dr. Setyanto,

Thanks for your feedback.

We are glad to know we will receive your resubmission tomorrow. We look forward to your good news.

Kind regards,
Joseph Wang
Assistant Editor

EMAIL – 3
[Future Internet] Manuscript ID: futureinternet-1797785 - Major Revisions

External
Inbox

F

Future Internet Editorial Office <futureinternet@mdpi.com> Tue, Jul 12, 9:26 AM

to me, Jimmy, Fawaz, Mohammed, Future

Dear Dr. Setyanto,

Thank you again for your manuscript submission:

Manuscript ID: futureinternet-1797785
Type of manuscript: Article
Title: CCrFS: Combine Correlation Features Selection for Detecting Phishing Websites Using Machine Learning
Authors: Jimmy Moedjahedy *, Arief Setyanto *, Fawaz Khaled Alarfaj *, Mohammed Alreshoodi
Received: 16 June 2022
E-mails: jimmy@unklab.ac.id, arief_s@amikom.ac.id, fkarfaj@imamu.edu.sa, mo.alreshoodi@qu.edu.sa

Your manuscript has now been reviewed by experts in the field. Please find your manuscript with the referee reports at this link:

https://susy.mdpi.com/user/manuscripts/resubmit/5c76d596d01b0c8b9b7c5e9ccbad2381

Please revise the manuscript according to the referees' comments and upload the revised file within 10 days.

Please use the version of your manuscript found at the above link for your revisions.

(I) Please check that all references are relevant to the contents of the manuscript.
(II) Any revisions to the manuscript should be marked up using the “Track Changes” function if you are using MS Word/LaTeX, such that any changes can be easily viewed by the editors and reviewers.
(III) Please provide a cover letter to explain, point by point, the details of the revisions to the manuscript and your responses to the referees’ comments.
(IV) If you found it impossible to address certain comments in the review

reports, please include an explanation in your rebuttal.
(V) The revised version will be sent to the editors and reviewers.

If one of the referees has suggested that your manuscript should undergo extensive English revisions, please address this issue during revision. We propose that you use one of the editing services listed at https://www.mdpi.com/authors/english or have your manuscript checked by a native English-speaking colleague.

Do not hesitate to contact us if you have any questions regarding the revision of your manuscript. We look forward to hearing from you soon.

Kind regards,
Mr. Joseph Wang
E-Mail: joseph.wang@mdpi.com

--
MDPI Wuhan Office No.6 Jingan Road, 5.5 Creative Industry Park, 25th Floor, Hubei Province, China

MDPI Future Internet Editorial Office
St. Alban-Anlage 66, 4052 Basel, Switzerland
E-Mail: futureinternet@mdpi.com
http://www.mdpi.com/journal/futureinternet

REVIEWER COMMENTS – 1

Review Report Form

Open Review
( ) I would not like to sign my review report
(x) I would like to sign my review report

English language and style
( ) English very difficult to understand/incomprehensible
( ) Extensive editing of English language and style required
( ) Moderate English changes required
(x) English language and style are fine/minor spell check required
( ) I don't feel qualified to judge about the English language and style

	Yes	Can be improved	Must be improved	Not applicable
Does the introduction provide sufficient background and include all relevant references?	(x)	( )	( )	( )
Are all the cited references relevant to the research?	(x)	( )	( )	( )
Is the research design appropriate?	(x)	( )	( )	( )
Are the methods adequately described?	( )	(x)	( )	( )
Are the results clearly presented?	( )	(x)	( )	( )
Are the conclusions supported by the results?	( )	(x)	( )	( )

Comments and Suggestions for Authors

The proposal presents a solid research work, which may be of the interest of Future Internet. Despite of this, the following enhancements are suggested prior to acceptation:

1 -The Introduction enumerates some state-of-the-art problems on the targeted research topics. However, this section do not provide insights of how the proposal (which contributions are also enumerated) is expected to outperform previous work on such issues.

