THE IMPACT OF SOFTWARE METRICS IN NASA METRIC DATA PROGRAM DATASET MODULES FOR SOFTWARE DEFECT PREDICTION

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

ADINDA AYU PUSPITA RAMADHANI

NIM 2011016120003

PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU

FEBRUARI 2024

THE IMPACT OF SOFTWARE METRICS IN NASA METRIC DATA PROGRAM DATASET MODULES FOR SOFTWARE DEFECT PREDICTION

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

ADINDA AYU PUSPITA RAMADHANI

NIM 2011016120003

PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU

FEBRUARI 2024

SKRIPSI

THE IMPACT OF SOFTWARE METRICS IN NASA METRIC DATA PROGRAM DATASET MODULES FOR SOFTWARE DEFECT PREDICTION

Oleh :

Adinda Ayu Puspita Ramadhani
NIM. 2011016120003

Telah dipertahankan di depan Dosen Penguji pada tanggal 19 Januari 2024.

Susunan Dosen Penguji:

Dosen Pembimbing I

Radityo Adi Nugroho, S.T., M.Kom.

NIP. 198212042008011006

Dosen Penguji I

Friska Abadi, S.Kom., M.Kom.

NIP. 198809132023211010

Dosen Pembimbing II

Mohammad Reza Faisal, S.T., M.T., Ph.D.

NIP. 197612202008121001

Dosen Penguji II

Rudy Herteno, S.Kom., M.Kom.

NIP. 198809252022031003

Koordinator PS Ilmu Komputer,

Irwan Budiman, S.T., M.Kom.

NIP. 197703252008121001

KATA PENGANTAR

Puji syukur saya panjatkan pada Allah SWT karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal yang berjudul "*The impact of software metrics in NASA Metric Data Program dataset modules for software defect prediction*" untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Penulis ingin mengucapkan ucapan terima kasih kepada semua pihak yang telah memberikan dukungan kepada penulis dalam pembuatan dan penyusunan jurnal ini, adapun yang dimaksud adalah sebagai berikut:

1. Bapak Radityo Adi Nugroho, S.T., M.Kom. selaku dosen pembimbing utama dan Bapak M. Reza Faisal, S. T., M. T., Ph. D. Selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.
2. Bapak Irwan Budiman, S. T., M. Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM
3. Ibu Dwi Kartini, S.Kom., M.Kom. selaku ketua panitia skripsi Porgram Studi Ilmu Komputer FMIPA ULM
4. Seluruh Dosen dan Staff Prodi Ilmu Komputer yaitu Ka Azizah atas bantuan dan saran beliau jurnal ini dapat diselesaikan dengan baik.
5. Keluarga terutama kedua orang tua yang selalu memberikan doa dan dukungan dalam penyelesaian jurnal.
6. Kakak Ajwa Helisa dan Bang Muhammad Meftah Mafazy yang telah membantu dalam pemberkasan sehingga jurnal ini bisa terselesaikan dengan baik.
7. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan jurnal.
8. Serta semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian jurnal ini.

Penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna. Meski demikian, diharapkan tulisan ini dapat memberikan manfaat bagi ilmu pengetahuan dan pembaca, serta mendapatkan berkah yang melimpah dari Tuhan Yang Maha Esa.

Banjarbaru, 23 Februari 2024

Adinda Ayu Puspita Ramadhani

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam jurnal ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 23 Februari 2024

Yang Menyatakan,

Adinda Ayu Puspita Ramadhani
NIM. 2011016210001

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET, DAN TEKNOLOGI
UNIVERSITAS LAMBUNG MANGKURAT
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM STUDI ILMU KOMPUTER
Jl. A.Yani Km 36, Banjarbaru, Kalimantan Selatan 70714 Telp/Fax. (0511) 4773112/4782899

BERITA ACARA
PERSETUJUAN REKOGNISI KEGIATAN PENGGANTI SKRIPSI

Pada hari ini, Senin Tanggal 26 Februari 2024 telah dilakukan evaluasi kelengkapan dokumen pengganti kegiatan skripsi mahasiswa S1 Ilmu Komputer Fakultas MIPA ULM

Nama	: Adinda Ayu Puspita Ramadhani
NIM	: 2011016120003
Jenis Kegiatan	: Jurnal Internasional
Kualifikasi Jurnal/Seminar	: Scopus Q3
Judul Publikasi	: The impact of software metrics in NASA Metric Data Program dataset modules for software defect prediction
Nama Jurnal/Penyelenggara	: TELKOMNIKA
Tanggal Publikasi/Presentasi	: 2 FEBRUARI 2024

Dokumen yang diajukan dinyatakan :*)

1. **Memenuhi syarat untuk diterima sebagai REKOGNISI PENGGANTI KEGIATAN SKRIPSI dan DINYATAKAN LULUS SKRIPSI.**
2. **Tidak memenuhi syarat untuk diterima sebagai REKOGNISI PENGGANTI KEGIATAN SKRIPSI.**

Demikian Berita Acara ini dibuat untuk digunakan sebagaimana mestinya.

No	Nama	Tanda Tangan	
1.	Friska Abadi S.Kom M.Kom	1.	
2.	Rudy Herteno S.Kom M.Kom		2.
3.	Radityo Adi Nugroho S.T., M.Kom	3.	
4.	Mohammad Reza Faisal, S.T, M.T, Ph.D.		4.

Koordinator PS Ilmu Komputer

Irwan Budiman, S.T., M.Kom.
NIP. 197703252008121001

Panitia Skripsi,

Dwi Kartini, S.Kom., M.Kom
NIP. 198704212012122003

LoA

UNIVERSITAS
LAMBUNG MANGKURAT

ADINDA AYU PUSPITA RAMADHANI <2011016120003@mhs.ulm.ac.id>

[TELKOMNIKA] Editor Decision "The impact of software metrics in NASA Metric Data Program dataset modules for software defect prediction"

1 message

Tole Sutikno <telkomnika@ee.uad.ac.id> Fri, Feb 2, 2024 at 9:41 PM
Reply-To: "Prof. Dr. Ir. Tole Sutikno, ASEAN Eng." <telkomnika@ee.uad.ac.id>
To: Adinda Ayu Puspita Ramadhani <2011016120003@mhs.ulm.ac.id>
Cc: Adinda Ayu Puspita Ramadhani <dindayupr@gmail.com>, Radityo Adi Nugroho <radityo.adi@ulm.ac.id>, Mohammad Reza Faisal <reza.faisal@ulm.ac.id>, Friska Abadi <friska.abadi@ulm.ac.id>, Rudy Herteno <rudy.herteno@ulm.ac.id>

-- URGENT !!! Paper ID# 25787

Dear Prof/Dr/Mr/Mrs Adinda Ayu Puspita Ramadhani,

It is my great pleasure to inform you that your paper entitled "The impact of software metrics in NASA Metric Data Program dataset modules for software defect prediction" is ACCEPTED and will be published on the TELKOMNIKA Telecommunication Computing Electronics and Control, ISSN 1693-6930; a SCOPUS/ScimagoJR indexed journal, CiteScore: 3.8 (Q2 CiteScore Quartile in Electrical and Electronic Engineering); SNIP: 0.558 and SJR: 0.286. Congratulations!

Thank you

Sincerely yours,
Prof. Dr. Ir. Tole Sutikno, ASEAN Eng.
Universitas Ahmad Dahlan
telkomnika@ee.uad.ac.id

Form Review

UNIVERSITAS LAMBUNG MANGKURAT

ADINDA AYU PUSPITA RAMADHANI <2011016120003@mhs.ulm.ac.id>

[TELKOMNIKA] Editor Decision: RESUBMIT for REVIEW

1 message

Tole Sutikno <telkomnika@ee.uad.ac.id> Fri, Dec 22, 2023 at 2:02 PM
Reply-To: "Prof. Dr. Tole Sutikno" <telkomnika@ee.uad.ac.id>
To: Adinda Ayu Puspita Ramadhani <2011016120003@mhs.ulm.ac.id>
Cc: Adinda Ayu Puspita Ramadhani <dindayupr@gmail.com>, Radityo Adi Nugroho <radityo.adi@ulm.ac.id>, Mohammad Reza Faisal <reza.faisal@ulm.ac.id>, Friska Abadi <friska.abadi@ulm.ac.id>, Rudy Herteno <rudy.herteno@ulm.ac.id>

-- Paper ID# 25787
-- Number of min references for research paper is 25 references
-- (and min 20 recently journal articles)
-- Adhere every detail of the guide of authors
(http://iaescore.com/gfa/telkomnika.docx)
==========

Dear Prof/Dr/Mr/Mrs: Adinda Ayu Puspita Ramadhani,

We have reached a decision regarding your submission entitled "The impact of software metrics in NASA Metric Data Program dataset modules for software defect prediction" to TELKOMNIKA (Telecommunication Computing Electronics and Control), ISSN 1693-6930; a SCOPUS/ScimagoJR indexed journal, CiteScore 2019: 1.7; SNIP 2019: 0.659 and SJR 2019: 0.214.

