

The 5th International Conference on Information Technology for Cyber and IT Service Management **CITSM 2017**



**STIKOM Bali
Convention Hall**

August 8, 2017 - August 10, 2017

**Proceedings of
CITSM 2017**

ISBN : 978-1-5386-2739-6

Organized by :



Partner & Co-Organizers :



2017 5th International Conference on Cyber and IT Service Management

Convention Hall, STIKOM Bali

August 8-10, 2017



ISBN : 978-1-5386-2737-2

IEEE Catalog Number : CFP1737Z-PRT

2017 5th International Conference On Cyber And IT Service Management (CITSM)

Convention Hall, STIKOM Bali

Phone: (0361) 244445

Email : contact.citsm@uinjkt.ac.id

Website : <http://citsm.id/>

August 8-10, 2017

ISBN : 978-1-5386-2737-2

IEEE Catalog Number : CFP1737Z-PRT

2017 5th International Conference On Cyber And IT Service Management (CITSM)

Copyright ©2017 by the Institute of Electrical and Electronics Engineers, Inc. All rights reserved.

Copyright and Reprint Permission

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or reproduction requests should be addressed to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331.

ISBN : 978-1-5386-2737-2

IEEE Catalog Number : CFP1737Z-PRT

Additional copies of this publication are available from

Curran Associates, Inc.

57 Morehouse Lane

Red Hook, NY 12571 USA

+1 845 758 0400

+1 845 758 2633 (FAX)

email: curran@proceedings.com

PREFACE

Assalaamu 'alaykum warahmatullahi wabarakaatuh,

The CITSM 2017 is in the general area of communication and information technology. It provides a forum for presenting and discussing the latest innovations, results and developments in IT Management & organizations, IT Applications, Cyber & IT Security, and ICT. The main objective of this conference is to provide a forum for engineers, academia, scientist, industry, and researchers to present the result of their research activities in the field of Computer and Information Technology. The primary focus of the conference is to create an effective medium for institutions and industries to share ideas, innovations, and problem solving techniques.

There are almost 205 papers submission and only 107 papers are accepted and 101 papers have been registered. Accepted papers will be presented in one of the regular sessions and will be published in the conference proceedings volume. All accepted papers are submitted to IEEEExplore. IEEE Conference Number: # 41401, IEEE Catalog Number: CFP1737Z-PRT, ISBN: 978-1-5386-2737-2, CFP1737Z-USB, ISBN: 978-1-5386-2738-9

On behalf of the CITSM organizers, we wish to extend our warm welcome and would like to thank for the all Keynote Speakers, Reviewers, authors, and Committees, for their effort, guidance, contribution and valuable support. Last but not least, thanks to all lecturers and staffs of the Faculty of Science & Technology, UIN Syarif Hidayatullah Jakarta and STIKOM BALI and other parties that directly and indirectly make this event successful.

Wa billahi taufiq wal hidaayah.

Wallahul muwaffiq ila aqwamit-tharieq.

Wasalaamu 'alaykumu warahmatullahi wabarakaatuh.

Husni Teja Sukmana (Organizing Chair)

Technical Program Committee

TPC Organization:

TPC Chair

- Ismail Khalil, Institute of Telecooperation Johannes Kepler University Linz, Austria

TPC Secretariat

Faculty of Science and Technology, Syarif Hidayatullah State Islamic University Jakarta, Indonesia

TPC Co-Chair:

- I Gede Putu Wirarama Wedashwara Wirawan, The School of Information Management and Computer Engineering (STIKOM) Bali, Indonesia
- Husni Teja Sukmana, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
- Suryadiputra Liawatimena, IEEE Indonesian Section Computer Society Chapter, Bina Nusantara University
- Aries Susanto, Ph.D, Syarif Hidayatullah State Islamic University Jakarta

TPC Member:

- Dwiza Riana, STMIK Nusa Mandiri, Indonesia
- Ankhaa Bayar, National University of Mongolia,
- TB Maulana, Gunadarma University, Indonesia
- Andrew Tanny Liem, Klabat University, Indonesia
- Khamis Alarabi, International Islamic University, Malaysia
- Nashrul Hakiem, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
- Masao Okano, Bunkyo University, Japan
- Adila A. Krisnadhi, Wright state University, USA
- Akram M. Zeki, International Islamic University, Malaysia
- Djoko Soetarno, Coris Research Center, Indonesia
- Afzan Salleh, International Islamic University, Malaysia
- Alfida Hasbullah, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
- Ali A. Alwan Aljuboori, International Islamic University, Malaysia
- Madihah S. Abd. Aziz, International Islamic University, Malaysia
- Taqwa Hariguna, STIMIK AMIKOM Purwokerto, Indonesia
- Bernardo Nugroho Yahya, Ulsan National Institute of Science and Technology, South Korea
- Muharman Lubis, Telkom University, Indonesia
- Kusri, Amikom Jogjakarta University, Indonesia
- Tedjo Darmanto, STIMIK AMIK Bandung, Indonesia
- Muhammad Izman Herdiansyah, Bina Darma University, Indonesia
- Rahmat Widia Sembiring, Medan State Polytechnic, Indonesia
- Sonny Zulhuda, International Islamic University Malaysia, Malaysia
- Murni Mahmud, International Islamic University Malaysia, Malaysia
- Aang Subiyakto, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
- Imam Marzuki Sofi, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
- Winda Astuti, Bina Nusantara University, Indonesia
- Lookman Adebisi, International Islamic University Malaysia, Malaysia
- Agus Rifai, International Islamic University Malaysia, Malaysia
- Sri Hartati, Gajah Mada University, Indonesia

- Rika Rosnelly, Potensi Utama Univerity, Indonesia
- Qomarul Huda, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
- Samsuryadi Sahmin, Sriwijaya University, Indonesia
- Nurhayati Buslim, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
- Diyah Puspitaningrum, Bengkulu University, Indonesia
- Dini Handayani, International Islamic University Malaysia, Malaysia
- Zeeshan Bhatti, University of Sindh, Pakistan
- Heru Agus Santoso, Dian Nuswantoro University, Indonesia
- Affandy, Dian Nuswantoro University, Indonesia
- Iwan Setyawan, IEEE Computer Society, Indonesia
- Raini Hasan, International Islamic University Malaysia, Malaysia
- Meyliana, Bina Nusantara University, Indonesia
- Syopiansyah Jayaputra, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
- Husnayati Hussin, International Islamic University Malaysia, Malaysia
- Roslina Othman, International Islamic University Malaysia, Malaysia
- Zahidah Zulkifli, International Islamic University Malaysia, Malaysia
- Martianus Frederic Ezerman, Nanyang Technology University, Singapore
- Leon Andretti Abdillah, Bina Darma University, Indonesia
- M Ary Heryanto, Dian Nuswantoro University, Indonesia
- Kim Jin Mook, Sunmoon University, South Korea
- Robert P Biuk-Aghai, University of Macau, Taiwan
- Suvdaa Batsuuri, School of engineering and applied sciences, national university of Mongolia, Mongolia
- Houari Sabirin, KDDI Research, Inc, Japan
- Jihye Bae, Sunmoon University, South Korea
- Noor Azura Zakaria, International Islamic University Malaysia, Malaysia
- Abdullah Alkalbani, University of Buraimi, Qatar
- Yudi Agusta, STIKOM Bali, Indonesia
- Ahmad Nurul Fajar, Bina Nusantara University, Indonesia
- Hamwira Yaacob, International Islamic University Malaysia, Malaysia
- Marini Othman, International Islamic University Malaysia, Malaysia
- Mohammad Rasheed, Kuala Lumpur Univerity, Malaysia
- Prihandoko, Gunadarma University, Indonesia
- Jarot Suroso, Bina Nusantara University, Indonesia
- Zaheer Khan, University of the west England, England
- Arief Setyanto, Amikom University, Indonesia
- Muhammad Fachrurrozi, Sriwijaya University, Indonesia
- Shingo Mabu, Yamaguchi University, Japan
- Okfalisa, UIN Riau, Indonesia
- Erna Utami, Amikom University, Indonesia
- Arief Zulianto, Langlanbuana University, Indonesia
- Muhammad Rusli, STIKOM Bali, Indonesia
- Agni Catur Bakti, Samperna University, Indonesia
- Lee Kyong Ohh, Sunmoon University, South Korea
- Hadi Purnawan Satria , Universitas Sriwijaya, Indonesia
- Teddy Mantoro, Samperna University, Indonesia
- Media, Sampoerna University, Indonesia
- Jeong Bae Lee, Busan University of Foreign Study, South Korea
- Fauzan Nurdin, International Islamic University, Malaysia