2 –The introduction must enumerate the core contributions of the paper

3 –The introduction should describe the organization of the rest of the document

4 -Why Tan's dataset is a suitable options against other OSINT related collections, also seconded by the research community? Please, indicate in the manuscript

5 –Related Works may extend the table towards express the traceability between the proposal and the cited state of the art solutions. Maybe including a new row, and some additional details may be enough

6 - Overall, the paper lacks of scientific soundness. This reviewer suggests to explicitly indicate a research hypothesis and how it is contrasted by the empirical/analytical results. Other supportive inputs may be a brief description of the research objectives, assumptions, limitations, etc.

7 –A new section (for example, Discussions), may Analytically (not empirically) review the proposal highlights and results against the state of the art.

8 -The conclusions may extend the suggestions for future work

Submission Date 16 June 2022

Date of this review 11 Jul 2022 23:07:28

CORRECTION FOR REVIEW 1

Dear Reviewer,

We appreciated very much the encouraging, critical and constructive comments on this manuscript by the reviewer. The comments have been very thorough and useful to improve the

manuscript. We strongly believe that the comments and suggestions have increased the scientific value of revised manuscript. We are submitting the revised manuscript in response to all the reviewer's comments as follows:

Point 1: The Introduction enumerates some state-of-the-art problems on the targeted research topics. However, this section do not provide insights of how the proposal (which contributions are also enumerated) is expected to outperform previous work on such issues.

Response 1: Thank you very much for the constructive comment, we list the issues on the state of the art and it is presented on page 2

"…Each of these studies used a unique method for selecting features. Recent dataset comes with several features, and the assumption is some researchers use 50-100% of the total features, with complex computational time. For instance, Alotaibi and Alotaibi [11] used 23 from 48 features, Naaz [10] used all the features for 49, Hutchinson et al. [12] used 16 from 30 features, Karabatak and Mustafa [13] used 27 from 30 features, and Zaini et al. [14] used 15 from 30 features.

This paper's hypothesis is reducing low-impact features lead to simpler computation with an insignificant decline in the recognition rate for the web phishing dataset. "

Point 2: The introduction must enumerate the core contributions of the paper

Response 2: Thank you very much for the comment. We revised the paper to show the list of contributions on page 2

"In this paper, three scenarios of feature selection followed by machine learning classification. Hence, the objectives and contributions of this paper are to propose the optimal feature selection scenario for detecting phishing websites using machine learning. We observe feature selection, and recursive feature elimination by gradually decreasing and making new subset features. Additionally, to conduct a comparative analysis of three correlation methods combined with a machine learning algorithm: Pearson, Power Predictive Score (PPS), and Maximal Information Coefficient (MICe) with Total Information Coefficient (TICe) on selecting features from a dataset. Moreover, a performance comparison of four Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT) and AdaBoost algorithms on selected features is evaluated."

Point 3: The introduction should describe the organization of the rest of the document

Response 3: Thank you very much for the comment, we add the organization of the paper in the last paragraph of introduction section as follow:

"The remaining paper is organized as follows, the second section discuss material and methods, the result is presented in section 3, section 4 presents the discussion and we draw our conclusion in section 5."

Point 4: Why Tan's dataset is a suitable options against other OSINT related collections, also seconded by the research community? Please, indicate in the manuscript

Response 4: Thank you very much for this constructive comment. Based on this comment, we decided to extend our experiment to the newest dataset and found the result still consistent. The correction can be seen in table 1. The result provided in section 4, page 14-18

*"**Table 1.** Dataset information.*

Dataset/category	*Features*	*Total data*	*Non-phishing*	*Phishing*	*Published*
Tan [15]/url	*48*	*10.000*	*5000*	*5000*	*2018*
Hannousse et. Al [16]/url	*87*	*11.430*	*5715*	*5715*	*2021*