Our decision is to RESUBMIT for REVIEW
Please find below the comments from the peer reviewers and editors.

Checklist for preparing your paper for publication:
http://telkomnika.uad.ac.id/index.php/TELKOMNIKA/about/editorialPolicies#custom-2

TELKOMNIKA Guide of authors: http://iaescore.com/gfa/telkomnika.docx

I look forward for hearing from you

Thank you

Best Regards,
Tole Sutikno, Ph.D.
Editor-in-Chief, TELKOMNIKA
email: telkomnika@ee.uad.ac.id

When revising your paper, keep the following points in mind:

1. Authors should have made substantial contributions to:

(a) the conception and design of the study, or acquisition of data, or analysis and interpretation of data
(b) drafting the article or revising it critically for important intellectual content

2. Introduction section

Explain the context of the study and state the precise objective
An Introduction should contain the following three parts:
- Background: Authors have to make clear what the context is. Ideally, authors should give an idea of the state-of-the art of the field the report is about.
- The Problem: If there was no problem, there would be no reason for writing a manuscript, and definitely no reason for reading it.
So, please tell readers why they should proceed reading. Experience shows that for this part a few lines are often sufficient.
- The Proposed Solution: Now and only now! - authors may outline the contribution of the manuscript. Here authors have to make sure readers point out what are the novel aspects of authors work. Authors should place the paper in proper context by citing relevant papers. At least, 10 references (recently journal articles) are used in this section.

3. Results and Discussion

The presentation of results should be simple and straightforward in style. This section reports the most important findings, including results analyses as appropriate and comparisons to other research results. This section should be supported suitable references.

4. Conclusion

Your conclusion should make your readers glad they read your paper. Summarize sentences the primary outcomes of the study in a paragraph (NOT in numbering).

5. References and Citations

We usually expect a minimum of 25 and 40 references primarily to recent journal papers for research/original paper and review paper, respectively.
Each citation should be written in the order of appearance in the text [1], [2], [3], [4], ... (Sequential order!!)
URGENT: SELF CITATION from same author or group authors is MAX 10%.

6. Paragraph

A paragraph is a sentence or group of sentences that support one main idea. Many authors have presented paragraphs in very long terms. Author should use simple sentences which are grammatically correct, but too many can make your writing less interesting.
Every paragraph in a paper should be:
- Unified: All of the sentences in a single paragraph should be related to a single controlling idea (often expressed in the topic sentence of the paragraph).

- Clear: The sentences should all refer to the central idea of the paper.
- Coherent: The sentences should be arranged in a logical manner and should follow a definite plan for development.
- Well-developed: Every idea discussed in the paragraph should be adequately explained and supported through evidence and details that work together to explain the paragraph's controlling idea.

Reviewer A:

Ethics (Plagiarism, Fraud, Other ethical concerns)
- If you suspect that an article is a substantial copy of another work, please let the editor know, citing the previous work in as much detail as possible.
- It is very difficult to detect the determined fraudster, but if you suspect the results in an article to be untrue, inform it
- Has there been a violation of the accepted norms in the ethical?
Please provide your detailed comments to the Author(s) on the following.:

Does the title of the paper accurately reflect the major focus contribution of this paper?:
Yes

If No, please suggest change of the title as appropriate within 10 words:

Is the abstract an appropriate and adequate digest of the work?:
Yes

Is the paper clear, concise, and well organized?:
No

Structure and Presentation (Layout and format, Title, Abstract, Introduction, Method, Results and Discussion, Conclusion, Language). Please provide your detailed comments to the Author(s) on the following:

Do authors place the paper in proper context by citing relevant papers? Is the paper free from obvious errors, misconceptions, or ambiguity? Is the paper written in correct English? Please note grammatical errors and suggest corrections:

Previous Research: If the article builds upon previous research does it reference that work appropriately? Are there any important works that have been omitted? Are the references accurate? Do authors place the paper in proper context by citing relevant papers? Please provide your detailed comments to the Author(s) on the following.:

Rate of the contribution strength to the field and the scientific quality is represented in this paper?:
Below average

Please mark appropriate scale for the overall grade for this paper?:
Below average

Reviewer's comments and suggestions how to improve the paper. (If it is not possible, kindly please use separate sheets or a copy of the paper for comments and suggestions for revision. Indicate whether revisions are mandatory or suggested. Please use word processing type format if possible, and then upload in next step or submit via email: telkomnika@uad.ac.id)

URGENT: Authors should consider all above items for preparing their

revision (and NOT only the below comments)
:

1. In your figures 1 and 2 section, I suggest providing an explanation of the vertical and horizontal axis to make it easier for the reader to understand what you mean.
2. Tables 1-3 are not mentioned in the text. I suggest that tables should be explained in the sentence before the table is displayed. All tables must be referenced in the text.
3. I found that some paragraphs in your paper consist of less than 3 sentences. I suggest to the writer to adjust the paragraphs to at least 3 sentences and a maximum of 7 sentences.
4. In your acknowledgments section, I recommend including the contract number of your research (if any).
5. The references [7], [17], [24], and [31] you used for your paper do not use the format suggested. I suggest to the author to readjust the reference writing in your paper to the IEEE format, you can learn more at the following link:
https://ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE-Reference-Guide.pdf

---

Reviewer I:

Ethics (Plagiarism, Fraud, Other ethical concerns)
- If you suspect that an article is a substantial copy of another work, please let the editor know, citing the previous work in as much detail as possible.
- It is very difficult to detect the determined fraudster, but if you suspect the results in an article to be untrue, inform it
- Has there been a violation of the accepted norms in the ethical?
Please provide your detailed comments to the Author(s) on the following.:

Does the title of the paper accurately reflect the major focus contribution of this paper?:
Yes

If No, please suggest change of the title as appropriate within 10 words:

Is the abstract an appropriate and adequate digest of the work?:
Yes

Is the paper clear, concise, and well organized?:
No

Structure and Presentation (Layout and format, Title, Abstract, Introduction, Method, Results and Discussion, Conclusion, Language). Please provide your detailed comments to the Author(s) on the following:

Do authors place the paper in proper context by citing relevant papers? Is the paper free from obvious errors, misconceptions, or ambiguity? Is the paper written in correct English? Please note grammatical errors and suggest corrections:

Previous Research: If the article builds upon previous research does it reference that work appropriately? Are there any important works that have been omitted? Are the references accurate? Do authors place the paper in proper context by citing relevant papers? Please provide your detailed comments to the Author(s) on the following.:

Rate of the contribution strength to the field and the scientific quality is represented in this paper?:

Good

Please mark appropriate scale for the overall grade for this paper?:

Below average

Reviewer's comments and suggestions how to improve the paper. (If it is not possible, kindly please use separate sheets or a copy of the paper for comments and suggestions for revision. Indicate whether revisions are mandatory or suggested. Please use word processing type format if possible, and then upload in next step or submit via email: telkomnika@uad.ac.id)

URGENT: Authors should consider all above items for preparing their revision (and NOT only the below comments)

:

-Tables should be described in the sentence before the table is displayed.

-Overall, this paper is well written and informative. The abstract contains research, objectives and results. Likewise, the conclusion is conveyed well.

UNIVERSITAS LAMBUNG MANGKURAT

ADINDA AYU PUSPITA RAMADHANI <2011016120003@mhs.ulm.ac.id>

[TELKOMNIKA] TELKOMNIKA Editor Decision - Revisions required !!

1 message

Tole Sutikno <telkomnika@ee.uad.ac.id> Fri, Jan 12, 2024 at 9:24 PM
Reply-To: "Prof. Dr. Tole Sutikno" <telkomnika@ee.uad.ac.id>
To: Adinda Ayu Puspita Ramadhani <2011016120003@mhs.ulm.ac.id>
Cc: Adinda Ayu Puspita Ramadhani <dindayupr@gmail.com>, Radityo Adi Nugroho <radityo.adi@ulm.ac.id>, Mohammad Reza Faisal <reza.faisal@ulm.ac.id>, Friska Abadi <friska.abadi@ulm.ac.id>, Rudy Herteno <rudy.herteno@ulm.ac.id>

-- Paper ID# 25787
-- Number of min references for research paper is 25 references (and min 20 recently journal articles)
-- Adhere every detail of the guide of authors (http://iaescore.com/gfa/telkomnika.docx)
==========

Dear Prof/Dr/Mr/Mrs: Adinda Ayu Puspita Ramadhani,

It is my great pleasure to inform you that your manuscript entitled "The impact of software metrics in NASA Metric Data Program dataset modules for software defect prediction" to TELKOMNIKA (Telecommunication Computing Electronics and Control), (a SCOPUS/ScimagoJR indexed journal, CiteScore: 2.9, SNIP: 0.579, SJR: 0.314) get positive response from Reviewers/Editors and has chance for publication, but revisions are required. Please update your manuscript in accordance with the Reviewers/Editors comments (in MS Word file format) adheres every detail of the guide of authors (http://iaescore.com/gfa/telkomnika.docx), and check it for spelling/grammatical mistakes. Describe the revisions to your manuscript in your response letter followed by point-by-point responses to the comments raised. The Editors will check whether the revision already address the reviewers' comments. Failing to do proper revision may lead to the rejection of your paper.