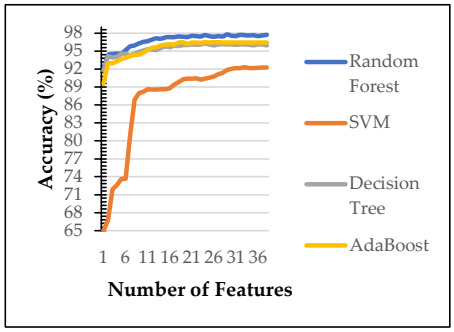


***Figure 3.** Comparison of various machine learning algorithms for dataset 1*

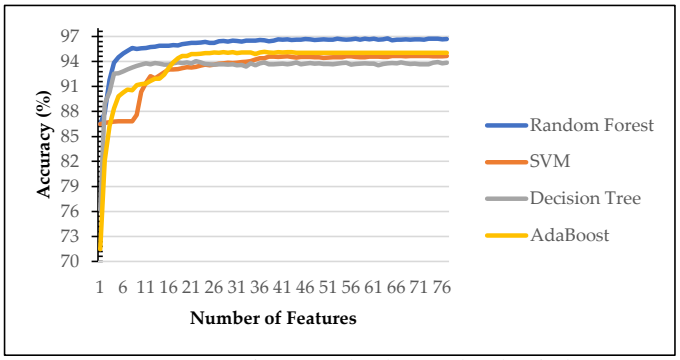


***Figure 4.** Comparison of various machine learning algorithms for dataset 2*

Point 5: Related Works may extend the table towards expressing the traceability between the proposal and the cited state of the art solutions. Maybe including a new row, and some additional details may be enough

Response 5: Thank you very much for this constructive comment. Based on this comment, we extend table 2 to page 5 and add several columns to show the dataset used, best achievement by accuracy and methods on each paper.

Commented [AS1]: Tampilkan tabelnya disini

Authors/Year	*Dataset*	*Number of features*	*Machine Learning Algorithm*	*Feature Selection algoritm*	*Accuracy*
Rao & Pais [35]/ 2019	*Build their own dataset using Phistank and Alexa*	*Not stated*	*RF, J48 Tree Decision, LR, Bayesian Network (BN), SVM, Sequential Minimal Optimization (SMO), AdaBoostM1, Multi-Layer Perceptron (MLP)*	*PCA-RF, obligate RF. PCF-RF performed the best accuracy*	*Experiment 1: 99.3%* *Experiment 2: 93.19%*
Chiew et al. [8]/ 2019	*Tan [15]*	*10*	*PART, JRip, RF, SVM, NB, C4.5*	*Hybrid Ensemble Feature Selection (HEFS), When HEFS is integrated with RF, the result outperforms another algorithm*	*94.6%*
Salihovic et al. [36] / 2019	*UCIs Phishing Websites Data Set and Spam Emails Dataset*	*10* *16*	*RF, K-Nearest Neighbor (KNN), SVM, LR, Artificial Neural Network (ANN), NB*	*BestFirst, CfsSubsEvaluation and Ranker, PrincipalComponents Optimization. Ranker + PrincipalComponents optimization using RF has the best accuracy*	*97.33%* *94.24%*
Sahingoz et al. [9] / 2019	*construct their own dataset*	*40*	*RF, DT, AdaBoost, SMO, NB, K-Star, KNN*	*Feature reduction mechanism using Natural Language Processing (NLP) features based. RF with NLP features outperform the other algorithms*	*97.98%*
Alotaibi & Alotaibi [38] / 2021	*Tan [15]*	*23*	*SVM, NB, AdaBoost, LightGBM*	*Consensus and majority voting feature selection. Consensus and voting with Adaboost and LightGMB get the best result*	*98.63%*
Haynes et al. [20] / 2021	*Shirazi Dataset*	*15*	*Artificial Neural Network (ANN), DT, Gaussian NB, Gradient Boosting (GB), KNN, RF, SVC, SVM*	*URL and HTML based, URL based only, Transformer (BERT, ELECTRA) NLP. BERT and ELECTRA using*	*More than 96%*

				ANN outperform another method	
Al-Sarem et al. [39] / 2021	*UCI dataset Tan [15] Mendeley dataset*	*Not stated*	*XGBoost, AdaBoost, RF, GradientBoost, Bagging, LightGBM*	*Ensemble, GA-based ensemble. GA-SVM outperforms another method*	*97.16%, 98.58%, 97.39%*

Point 6: Overall, the paper lacks of scientific soundness. This reviewer suggests to explicitly indicate a research hypothesis and how it is contrasted by the empirical/analytical results. Other supportive inputs may be a brief description of the research objectives, assumptions, limitations, etc.

Response 6: Thank you very much for this constructive comment. Based on this comment, we add research hypothesis objective, assumption and limitation on page 2 as follows:

".... Each of these studies used a unique method for selecting features. Recent dataset comes with several features, and the assumption is some researchers use 80-100% of the total features, with complex computational time. This paper's hypothesis is reducing low-impact features lead to simpler computation with an insignificant decline in the recognition rate for the web phishing dataset.

Hence, the objectives and contributions of this paper are to propose the optimal model for detecting phishing websites using machine learning, feature selection, and recursive feature elimination by gradually decreasing and making new subset features "

We also provide the proof of the hypothesis on table 15, 16 and also figure 5, 6 on page 15-16

"..as we can see the accuracy for dataset 1, on the full features are achieved at 98% and it is decline insignificantly from 97% to 96% on 38, 20 features to 10 features. For dataset 2, on the full features are achieved at 96% and it is decline insignificantly from 96% to 95% on 77, 60, 50, 40, 30, 20 features to 10 features".

Point 7: A new section (for example, Discussions), may Analytically (not empirically) review the proposal highlights and results against the state of the art.

Response 7: Thank you very much for this constructive comment. Based on this comment, we have already improved the discussion section (section 4) discuss the result and comparison three of our proposed approaches and comparison to the state of the art. We also identify the theoretical reason of the result and identify important features according to our study. We elaborate all the discussion on pages 15-19.

Point 8: The conclusions may extend the suggestions for future work

Thank you very much for this constructive comment, we add suggestion of future work based on the conclusion on page 18.

"…We aware that reducing the feature will definitely reduce the amount of information and our experiment result shows that consequent although the gap between full features and minimum subset (10 features) is less than 2% of accuracy. However, if the accuracy is be-ing concerned, the option of using dimensionality reduction such as principle component analysis (PCA) or autoencoder would be interesting exploration. Another

remaining problem of phishing is the evolving technique of phishing itself; therefore, exercising the new evidence by providing the latest dataset is always a challenge in the future. "

Review Report Form

Open Review
(x) I would not like to sign my review report
( ) I would like to sign my review report

English language and style
( ) English very difficult to understand/incomprehensible
( ) Extensive editing of English language and style required
( ) Moderate English changes required
(x) English language and style are fine/minor spell check required
( ) I don't feel qualified to judge about the English language and style

	Yes	Can be improved	Must be improved	Not applicable
Does the introduction provide sufficient background and include all relevant references?	( )	( )	(x)	( )
Are all the cited references relevant to the research?	(x)	( )	( )	( )
Is the research design appropriate?	( )	(x)	( )	( )
Are the methods adequately described?	( )	(x)	( )	( )
Are the results clearly presented?	(x)	( )	( )	( )
Are the conclusions supported by the results?	( )	(x)	( )	( )

Comments and Suggestions for Authors: Please see attached.

peer-review-20451151.v2.pdf

Submission Date: 16 June 2022

Date of this review: 22 Jun 2022 09:11:11

The first paragraph of the Introduction contains a number of points and it would be better if the statements were underpinned with evidence and or extended in some way. In addition, the sentence relating to visually impaired people could start the second paragraph as this is important and needs more explanation in terms of how visually impaired people are at risk and if they are a special target group. It may not be a good idea to group references as the reader may wish to follow up a specific point and references source.