Please submit your revised manuscript within 8 weeks by uploading to our online system (at same paper ID, as author version under our decision)

I look forward for your response

Sincerely yours,
Prof. Dr. Tole Sutikno
Universitas Ahmad Dahlan
telkomnika@ee.uad.ac.id

:: M.Eng. (S2) in Electrical Engineering (Intelligent Systems Engineering)
::

Want to study Master's Program of Electrical Engineering at Universitas Ahmad Dahlan, Yogyakarta?

We are pleased to announce that the Indonesia Ministry of Education, Culture, Research, and Technology has approved (https://s.id/1G0RL) the establishment of the Master of Electrical Engineering in Intelligent Systems Engineering (ISE) Program at Universitas Ahmad Dahlan. In our Master in Intelligent Systems Engineering, you will gain technical expertise in computational engineering and artificial intelligence—the future of engineering and problem solving. Please take a look at the attached flyer #1

(https://s.id/1G0O8) or flyer #2 (https://s.id/1G0On).

::::::::::::::::::::: EDITOR COMMENTS :::::::::::::::::::::

Your current paper version is not satisfactory yet for publication. Please read carefully and prepare your paper as instructed below. The paper that does not follow the instruction and failing to do proper revision may lead to the rejection of your paper.

1). PLEASE ADHERE STRICTLY to THE GUIDE OF AUTHORS http://iaescore.com/gfa/telkomnika.docx, and the checklist for preparing your FINAL paper for publication: http://telkomnika.uad.ac.id/index.php/TELKOMNIKA/about/editorialPolicies#custom-2

2). It is mandatory to present a final paper with article structure in "IMRADC style":
1. INTRODUCTION
2. The Proposed Method/Algorithm/Procedure specifically designed (optional)
3. METHOD
4. RESULTS AND DISCUSSION
5. CONCLUSION
See http://iaescore.com/gfa/telkomnika.docx

3). Please ensure that all references have been cited in your text. Use a tool such as EndNote, Mendeley, or Zotero for reference management and formatting, and choose IEEE style. Each citation should be written in the order of appearance in the text in square brackets. For example, the first citation [1], the second citation [2], and the third and fourth citations [3], [4]. When citing multiple sources at once, the preferred method is to list each number separately, in its own brackets, using a comma or dash between numbers, as such: [1], [3], [5]. It is not necessary to mention an author's name, pages used, or date of publication in the in-text citation. Instead, refer to the source with a number in a square bracket, e.g. [9], that will then correspond to the full citation in your reference list.
Examples of in-text citations:
This theory was first put forward in 1970 [9].
Zadeh [10] has argued that ...
Several recent studies [7], [9], [11]-[15] have suggested that....
... end of the line for my research [16].

4). Please present all references as complete as possible and use IEEE style (include information of DOIs, volume, number, pages, etc). If it is available, DOI information is mandatory!! See http://iaescore.com/gfa/telkomnika.docx

5). Complete the biographies of authors as our template (include link to the authors' profiles, do not delete any icons in the template). It is mandatory!! See http://iaescore.com/gfa/telkomnika.docx

6). Use different PATTERNS for presenting different results in your figures/graphics (instead of different colors). It is mandatory!! See http://iaescore.com/gfa/telkomnika.docx

7). Prepare all tables as our template (in MS Word file format and editable, NOT as figure)

Reviewer A:

Ethics (Plagiarism, Fraud, Other ethical concerns)
- If you suspect that an article is a substantial copy of another work, please let the editor know, citing the previous work in as much detail as possible.
- It is very difficult to detect the determined fraudster, but if you suspect the results in an article to be untrue, inform it
- Has there been a violation of the accepted norms in the ethical?
Please provide your detailed comments to the Author(s) on the following.:

Does the title of the paper accurately reflect the major focus contribution of this paper?:
Yes

If No, please suggest change of the title as appropriate within 10 words:

Is the abstract an appropriate and adequate digest of the work?:
Yes

Is the paper clear, concise, and well organized?:
No

Structure and Presentation (Layout and format, Title, Abstract, Introduction, Method, Results and Discussion, Conclusion, Language). Please provide your detailed comments to the Author(s) on the following:

Do authors place the paper in proper context by citing relevant papers? Is the paper free from obvious errors, misconceptions, or ambiguity? Is the paper written in correct English? Please note grammatical errors and suggest corrections:

Previous Research: If the article builds upon previous research does it reference that work appropriately? Are there any important works that have been omitted? Are the references accurate? Do authors place the paper in proper context by citing relevant papers? Please provide your detailed comments to the Author(s) on the following.:

Rate of the contribution strength to the field and the scientific quality is represented in this paper?:
Below average

Please mark appropriate scale for the overall grade for this paper?:
Below average

Reviewer's comments and suggestions how to improve the paper. (If it is not possible, kindly please use separate sheets or a copy of the paper for comments and suggestions for revision. Indicate whether revisions are mandatory or suggested. Please use word processing type format if possible, and then upload in next step or submit via email: telkomnika@uad.ac.id)

URGENT: Authors should consider all above items for preparing their revision (and NOT only the below comments)
:

1. I found that the citations in your paper are not done sequentially. I suggest numbering the citations consecutively according to the first mention of each source in the text (sequential order). The first citation in your text should be [1], the second should be [2], the third should be [3], and so on.
2. The references you used for your paper do not use the format suggested. I suggest to the author to readjust the reference writing in your paper to the

IEEE format, you can learn more at the following link:
https://ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE-Reference-Guide.pdf
3. I found some references [25] and [30] that were not referenced in the body text/not cited. I suggest to the author to cite every existing reference or delete uncited references.
4. In your figures section, I suggest providing an explanation of the vertical and horizontal axis to make it easier for the reader to understand what you mean.
5. I found that some paragraphs in your paper consist of less than 3 sentences. I suggest to the writer to adjust the paragraphs to at least 3 sentences.
6. In your acknowledgments section, I recommend including the contract number of your research (if any).

Reviewer B:

Ethics (Plagiarism, Fraud, Other ethical concerns)
- If you suspect that an article is a substantial copy of another work, please let the editor know, citing the previous work in as much detail as possible.
- It is very difficult to detect the determined fraudster, but if you suspect the results in an article to be untrue, inform it
- Has there been a violation of the accepted norms in the ethical?
Please provide your detailed comments to the Author(s) on the following.:

Does the title of the paper accurately reflect the major focus contribution of this paper?:
Yes

If No, please suggest change of the title as appropriate within 10 words:

Is the abstract an appropriate and adequate digest of the work?:
Yes

Is the paper clear, concise, and well organized?:
No

Structure and Presentation (Layout and format, Title, Abstract, Introduction, Method, Results and Discussion, Conclusion, Language). Please provide your detailed comments to the Author(s) on the following:

Do authors place the paper in proper context by citing relevant papers? Is the paper free from obvious errors, misconceptions, or ambiguity? Is the paper written in correct English? Please note grammatical errors and suggest corrections:

Previous Research: If the article builds upon previous research does it reference that work appropriately? Are there any important works that have been omitted? Are the references accurate? Do authors place the paper in proper context by citing relevant papers? Please provide your detailed comments to the Author(s) on the following.:

Rate of the contribution strength to the field and the scientific quality is represented in this paper?:
Good

Please mark appropriate scale for the overall grade for this paper?:

Good

Reviewer's comments and suggestions how to improve the paper. (If it is not possible, kindly please use separate sheets or a copy of the paper for comments and suggestions for revision. Indicate whether revisions are mandatory or suggested. Please use word processing type format if possible, and then upload in next step or submit via email: telkomnika@uad.ac.id)

URGENT: Authors should consider all above items for preparing their revision (and NOT only the below comments)
:

- Missing mention number citation [25] and [30]

-Overall, this paper is well written and informative. The abstract contains research, objectives and results. Likewise, the conclusion is conveyed well..
The paper is also equipped with well-explained tables and figures

---

Bukti Pembayaran

Aplikasi setoran/Transfer/Kliring/Inkaso
Deposit / Transfer / Clearing / Collection Application