Page 2 line 50: why are these defined as "five industries: banking, email, cloud, payment services, and SaaS"? Are they industries?

Page 2 line 53: what does this mean: "the website's operational lifespan is minimal"?

Who is carrying out the monitoring and analysis of websites? Is it government? This is important and attention needs to be made to this by drawing on known examples.

Section 2.1: the data referred to (source 11) is over two years old. Is it still relevant? Why use this data set?

Section 2 should contain more information about past studies using this type of methodological approach. This would place the present study in a better context. What were the main factors taken into account? What are the known drawbacks if any?

Section 4: Discussion is too short. Attention can be given also to the simulation approach and some discussion entered into. It is normal to revisit the literature and to offer explanations and insights. Also, what did the findings not indicate? Why was this the case.

What else does the reader need to know as regards Table 10?

The Conclusion is rather matter of fact and should not be a summary. It needs to be more convincing and offer some interpretation.

On reflection various aspects of the paper need to be extended and developed. More needs to be included about machine learning capability for example. The abstract can be reviewed for completeness also.

CORRECTION FOR REVIEW 2

Dear Reviewer,
We appreciated very much the encouraging, critical and constructive comments on this manuscript by the reviewer. The comments have been very thorough and useful to improve the manuscript. We strongly believe that the comments and suggestions have increased the scientific value of revised manuscript. We are submitting the revised manuscript in response to all the reviewer's comments as follows:

Point 1: The first paragraph of the Introduction contains a number of points and it would be better if the statements were underpinned with evidence and or extended in some way. In addition, the sentence relating to visually impaired people could start the second paragraph as this is important and needs more explanation in terms of how visually impaired people are at risk and if they are a special target group. It may not be a good idea to group references as the reader may wish to follow up a specific point and references source.

Response 1: Thank you very much for this constructive comment, I agree and we revised the sentence in first paragraph on page 1. We remove the statement "*visually impaired people...*" and also revise the group reference to single reference.

Point 2: Page 2 line 50: why are these defined as "five industries: banking, email, cloud, payment services, and SaaS"? Are they industries?

Response 2: Thank you very much for this constructive comment, I agree and we revise the sentence in first paragraph in page 2.

"Meanwhile, according to Phislabs' data [6], during 2019 83.9% of phishing attacked services are financial, email, cloud, payment, and SaaS."

Point 3: Page 2 line 53: what does this mean: "the website's operational lifespan is minimal"?

Response 3: Thank you very much for this constructive comment, we revised the sentence as follows:

"and the website's operational lifespan is limited"

Point 4: Who is carrying out the monitoring and analysis of websites? Is it government? This is important and attention needs to be made to this by drawing on known examples.

Response 4: Thank you very much for this constructive comment.

The monitoring ana analysis is becoming concern of all the company and institution. An international coalition called Anti-Phishing Working group (APWG), was founded in 2003, APWG as an international coalition of counter-cybercrime responders, forensic investigators, law enforcement agencies, technology companies, financial services firms, university researchers, NGOs and multilateral treaty organizations operating as a non-profit organization. Its directors, managers and research fellows advise national and sub-national governments as well as the United Nations (Office on Drugs and Crime) as recognized experts (as defined by the Doha Declaration of 2010 and Salvador Declaration of 2015) as well as multilateral bodies and organizations. We include this coalition report on page 2.

Point 5: Section 2.1: the data referred to (source 11) is over two years old. Is it still relevant? Why use this data set?

Response 5: Thank you very much for this constructive comment, we have decided to extend the study with a newer dataset. The detail of dataset is presented in table 1, the result is presented and discuss throughout Result and Discusion section.

Table 1. Dataset information.