BSI BANK SYARIAH INDONESIA

Bismillahirrahmanirrahiim

Kepada *to* **PT Bank Syariah Indonesia**

harap dilakukan transaksi berikut *Please do this transaction:*

Cabang *Branch* _____ No.: 349276

Tanggal *Date* _____

Jenis Transaksi *Transaction*: ☐ Setoran *Deposit* ☐ Pemindahbukuan *Overbooking* ☐ RTGS *RTGS* ☐ Kliring-inkaso *clearing-collection* ☐ SWIFT *SWIFT* ☐ Wesel *Demand Draft* ☐ Lainnya *Others* _____

harap ditulis dengan huruf cetak *fill in with block letters*

VALIDASI *Validation*

PENGIRIM: ADINDA AYU PUSPITA RAMADHANI

KE REKENING: 6410950070

PENERIMA: TOLE SUTIKNO

BANK PENERIMA: Bank Syariah Indonesia

JUMLAH: Rp. 5.400.000,00

BIAYA: Rp. 0,00

BERITA: PEMBAYARAN JURNAL PAPER ID 25707

Tel: (Pemohon/Applicant)

PENERIMA *Beneficiary*: ☐ Penduduk *Resident* ☐ Bukan penduduk *Non-Resident*

Nama *Name*: TOLE SUTIKNO

Nomor rekening *Account Number*: 6410950070

Bank *Bank*: BANK SYARIAH INDONESIA

Alamat & Nomor Telepon *Address & Telephone Number*: YOGYAKARTA

PENGIRIM *Applicant*: ☐ Penduduk *Resident* ☐ Bukan Penduduk *Non-Resident* ☐ Nasabah *Customer* ☐ Bukan Nasabah *Walk in Customer*

Nama *Name*: ADINDA AYU PUSPITA RAMADHANI

Nomor Identitas *Identity Number*: 6 3 7 2 0 2 4 7 1 2 0 1 0 0 0 6

Alamat & Nomor Telepon *Address & Telephone Number*: _____

Mata Uang *Currency*: ☑ Rupiah *Local Currency* ☐ Valuta Asing *Foreign Currency* ☐☐☐

SUMBER DANA TRANSAKSI *Source of Fund*: ☐ Tunai *Cash* ☐ Cek/Bilyet giro *Cheque* ☐ Debet Rekening: *Debet Account:* ☐☐☐☐☐☐☐☐☐☐

BERITA UNTUK PENERIMA *Message for Beneficiary*

Bank Tertarik *Drawee Bank*	Nomor cek/BG *Cheque Number*	Valuta *Currency*	Nominal *Amount*

Jumlah *Amount*: Rp 5.400.000

Terbilang *In Words*: LIMA JUTA EMPAT RATUS RIBU RUPIAH

diisi oleh Bank *filled out by bank*

Jumlah Transfer *Amount of Transfer*	
Komisi *Commission*	
Biaya Pengiriman (SWIFT/RTGS/SKN)	
Biaya Koresponden *Correspondent Charge*	
Sub Total	
Kurs *Rate*	
Total	

BIAYA TRANSAKSI *Handling Charge*: ☐ Tunai *Cash* ☐ Debet Rekening: *Debet Account:* ☐☐☐☐☐☐☐☐☐☐

Biaya Bank Koresponden *Correspondent Charge*: ☐ Pengirim *Applicant* ☐ Penerima *Beneficiary* ☐ Lainnya *Others* _____

Pemohon telah membaca serta memahami, dan dengan ini menyetujui sepenuhnya syarat-syarat dan ketentuan yang tercantum dibalik formulir transaksi ini *Applicant have read and understand, and with this applicant uncondition ally accept all term and condition on the reserve of this transaction form.*

Pengesahan Bank *Bank's Authorization*	Tanda tangan pemohon *Applicant's Signature*

TUJUAN TRANSAKSI *Underlying Transaction*: PEMBAYARAN JURNAL PAPER ID #25707

Perhatian : Apabila penerima adalah bukan Nasabah/ ***Walk in Customer*** maka wajib mengisi alamat dan nomor telepon penerima.

AG 001

Lembar 1 :untuk Teller

Naskah Jurnal

TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 99, No. 1, Month 2099, pp. 1~1x
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v99i1.paperID

The impact of software metrics in NASA Metric Data Program dataset modules for software defect prediction

Adinda Ayu Puspita Ramadhani, Radityo Adi Nugroho, Mohammad Reza Faisal, Friska Abadi, Rudy Herteno
Department of Computer Science, Faculty of Mathematics and Natural Science, Lambung Mangkurat University, Banjarbaru, Indonesia

Article Info

Article history:

Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy

Keywords:

Software Defect Prediction
Software Defect Metrics
NASA MDP
k-Nearest Neighbor

ABSTRACT

This paper discusses software metrics and their impact on software defect prediction values in the NASA Metric Data Program (MDP) dataset. The NASA MDP dataset consists of four categories of software metrics: Halstead, McCabe, LoC, and Misc. However, there is no study showing which metrics participate in increasing the AUC value of the NASA MDP dataset. This study utilizes 12 modules from the NASA MDP dataset, where these 12 modules are being tested into 14 relationships of software metrics derived from the four existing metric categories. Subsequently, classification is performed using the kNN method. The research concludes that software metrics have a significant impact on the AUC value, with the LoC + McCabe + Misc metrics relationship influencing the improvement of the AUC value. However, the metrics relationship that has the most impact on achieving less optimal AUC values is McCabe. Halstead metric also plays a role in decreasing the performance of other metrics.

Corresponding Author:

Radityo Adi Nugroho
Department of Computer Science, Faculty of Mathematics and Natural Science
Lambung Mangkurat University
A. Yani St., KM 36, Banjarbaru, South Kalimantan, Indonesia
Email: radityo.adi@ulm.ac.id

1. INTRODUCTION

Over the past two decades, there has been a rapid increase in the demand for software development in all sectors [1]. Along with this, the quality of the software becomes extremely important, particularly where software defects are concerned [2]. A software defect is generally defined as a deviation from the specifications or requirements that should be possessed by software, this defect can cause the software to fail the functions that are typically reported during the software testing phase [3], [4]. In 2018, it was reported that the identification and refinement of software defects and losses due to software failures in production accounted for 50% of the total cost of software development [5]. It has also been reported that a minor error in data conversion caused a $125 million NASA spacecraft to disappear in space [6]. All of these reports have made research into software defects and software defect prediction one of the most extensive in

recent years. Software Defect Prediction (SDP) itself is a classification process performed to detect code or modules of software that are likely to fail or have defects. This process is implemented using previously available defect data [7].

Since the beginning of 2000, software defect prediction studies have widely used data from the NASA repository. This repository is known as NASA MDP. NASA MDP become popular among researchers because its original format is easy to use in research of Software Defect Prediction, moreover, the difficulty of obtaining software defect data has made public repositories of software defect data very popular [8], [9]. The NASA MDP dataset consists of 12 pre-cleaned module datasets, and every dataset module represents a system/subsystem of NASA's software. [10], [11]. Each module in the NASA MDP dataset contains a set of static software metrics which used to determine the presence of defects in the software module [12], [13]. Software metrics themselves are widely used in software defect prediction studies and have a crucial role in the construction of SDP models as one of the indicators for the measurement of software quality and the estimation of the presence of defects and vulnerabilities in a code [14], [15]. Some studies that use software metrics already conducted include [16]-[21].

In research using software metrics as indicators, it is important to have an understanding of software metrics. The understanding of metrics relies on the concepts of the metric itself as well as its relations to other metrics, each metrics relationship will have an impact on each other [21]. Several research studying the relationships between metrics have already been conducted, such as a comparison between static metrics and dynamic weighted metrics performed by [22], the analysis of metrics on OOS to evaluate and identify the relationship between metrics value with software security performed by [23] and study about the relationship between two types of metrics, namely object-oriented metrics with procedural-oriented metrics performed by [24]. Although research on the metrics relationship has been widely conducted, most studies do not provide clear evidence on how a metric can influence other metrics, or which metrics have the most significant impact on the relationships between them [21]. Moreover, it is not known whether all existing metric attributes are relevant in Software Defect Prediction [13].

Based on the explanation above, this research will study the influence of software metrics on the accuracy of software defect prediction in the kNN classification method. kNN is chosen to be the classification method because it is one of the methods widely used in various studies of Software Defect Prediction, as conducted by [26]-[29].

2. METHOD

2.1 Collecting Data

The dataset utilized in this study originates from the NASA repository known as the NASA Metric Data Program (MDP). This particular dataset was selected due to its widespread use in software metrics research and its explicit design for research in this area [9], [29]. NASA MDP provides two versions of clean dataset, namely D' (which includes duplicate and inconsistent instances) and D'' (which excludes duplicate and inconsistent instances) [12]. The NASA MDP dataset version used in this study is D'' which is taken from (https://github.com/klainfo/NASADefectDataset). Amount of data for each dataset shown in Table 1.