Dataset/category	Features	Total data	Non-phishing	Phishing	Published
Tan [15]/url	48	10.000	5000	5000	2018
Hannousse et. Al [16]/url	87	11.430	5715	5715	2021

Point 6: Section 2 should contain more information about past studies using this type of methodological approach. This would place the present study in a better context. What were the main factors taken into account? What are the known drawbacks if any?

Response 6: Thank you very much for this constructive comment. Based on this comment we extend table 2 to page 5, we add 2 columns to show the dataset used, best achievement and methods on each paper best achievement.

Past study : Each of these studies used a unique method for selecting features. Recent datasets come with several features, and some researchers use 80-100% of the total features.

This Study : This paper's hypothesis is reducing low-impact features leads to simpler computation with an insignificant decline in the recognition rate for the web phishing dataset. Therefore, we try to observe the possibility to use only 12% to 25% only and maintain high accuracy.

Point 7: Section 4: Discussion is too short. Attention can be given also to the simulation approach and some discussion entered into. It is normal to revisit the literature and to offer explanations and insights. Also, what did the findings not indicate? Why was this the case.

Response 7: Thank you very much for this constructive comment. Based on this comment, we have already improved the discussion section (section 4) discuss the result and comparison three of our proposed approaches and comparison to the state of the art. We also identify the theoretical reason of the result and identify important features according to our study. We elaborate all the discussion on pages 15-19.

Point 8: What else does the reader need to know as regards Table 10?

Response 8: Thank you very much for this constructive comment. This table's number changed to Table 17 because we added a new dataset. Table 17 shows several research that used the same dataset and compared the accuracy and number of features selected.

Table 17. The studies feature set and accuracy comparison on Tan's dataset

No	Feature Typeset	Features	Accuracy
Chiew et al. [8]	Full Features	48	96.17%
	Baseline	10	94.60%
	Full Features HEFS	30	92.40%
	Baseline HEFS	5	93.22%
Khan et al. [17]	Full Features	48	97.87%
	Using PCA	30	94.90%
Dangwal et al. [18]	RF	30	93.7%
Ours	Full features	48	**98.1%**
	PPS Correlation Only	38	**98.16%**
	PPS+RFE	30	**97.96%**
	PPS+RFE	20	**97.63%**
	PPS+RFE	10	**96.96%**
	MICe TICe Correlation Only	38	**97.86%**
	MICe TICe + RFE	30	**97.53%**
	MICe TICe + RFE	20	**97.53%**
	MICe TICe + RFE	10	**97.06%**
	Spearman Correlation Only	38	**97.86%**
	Spearman + RFE	30	**97.83%**
	Spearman + RFE	20	**97.6%**
	Spearman + RFE	10	**96.63%**

Point 9: The Conclusion is rather matter of fact and should not be a summary. It needs to be more convincing and offer some interpretation.

Response 9: Thank you very much for this constructive comment. Based on this comment, we have already improved the conclusion section.

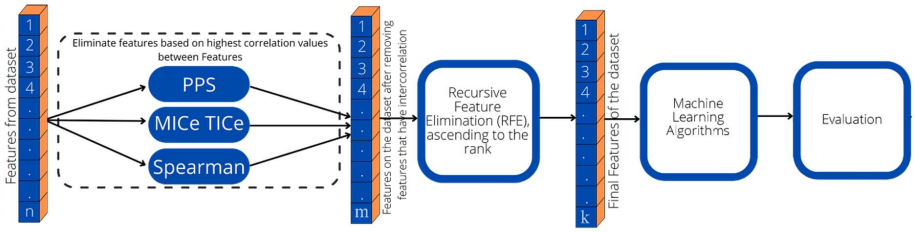
“In conclusion, Removal of inter-correlated features and low and negative correlation features to the output label leads to a better recognition rate of phishing dataset. The subset of phishing dataset selected with PPS+RFE scenario slightly overperforms compared to MICe TICe+RFE and Spearman+RFE. According to the experimental result, Random Forest (RF) achieves better accuracy in recognizing the phishing website in Tan's [15] and Hannousse and Yahiouche ’s [16] dataset. Our approach to feature selection and classification using ran-dom forest achieves slightly better accuracy on Tan's [15] dataset at 96,96 % of accuracy, compared to the result reported in [7] and [18] at 94,6% and 93.7%, respectively. On Hannousse and Yahiouche's [16] dataset, our method of feature selection and classification using random forest also produces marginally better accuracy at 97.96% “

Point 10: On reflection various aspects of the paper need to be extended and developed. More needs to be included about machine learning capability, for example. The abstract can be reviewed for completeness also.

Response 10: Thank you very much for this constructive comment. Based on this comment, we have revised our abstract we add our proposed approach on feature selection, our approach on machine learning algorithms, best achievement of accuracy on reduced features on two datasets.

“In this work, we proposed a method that combines correlation and recursive feature elimination to determine which URL characteristics are useful for identifying phishing websites by gradually decreasing the number of features while maintaining accuracy value. In this paper, we use two datasets that contain 48 and 87 features. The first scenario combines power predictive score correlation and recursive feature elimination; the second scenario is the maximal information coefficient correlation and recursive feature elimination. The third scenario combines spearman correlation and recursive feature elimination. All three scenarios from the combined findings of the proposed methodologies achieve a high level of accuracy even with the smallest feature subset.”

On methods, we add machine learning and evaluation after two round feature selection



On result, an additional figure to shows the accuracy on gradually increase number of features on two datasets

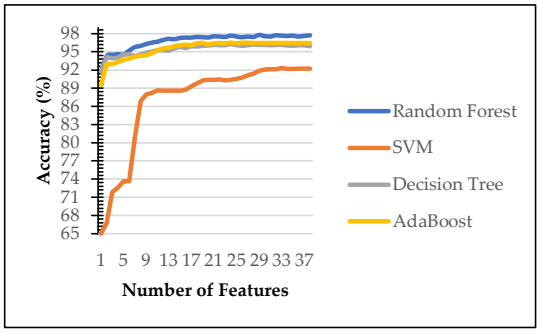


Figure 3. Comparison of various machine learning algorithms for dataset 1

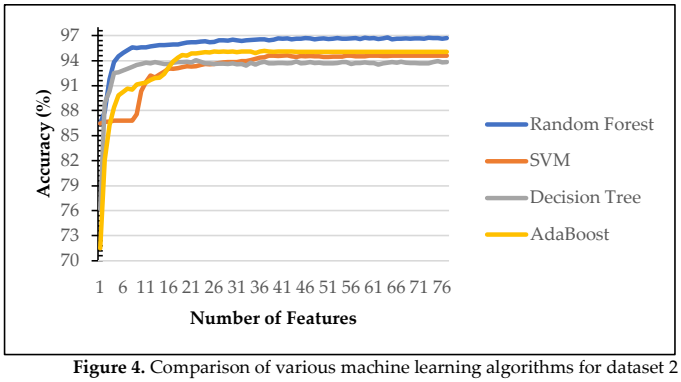


Figure 4. Comparison of various machine learning algorithms for dataset 2

We add an analysis of our result and its comparison of the latest work on the same effort on the same dataset with different features selection and machine learning algorithm.

EMAIL – 2
[Future Internet] Manuscript ID: futureinternet-1797785 - Assistant Editor Assigned

External

Inbox

J

Joseph Wang <joseph.wang@mdpi.com> Fri, Jun 17, 3:42 PM

to me, Joseph, Jimmy, Fawaz, Mohammed, Future

Dear Dr. Setyanto,

Your paper has been assigned to Joseph Wang, who will be your main point of contact as your paper is processed further.