Table 1. NASA MDP D'' Datasets [11]

Dataset	Attributes	Modules	Defective	Non-Defective	Defective (%)
CM1	38	327	42	285	12.8
JM1	22	7,720	1,612	6,108	20.8
KC1	22	1,162	294	868	25.3
KC2	40	194	36	158	18.5
MC1	39	1,952	36	1,916	1.8
MC2	40	124	44	80	35.4
MW1	38	250	25	225	10
PC1	38	679	55	624	8.1
PC2	37	722	16	706	2.2
PC3	38	1,053	130	923	12.3

PC4	38	1,270	176	1,094	13.8
PC5	39	1,694	458	1,236	27.0

2.2 Software Metrics

Certain studies suggest that software defect prediction models formulated based on software metrics are effective in predicting defects in source code [31]. The datasets module in NASA MDP are formed by 4 software metric categories, namely LoC, Halstead, McCabe, and Misc [32]. In Table 2 it shown 4 software metric categories with features in each category.

2.3 Preprocessing Data

Data preprocessing techniques are often employed to minimize the interference of raw data, including noisy data, and transform it into a form that is more understandable [33], [34]. During the preprocessing stage, pivotal procedures encompass feature selection and sampling methods [35]. Feature selection is one of the steps in data preprocessing where relevant features are selected in the learning model [36].

In this research, preprocessing data is conducted by performing feature selection based on 4 software metrics categories available in NASA MDP dataset modules. These 4 software metrics categories are paired into the following pairs which shown in Table 3.

Table 2. Software Metrics in NASA MDP [32]

Metrics Category	Features	Number of Features
LoC	CYCLOMATIC_COMPLEXITY CYCLOMATIC_DENSITY DESIGN_COMPLEXITY ESSENTIAL_COMPLEXITY	6
Misc	BRANCH_COUNT CALL_PAIRS CONDITION_COUNT DECISION_COUNT DECISION_DENSITY DESIGN_DENSITY EDGE_COUNT ESSENTIAL_DENSITY PARAMETER_COUNT GLOBAL_DATA_COMPLEXITY GLOBAL_DATA_DENSITY MAINTENANCE_SEVERITY MODIFIED_CONDITION_COUNT MULTIPLE_CONDITION_COUNT NODE_COUNT NORMAL_CYLOMATIC_COMPLEXITY PERCENT_COMMENTS	17
McCabe	LOC_BLANK LOC_CODE_AND_COMMENT LOC_COMMENTS LOC_EXECUTABLE NUMBER_OF_LINES LOC_TOTAL	6
Halstead	HALSTEAD_CONTENT HALSTEAD_DIFFICULTY HALSTEAD_EFFORT HALSTEAD_ERROR_EST HALSTEAD_PROG_TIME HALSTEAD_VOLUME NUM_OPERANDS NUM_OPERATORS NUM_UNIQUE_OPERANDS NUM_UNIQUE_OPERATORS HALSTEAD_LENGTH HALSTEAD_LEVEL	12

Table 3. Metrics Pairs Used in the Study

Metric Quantity	Metrics Pairs
1 Metric	Halstead LoC McCabe Misc
2 Metrics	Halstead+LoC Halstead+McCabe Halstead+Misc LoC+McCabe LoC+Misc McCabe+Misc
3 Metrics	LoC+McCabe+Misc Halstead+McCabe+Misc Halstead+LoC+Misc Halstead+LoC+McCabe

2.4 Classification

k-Nearest Neighbor (kNN) is a classification method based on supervised learning [37]. The fundamental concept of kNN is to classify training data into specific classes through majority voting based on the k value, the k value itself is a constant value defined by the user, and the class that represents most of the training data will be assigned to the test pattern [38], [39]. In this study, the k values used are k=3, k=5, k=7, k=9, k=11, k=13, and k=15. Before classification with kNN is performed, the dataset is divided into training data and testing data using 10-fold Cross-Validation.

2.5 Evaluation

To evaluate the results obtained in this study, we will compare the area under the curve (AUC) values for each metric relationship within the same dataset module with those of other metric relationships. The highest AUC value for each k value in a metric relationship will be selected and then compared to the highest value held by other metric relationships. The AUC values will also be compared with AUC values from the dataset modules containing all metrics.

3. RESULTS AND DISCUSSION

The classification method used in this study is limited to k-Nearest Neighbor (kNN) along with the NASA MDP dataset from the NASA repository, consisting of 12 dataset modules. The testing process is performed on each dataset module, both on the original dataset and on the relationships between each category of software metrics within the NASA MDP dataset. On each dataset module, testing is carried out with several values of k, namely k=3, k=5, k=7, k=9, k=11, k=13, and k=15. This testing phase is conducted to determine the most optimal AUC value for each module dataset and to understand the impact of the relationships between metrics.

As seen in the testing results in Table 4, each pair of metrics for each dataset module has an impact on the AUC value of a dataset. Some metric pairs influence increasing the AUC value. However, some metric pairs result in values that are not better than the AUC value of the dataset composed of all metric categories. The highest value for each dataset module is shown in Fig. 1.

Table 4. Testing Result

Data set	k	All Metrics	Halstead	LoC	McCabe	Misc	Halstead + LoC	Halstead + McCabe	Halstead + Misc	LoC + McCabe	LoC + Misc	McCabe + Misc	LoC + McCabe + Misc	Halstead + McCabe + Misc	Halstead + LoC + Misc	Halstead + LoC + McCabe
CM1	3	0.56	0.56	0.676	0.498	0.505	0.558	0.56	0.559	0.672	0.676	0.518	0.656	0.559	0.56	0.558
	5	0.605	0.605	0.695	0.482	0.568	0.604	0.605	0.606	0.693	0.691	0.542	0.685	0.606	0.606	0.604
	7	0.62	0.616	0.72	0.484	0.57	0.616	0.616	0.617	0.724	0.697	0.561	0.697	0.617	0.62	0.616
	9	0.617	0.614	0.726	0.516	0.587	0.616	0.614	0.616	0.719	0.711	0.583	0.708	0.616	0.617	0.616
	11	0.634	0.633	0.723	0.524	0.588	0.633	0.633	0.633	0.711	0.692	0.603	0.69	0.633	0.634	0.633
	13	0.668	0.664	0.73	0.494	0.595	0.666	0.664	0.667	0.735	0.7	0.617	0.697	0.667	0.669	0.666
	15	0.657	0.651	0.739	0.514	0.607	0.652	0.651	0.654	0.738	0.71	0.612	0.709	0.654	0.657	0.652
JM1	3	0.574	0.574	0.596	0.51	0.507	0.573	0.572	0.573	0.628	0.616	0.511	0.629	0.574	0.574	0.573
	5	0.591	0.59	0.615	0.53	0.526	0.591	0.588	0.589	0.648	0.633	0.533	0.652	0.59	0.591	0.591
	7	0.603	0.602	0.628	0.548	0.544	0.604	0.603	0.604	0.659	0.646	0.55	0.664	0.604	0.604	0.604
	9	0.61	0.607	0.639	0.558	0.556	0.611	0.607	0.607	0.666	0.653	0.558	0.669	0.607	0.611	0.611
	11	0.618	0.615	0.647	0.566	0.568	0.618	0.615	0.616	0.672	0.66	0.569	0.671	0.615	0.619	0.618
	13	0.624	0.621	0.647	0.574	0.569	0.624	0.622	0.622	0.672	0.665	0.575	0.676	0.622	0.624	0.624
	15	0.629	0.625	0.65	0.581	0.572	0.628	0.652	0.625	0.676	0.67	0.58	0.679	0.625	0.629	0.628
KC1	3	0.524	0.525	0.547	0.533	0.525	0.528	0.52	0.519	0.543	0.582	0.533	0.56	0.526	0.543	0.668
	5	0.587	0.591	0.591	0.57	0.551	0.597	0.593	0.59	0.608	0.621	0.575	0.614	0.595	0.59	0.598
	7	0.617	0.618	0.602	0.621	0.598	0.617	0.619	0.619	0.623	0.63	0.619	0.634	0.622	0.616	0.617
	9	0.628	0.625	0.611	0.617	0.621	0.626	0.626	0.624	0.632	0.637	0.61	0.635	0.627	0.628	0.628

	11	0.639	0.642	0.609	0.623	0.612	0.639	0.641	0.641	0.634	0.634	0.616	0.634	0.641	0.639	0.64
	13	0.651	0.649	0.621	0.624	0.616	0.65	0.649	0.649	0.636	0.637	0.619	0.645	0.65	0.65	0.649
	15	0.652	0.647	0.617	0.616	0.609	0.652	0.648	0.648	0.636	0.646	0.612	0.642	0.648	0.652	0.652
KC3	3	0.439	0.468	0.49	0.391	0.574	0.452	0.48	0.471	0.469	0.464	0.544	0.455	0.513	0.483	0.439
	5	0.515	0.497	0.554	0.552	0.604	0.54	0.504	0.537	0.524	0.55	0.612	0.572	0.499	0.532	0.55
	7	0.579	0.571	0.648	0.639	0.705	0.575	0.594	0.578	0.658	0.656	0.674	0.671	0.59	0.591	0.587
	9	0.585	0.591	0.668	0.647	0.687	0.598	0.589	0.586	0.669	0.692	0.691	0.702	0.606	0.583	0.604
	11	0.618	0.62	0.705	0.663	0.694	0.61	0.626	0.614	0.66	0.732	0.7	0.719	0.615	0.616	0.613
	13	0.637	0.641	0.691	0.663	0.713	0.641	0.643	0.638	0.648	0.718	0.717	0.703	0.645	0.637	0.641
	15	0.645	0.645	0.676	0.671	0.699	0.642	0.641	0.644	0.655	0.748	0.72	0.748	0.639	0.642	0.644
MC1	3	0.671	0.628	0.778	0.613	0.723	0.661	0.638	0.651	0.779	0.795	0.723	0.796	0.651	0.671	0.661
	5	0.681	0.632	0.769	0.667	0.796	0.673	0.634	0.67	0.779	0.815	0.782	0.804	0.67	0.681	0.672
	7	0.701	0.653	0.768	0.696	0.803	0.682	0.655	0.699	0.773	0.806	0.81	0.806	0.699	0.701	0.682
	9	0.709	0.66	0.759	0.681	0.825	0.689	0.672	0.706	0.764	0.8	0.823	0.8	0.706	0.709	0.689
	11	0.728	0.689	0.752	0.695	0.823	0.706	0.701	0.726	0.758	0.792	0.826	0.792	0.726	0.728	0.706
	13	0.733	0.703	0.755	0.691	0.838	0.709	0.706	0.73	0.766	0.786	0.825	0.787	0.731	0.733	0.709
	15	0.745	0.712	0.759	0.722	0.842	0.719	0.716	0.743	0.766	0.781	0.828	0.791	0.743	0.745	0.719
MC2	3	0.628	0.625	0.674	0.647	0.681	0.624	0.625	0.627	0.669	0.715	0.67	0.717	0.627	0.628	0.624
	5	0.631	0.629	0.684	0.624	0.711	0.629	0.629	0.629	0.686	0.717	0.714	0.702	0.629	0.631	0.629
	7	0.62	0.62	0.642	0.607	0.654	0.62	0.62	0.62	0.661	0.724	0.666	0.735	0.62	0.62	0.62
	9	0.616	0.616	0.668	0.591	0.676	0.616	0.616	0.616	0.674	0.717	0.659	0.714	0.616	0.616	0.616
	11	0.635	0.635	0.642	0.612	0.689	0.635	0.635	0.635	0.646	0.722	0.693	0.728	0.635	0.635	0.635
	13	0.618	0.618	0.654	0.644	0.662	0.618	0.618	0.618	0.656	0.701	0.653	0.696	0.618	0.618	0.618
	15	0.612	0.612	0.668	0.615	0.648	0.612	0.612	0.612	0.669	0.699	0.653	0.702	0.612	0.612	0.612
MW 1	3	0.512	0.508	0.575	0.66	0.66	0.508	0.508	0.512	0.583	0.635	0.648	0.628	0.512	0.512	0.508
	5	0.523	0.522	0.639	0.632	0.681	0.522	0.522	0.526	0.709	0.698	0.682	0.691	0.526	0.523	0.522
	7	0.56	0.539	0.678	0.694	0.671	0.539	0.539	0.559	0.693	0.704	0.671	0.702	0.559	0.56	0.539
	9	0.598	0.598	0.676	0.671	0.703	0.599	0.598	0.598	0.687	0.688	0.709	0.684	0.598	0.598	0.599
	11	0.589	0.589	0.678	0.671	0.721	0.589	0.589	0.589	0.675	0.695	0.733	0.697	0.589	0.589	0.589
	13	0.599	0.596	0.677	0.667	0.723	0.596	0.596	0.599	0.677	0.701	0.74	0.703	0.599	0.599	0.596
	15	0.6	0.6	0.661	0.682	0.732	0.6	0.6	0.6	0.668	0.714	0.736	0.71	0.6	0.6	0.6
PC1	3	0.507	0.507	0.71	0.63	0.662	0.508	0.507	0.508	0.713	0.723	0.672	0.707	0.508	0.507	0.508
	5	0.519	0.525	0.755	0.637	0.689	0.524	0.525	0.519	0.768	0.783	0.703	0.789	0.519	0.519	0.524
	7	0.533	0.534	0.785	0.645	0.703	0.534	0.533	0.533	0.791	0.816	0.716	0.822	0.533	0.533	0.534
	9	0.554	0.551	0.8	0.671	0.73	0.553	0.551	0.552	0.791	0.821	0.717	0.821	0.551	0.554	0.553
	11	0.566	0.568	0.804	0.653	0.72	0.566	0.568	0.567	0.819	0.814	0.727	0.814	0.567	0.566	0.566
	13	0.572	0.571	0.812	0.642	0.701	0.573	0.571	0.571	0.818	0.826	0.708	0.822	0.571	0.572	0.573
	15	0.591	0.591	0.806	0.637	0.699	0.592	0.591	0.591	0.816	0.849	0.695	0.847	0.591	0.591	0.592
PC2	3	0.519	0.519	0.72	0.466	0.473	0.52	0.519	0.519	0.703	0.523	0.471	0.499	0.519	0.519	0.52
	5	0.507	0.507	0.729	0.47	0.476	0.507	0.507	0.507	0.714	0.509	0.451	0.512	0.507	0.507	0.507
	7	0.517	0.517	0.728	0.519	0.523	0.517	0.517	0.517	0.747	0.55	0.522	0.568	0.517	0.517	0.517
	9	0.508	0.508	0.724	0.5	0.509	0.508	0.508	0.508	0.745	0.579	0.525	0.582	0.508	0.508	0.508
	11	0.494	0.494	0.719	0.482	0.56	0.494	0.494	0.494	0.752	0.66	0.515	0.673	0.494	0.494	0.494
	13	0.592	0.592	0.77	0.464	0.597	0.592	0.592	0.591	0.792	0.681	0.574	0.684	0.591	0.592	0.592
	15	0.642	0.643	0.787	0.474	0.614	0.642	0.643	0.641	0.792	0.679	0.637	0.677	0.641	0.642	0.642
PC3	3	0.613	0.609	0.681	0.566	0.654	0.614	0.608	0.613	0.708	0.661	0.65	0.665	0.613	0.613	0.615
	5	0.658	0.655	0.719	0.585	0.682	0.654	0.75	0.657	0.739	0.697	0.693	0.706	0.657	0.658	0.654
	7	0.666	0.665	0.741	0.607	0.689	0.665	0.665	0.661	0.764	0.715	0.704	0.721	0.661	0.666	0.665

	9	0.676	0.671	0.744	0.621	0.719	0.671	0.671	0.674	0.769	0.732	0.728	0.741	0.674	0.676	0.671
	11	0.674	0.667	0.746	0.651	0.721	0.673	0.667	0.673	0.763	0.733	0.738	0.746	0.673	0.674	0.673
	13	0.671	0.663	0.748	0.666	0.725	0.668	0.665	0.666	0.758	0.75	0.732	0.762	0.666	0.671	0.668
	15	0.667	0.666	0.751	0.666	0.739	0.665	0.666	0.667	0.759	0.764	0.738	0.768	0.667	0.667	0.665
PC4	3	0.592	0.565	0.775	0.635	0.725	0.567	0.564	0.584	0.809	0.817	0.732	0.813	0.584	0.592	0.568
	5	0.592	0.568	0.806	0.689	0.774	0.574	0.567	0.588	0.829	0.829	0.771	0.838	0.588	0.592	0.575
	7	0.606	0.574	0.824	0.715	0.782	0.579	0.577	0.598	0.843	0.852	0.797	0.852	0.598	0.604	0.579
	9	0.602	0.58	0.828	0.709	0.807	0.579	0.58	0.601	0.836	0.854	0.813	0.852	0.601	0.603	0.579
	11	0.607	0.588	0.827	0.708	0.81	0.583	0.588	0.609	0.838	0.856	0.816	0.858	0.609	0.607	0.583
	13	0.611	0.592	0.826	0.707	0.816	0.591	0.592	0.611	0.839	0.857	0.819	0.859	0.611	0.611	0.591
	15	0.619	0.6	0.826	0.706	0.821	0.602	0.6	0.615	0.84	0.857	0.823	0.857	0.615	0.619	0.602
PC5	3	0.57	0.571	0.623	0.603	0.631	0.577	0.571	0.57	0.639	0.647	0.652	0.654	0.573	0.576	0.577
	5	0.641	0.642	0.665	0.669	0.702	0.646	0.638	0.631	0.68	0.704	0.694	0.692	0.636	0.633	0.643
	7	0.654	0.654	0.693	0.692	0.7	0.654	0.65	0.654	0.704	0.723	0.715	0.712	0.655	0.654	0.65
	9	0.671	0.667	0.704	0.7	0.706	0.668	0.665	0.671	0.717	0.732	0.717	0.732	0.67	0.67	0.668
	11	0.679	0.677	0.707	0.718	0.711	0.679	0.676	0.679	0.719	0.729	0.714	0.73	0.678	0.68	0.679
	13	0.685	0.683	0.709	0.723	0.711	0.684	0.683	0.685	0.721	0.731	0.723	0.733	0.684	0.685	0.685
	15	0.688	0.685	0.715	0.729	0.712	0.687	0.684	0.684	0.723	0.736	0.723	0.737	0.685	0.688	0.687