Journal: Future Internet
Manuscript ID: futureinternet-1797785
Title: CCrFS: Combine Correlation Features Selection for Detecting Phishing Websites Using Machine Learning
Authors: Jimmy Moedjahedy *, Arief Setyanto *, Fawaz Khaled Alarfaj *, Mohammed Alreshoodi

Received: 16 June 2022
E-mails: jimmy@unklab.ac.id, arief_s@amikom.ac.id, fkarfaj@imamu.edu.sa, mo.alreshoodi@qu.edu.sa

You can find it here:
https://susy.mdpi.com/user/manuscripts/review_info/5c76d596d01b0c8b9b7c5e9ccbad2381

Best regards,
Mr. Joseph Wang
E-Mail: joseph.wang@mdpi.com

--
MDPI Wuhan Office No.6 Jingan Road, 5.5 Creative Industry Park, 25th Floor, Hubei Province, China

MDPI Future Internet Editorial Office
St. Alban-Anlage 66, 4052 Basel, Switzerland

E-Mail: futureinternet@mdpi.com
http://www.mdpi.com/journal/futureinternet

EMAIL – 1
[Future Internet] Manuscript ID: futureinternet-1797785 - Submission Received

External

Inbox

F

Editorial Office <futureinternet@mdpi.com> Thu, Jun 16, 7:41 PM

to me, Jimmy, Fawaz, Mohammed

Dear Dr. Setyanto,

Thank you very much for uploading the following manuscript to the MDPI submission system. One of our editors will be in touch with you soon.

Journal name: Future Internet
Manuscript ID: futureinternet-1797785
Type of manuscript: Article
Title: CCrFS: Combine Correlation Features Selection for Detecting Phishing Websites Using Machine Learning
Authors: Jimmy Moedjahedy *, Arief Setyanto *, Fawaz Khaled Alarfaj *, Mohammed Alreshoodi
Received: 16 June 2022
E-mails: jimmy@unklab.ac.id, arief_s@amikom.ac.id, fkarfaj@imamu.edu.sa, mo.alreshoodi@qu.edu.sa

You can follow progress of your manuscript at the following link (login required):
https://susy.mdpi.com/user/manuscripts/review_info/5c76d596d01b0c8b9b7c5e9ccbad2381

The following points were confirmed during submission:

1. Future Internet is an open access journal with publishing fees of 1400 CHF for an accepted paper (see https://www.mdpi.com/about/apc/ for details). This manuscript, if accepted, will be published under an open access Creative Commons CC BY license (https://creativecommons.org/licenses/by/4.0/), and I agree to pay the Article Processing Charges as described on the journal webpage (https://www.mdpi.com/journal/futureinternet/apc). See https://www.mdpi.com/about/openaccess for more information about open access publishing.

Please note that you may be entitled to a discount if you have previously received a discount code or if your institute is participating in the MDPI Institutional Open Access Program (IOAP), for more information see

https://www.mdpi.com/about/ioap. If you have been granted any other special discounts for your submission, please contact the Future Internet editorial office.

2. I understand that:

a. If previously published material is reproduced in my manuscript, I will provide proof that I have obtained the necessary copyright permission. (Please refer to the Rights & Permissions website: https://www.mdpi.com/authors/rights).

b. My manuscript is submitted on the understanding that it has not been published in or submitted to another peer-reviewed journal. Exceptions to this rule are papers containing material disclosed at conferences. I confirm that I will inform the journal editorial office if this is the case for my manuscript. I confirm that all authors are familiar with and agree with submission of the contents of the manuscript. The journal editorial office reserves the right to contact all authors to confirm this in case of doubt. I will provide email addresses for all authors and an institutional e-mail address for at least one of the co-authors, and specify the name, address and e-mail for invoicing purposes.

If you have any questions, please do not hesitate to contact the Future Internet editorial office at futureinternet@mdpi.com

Kind regards,
Future Internet Editorial Office
St. Alban-Anlage 66, 4052 Basel, Switzerland
E-Mail: futureinternet@mdpi.com
Tel. +41 61 683 77 34
Fax: +41 61 302 89 18