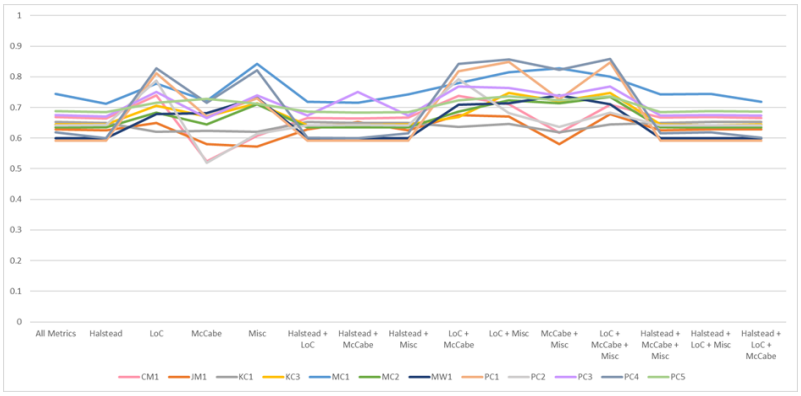

Figure 1. Comparison of Highest Value in Each Metrics Relationship

In Fig. 1 and Fig. 2 the horizontal color line shows each dataset value (where the value shown in vertical axis) for metrics relationship (where the value shown in horizontal axis). In Fig. 1 it shows that the relationships among metrics have a significant impact on the difference in AUC values. The metric relationship that has the most significant role in increasing the AUC value is the relationship between LoC + McCabe + Misc. This metric relationship provides the best AUC value in the module datasets JM1, KC3, MC2, PC4, and PC5. Meanwhile, in module datasets PC1, PC2, and PC3, the highest AUC values are obtained in the LoC + McCabe metric relationship. In the CM1 module dataset, the highest AUC value is associated with the LoC metric. The KC1 module dataset achieves the highest AUC value in the Halstead + LoC + McCabe metric relationship, and the MC2 module dataset obtains the highest AUC value in the McCabe + Misc metric relationship.

Meanwhile, in Fig 2. It is shown that the McCabe metric has been found to play a role in producing less optimal AUC values, as evidenced by the AUC values in the CM1, KC1, KC3, MC2, and PC3 module datasets. On the other hand, the Halstead metric plays a role in decreasing the performance of other metrics

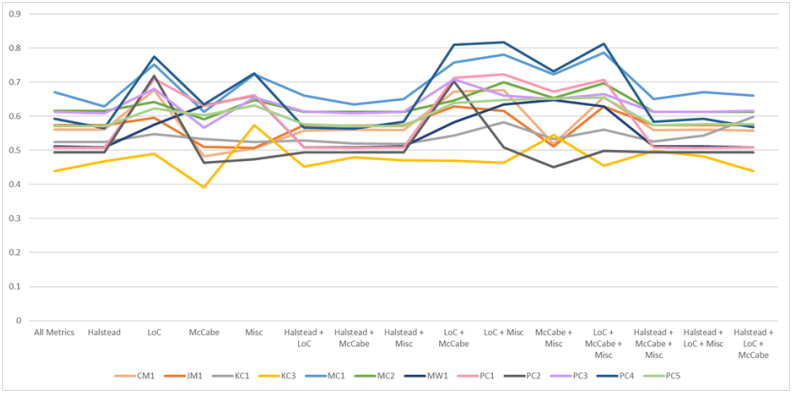

Figure 2. Comparison of Lowest Value in Each Metrics Relationship

4. CONCLUSION

Based on the test results, it can be concluded that the relationship for each software metric in the NASA MDP dataset significantly affects the classification AUC value. Furthermore, the choice of the k value in the kNN classification method also affects the AUC value in each software metrics relationship. The software metrics relationship that has the most impact on achieving a high AUC value is the relationship between the LoC + McCabe + Misc metric categories, however, the metrics relationship that has the most impact on achieving less optimal AUC values is McCabe. The Halstead metric also plays a role in decreasing the performance of other metrics.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Computer Science, Faculty of Mathematics and Natural Science, Lambung Mangkurat University for their support. The completion of this research would not be possible without any help from so many people whose names may not all be enumerated.

REFERENCES

[1] C. Manjula and L. Florence, "Deep neural network based hybrid approach for software defect prediction using software metrics," *Cluster Comput.*, vol. 22, pp. 9847–9863, Jul. 2019, doi: 10.1007/s10586-018-1696-z.

[2] L. Zhou, R. Li, S. Zhang, and H. Wang, "Imbalanced Data Processing Model for Software Defect Prediction," *Wirel. Pers. Commun.*, vol. 102, no. 2, pp. 937–950, Sep. 2018, doi: 10.1007/s11277-017-5117-z.

[3] L. Qiao, X. Li, Q. Umer, and P. Guo, "Deep learning based software defect prediction," *Neurocomputing*, vol. 385, pp. 100–110, Apr. 2020, doi: 10.1016/j.neucom.2019.11.067.

[4] S. Patil and B. Ravindran, "Predicting software defect type using concept-based classification," *Empir. Softw. Eng.*, vol. 25, no. 2, pp. 1341–1378, Mar. 2020, doi: 10.1007/s10664-019-09779-6.

[5] J. Pachouly, S. Ahirrao, K. Kotecha, G. Selvachandran, and A. Abraham, "A systematic literature review on software defect prediction using artificial intelligence: Datasets, Data Validation Methods, Approaches, and Tools," *Eng. Appl. Artif. Intell.*, vol. 111, p. 104773, May 2022, doi: 10.1016/J.ENGAPPAI.2022.104773.

[6] Ö. F. Arar and K. Ayan, "A feature dependent Naive Bayes approach and its application to the software defect prediction problem," *Appl. Soft Comput. J.*, vol. 59, pp. 197–209, Oct. 2017, doi: 10.1016/j.asoc.2017.05.043.

[7] T. Wang, Z. Zhang, X. Jing, and L. Zhang, "Multiple kernel ensemble learning for software defect prediction," *Autom. Softw. Eng.*, vol. 23, no. 4, pp. 569–590, Dec. 2016, doi: 10.1007/s10515-015-0179-1.

[8] J. Petrić, D. Bowes, T. Hall, B. Christianson, and N. Baddoo, "The Jinx on the NASA software defect data sets," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jun. 2016. doi: 10.1145/2915970.2916007.

[9] D. Gray, D. Bowes, N. Davey, Y. Sun, and B. Christianson, "The Misuse of the NASA Metrics Data Program Data Sets for Automated Software Defect Prediction." [Online]. Available: http://mdp.ivv.nasa.gov/

[10] D. Gray, D. Bowes, N. Davey, Y. Sun, and B. Christianson, "Reflections on the NASA MDP data sets," *IET Softw.*, vol. 6, no. 6, pp. 549–558, Dec. 2012, doi: 10.1049/iet-sen.2011.0132.

[11] A. Iqbal and S. Aftab, "A classification framework for software defect prediction using multi-filter feature selection technique and MLP," *Int. J. Mod. Educ. Comput. Sci.*, vol. 12, no. 1, pp. 18–25, 2020, doi: 10.5815/ijmecs.2020.01.03.

[12] D. Gray, D. Bowes, N. Davey, Y. Sun, and B. Christianson, "CCIS 43 - Using the Support Vector Machine as a Classification Method for Software Defect Prediction with Static Code Metrics." [Online]. Available: http://mdp.ivv.nasa.gov/

[13] Khadijah, A. Adorada, P. W. Wirawan, and K. Kurniawan, "The Comparison of Feature Selection Methods in Software Defect Prediction," in *ICICoS 2020 - Proceeding: 4th International Conference on Informatics and Computational Sciences*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/ICICoS51170.2020.9299022.
[14] E. Sreedevi, V. Premalatha, S. Sivakumar, and S. R. Nayak, "A comparative study on new classification algorithm using NASA MDP datasets for software defect detection," *Proc. Int. Conf. Intell. Sustain. Syst. ICISS 2019*, no. Iciss, pp. 312–317, 2019, doi: 10.1109/ISS1.2019.8908096.
[15] N. Medeiros, N. Ivaki, P. Costa, and M. Vieira, "Vulnerable Code Detection Using Software Metrics and Machine Learning," *IEEE Access*, vol. 8, pp. 219174–219198, 2020, doi: 10.1109/ACCESS.2020.3041181.
[16] Y. Özçevik and O. Altay, "MetricHunter: A software metric dataset generator utilizing SourceMonitor upon public GitHub repositories," *SoftwareX*, vol. 23, Jul. 2023, doi: 10.1016/j.softx.2023.101499.
[17] "Specifications table", doi: 10.5281/zenodo.7011334.
[18] M. A. S. Bigonha, K. Ferreira, P. Souza, B. Sousa, M. Januário, and D. Lima, "The usefulness of software metric thresholds for detection of bad smells and fault prediction," *Inf. Softw. Technol.*, vol. 115, no. March 2018, pp. 79–92, 2019, doi: 10.1016/j.infsof.2019.08.005.
[19] Y. Bugayenko *et al.*, "Qualitative Clustering of Software Repositories Based on Software Metrics," *IEEE Access*, vol. 11, pp. 14716–14727, 2023, doi: 10.1109/ACCESS.2023.3244495.
[20] K. Phung, E. Ogunshile, and M. Aydin, "Error-Type - A Novel Set of Software Metrics for Software Fault Prediction," *IEEE Access*, vol. 11, pp. 30562–30574, 2023, doi: 10.1109/ACCESS.2023.3262411.
[21] M. M. A. Dabdawb and B. Mahmood, "On the relations among object-oriented software metrics: A network-based approach," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 901–915, 2021, doi: 10.12785/ijcds/100182.
[22] H. Schnoor and W. Hasselbring, "Comparing static and dynamic weighted software coupling metrics," *Computers*, vol. 9, no. 2, 2020, doi: 10.3390/computers9020024.
[23] K. Kuk, "Object – oriented software metrics in software code vulnerability analysis," in *2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 2020.
[24] L. Prasad and A. Nagar, "Experimental analysis of different metrics (object-oreinted and structural) of software," in *2009 1st International Conference on Computational Intelligence, Communication Systems and Networks, CICSYN 2009*, 2009, pp. 235–240. doi: 10.1109/CICSYN.2009.22.
[25] S. Goyal, *Handling Class-Imbalance with KNN (Neighbourhood) Under-Sampling for Software Defect Prediction*, vol. 55, no. 3. Springer Netherlands, 2022. doi: 10.1007/s10462-021-10044-w.
[26] Q. Yu, S. juan Jiang, R. cun Wang, and H. yang Wang, "A feature selection approach based on a similarity measure for software defect prediction," *Front. Inf. Technol. Electron. Eng.*, vol. 18, no. 11, pp. 1744–1753, Nov. 2017, doi: 10.1631/FITEE.1601322.
[27] W. Wen *et al.*, "Cross-Project Software Defect Prediction Based on Class Code Similarity," *IEEE Access*, vol. 10, pp. 105485–105495, 2022, doi: 10.1109/ACCESS.2022.3211401.
[28] T. T. Khuat and M. H. Le, "Evaluation of Sampling-Based Ensembles of Classifiers on Imbalanced Data for Software Defect Prediction Problems," *SN Comput. Sci.*, vol. 1, no. 2, Mar. 2020, doi: 10.1007/s42979-020-0119-4.
[29] D. Sharma and P. Chandra, "A comparative analysis of soft computing techniques in software fault prediction model development," *Int. J. Inf. Technol.*, vol. 11, no. 1, pp. 37–46, Mar. 2019, doi: 10.1007/s41870-018-0211-3.
[30] J. Murillo-Morera, C. Quesada-López, C. Castro-Herrera, and M. Jenkins, "An Empirical Evaluation of NASA-MDP Data sets using a Genetic Defect-Proneness Prediction Framework," 2016. doi: 10.1109/CONCAPAN.2016.7942359.
[31] R. A. Coelho, F. D. R. N. Guimaraes, and A. A. A. Esmin, "Applying swarm ensemble clustering technique for fault prediction using software metrics," in *Proceedings - 2014 13th International Conference on Machine Learning and Applications, ICMLA 2014*, Institute of Electrical and Electronics Engineers Inc., Feb. 2014, pp. 356–361. doi: 10.1109/ICMLA.2014.63.
[32] R. A. Nugroho, F. Abadi, M. R. Faisal, R. Herteno, and R. Ramadhani, "Metrics Based Feature Selection for Software Defect Prediction," *J. Komputasi*, vol. 8, no. 2, 2020, doi: 10.23960/komputasi.v8i2.2670.
[33] J. Wang, Y. Wang, Z. Li, H. Li, and H. Yang, "A combined framework based on data preprocessing, neural networks and multi-tracker optimizer for wind speed prediction," *Sustain. Energy Technol. Assessments*, vol. 40, Aug. 2020, doi: 10.1016/j.seta.2020.100757.
[34] S. A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, "Data preprocessing in predictive data mining," *Knowl. Eng. Rev.*, vol. 34, 2019, doi: 10.1017/S026988891800036X.
[35] S. K. Pandey and A. K. Tripathi, "An empirical study toward dealing with noise and class imbalance issues in software defect prediction," *Soft Comput.*, vol. 25, no. 21, pp. 13465–13492, Nov. 2021, doi: 10.1007/s00500-021-06096-3.
[36] K. Gao, T. M. Khoshgoftaar, and A. Napolitano, "Combining feature subset selection and data sampling for coping with highly imbalanced software data," in *Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE*, Knowledge Systems Institute Graduate School, 2015, pp. 439–444. doi: 10.18293/SEKE2015-182.
[37] K. P. Sanal Kumar and R. Bhavani, "Human activity recognition in egocentric video using PNN, SVM, kNN and SVM+kNN classifiers," *Cluster Comput.*, vol. 22, pp. 10577–10586, Sep. 2019, doi: 10.1007/s10586-017-1131-x.
[38] F. Bulut and M. F. Amasyali, "Locally adaptive k parameter selection for nearest neighbor classifier: one nearest cluster," *Pattern Anal. Appl.*, vol. 20, no. 2, pp. 415–425, May 2017, doi: 10.1007/s10044-015-0504-0.
[39] J. Gou, Z. Yi, L. Du, and T. Xiong, "A local mean-based k-nearest centroid neighbor classifier," *Comput. J.*, vol. 55, no. 9, pp. 1058–1071, Sep. 2012, doi: 10.1093/comjnl/bxr131.

BIOGRAPHIES OF AUTHORS

	**Adinda Ayu Puspita Ramadhani** is an undergraduate student in the Department of Computer Science, at Lambung Mangkurat University. Her research interest is centered on software defect prediction. She can be contacted at email: dindayupr@gmail.com.
	**Radityo Adi Nugroho** received his bachelor's degree in Informatics from the Islamic University of Indonesia and a master's degree in Computer Science from Gadjah Mada University. Currently, he is an assistant professor in the Department of Computer Science at Lambung Mangkurat University. His research interests include software defect prediction and computer vision. He can be contacted at email: radityo.adi@ulm.ac.id.
	**Mohammad Reza Faisal** received the B.Sc. and M.Eng. degrees in physics and informatics from Bandung Institute of Technology, Bandung, Indonesia, in 2004 and 2013. He also received a B.Eng. degree in informatics from Pasundan University, Bandung, Indonesia, in 2002 and a Ph.D. in computer science from Kanazawa University, Ishikawa, Japan, in 2018. He is currently a lecturer in the Computer Science Department, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University in Banjarbaru, Indonesia. His research interests include artificial intelligence applications, text mining, and software engineering. He can be contacted at email: reza.faisal@ulm.ac.id.
	**Friska Abadi** received his bachelor's degree in computer science from Lambung Mangkurat University, Banjarbaru, Indonesia, in 2011. He also received a master's degree in informatics from STMIK Amikom, Yogyakarta, in 2016. He is currently a lecturer in the Computer Science Department, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarbaru, Indonesia. His research interests include data mining and software engineering. He can be contacted at email: friska.abadi@ulm.ac.id.
	**Rudy Herteno** is currently a lecturer in the Faculty of Mathematics and Natural Science, at Lambung Mangkurat University. He received his bachelor's degree in Computer Science from Lambung Mangkurat University and a master's degree in Informatics from STMIK Amikom University. His research interests include software engineering, software defect prediction, and deep learning. Email: rudy.herteno@ulm.ac.id