- Ahmad Nurul Fajar, Bina Nusantara University, Indonesia
- Desmon Alexander, Foresec, Singapore
- Elly Matul Imah, University of Indonesia, Indonesia
- Fauzan Nurdin, International Islamic University Malaysia,
- Ford Lumban Gaol, Bina Nusantara University, Indonesia
- Habib Kassim, PASAS Singapore,
- Ivan Lanovara, Infrastructure University Kuala Lumpur,
- Kuncoro Wastuwibowo, IEEE Indonesia Section
- Marimin, Bogor Agricultural Institute, Indonesia
- Rizal Isnanto, University of Diponegoro, Indonesia
- Sigit Puspito Wigati Jarot, Commissioner, Ministry of Communication & Information Technology
- SM Syed Ali, PASAS Singapore
- Wikan Danar Sunindyo, Bandung Institute of Technology, Indonesia

TABLE OF CONTENT

	Front Matter	ii-iv
	Preface	v
	List of Reviewers	vi-viii
	Table of Content	ix-xix
1	Analysis of Travel Time Computation Accuracy from Crowdsourced Data of Hospitality Application in South of Tangerang City with Estimated Travel Time Method <i>Rizal Broer Bahaweres, Arini, Muhamad Rizka Akbar</i>	1-5
2	Combining of Feature Extraction for Real-time Facial Authentication System <i>I. Intan</i>	6-11
3	Incremental Technique with Set of Frequent Word Item sets for Mining Large Indonesian Text Data <i>Dian Sa'adillah Maylawati, Muhammad Ali Ramdhani, Ali Rahman, Wahyudin Darmalaksana</i>	12-17
4	Crawling and Cluster Hidden Web Using Crawler Framework and Fuzzy-KNN <i>I Gede Surya Rahayuda, Ni Putu Linda Santiari</i>	18-24
5	Internet Service Providers Liability for Third Party Content: Freedom to Operate? <i>Ida Madieha Abdul Ghani Azmi, Suzi Fadhilah Ismail, Mahyuddin Daud</i>	25-29
6	Trust, Risk and Public Key Infrastructure Model on E-Procurement Adoption <i>Herlino Nanang, Ahmad F. Misman, Zahidah Zulkifli</i>	30-35
7	Wiki-MetaSemantik: A Wikipedia-derived Query Expansion Approach based on Network Properties <i>Diyah Puspitaningrum, Gries Yulianti, I.S.W.B. Prasetya</i>	36-41
8	Fast and Efficient Image Watermarking Algorithm using Discrete Tchebichef Transform <i>De Rosal Ignatius Moses Setiadi, T. Sutojo, Eko Hari Rachmawanto, Christy Atika Sari</i>	42-46

9	Diagnosis of Toddler Digestion Disorder using Forward Chaining Method <i>Kasman Rukun, B. Herawan Hayadi, Isra Mouludi, Adyanata Lubis, Safril, Jufri</i>	47-49
10	The Comparation of Text Mining With Naive Bayes Classifier, Nearest Neighbor, and Decision Tree to Detect Indonesian Swear Words on Twitter <i>Wildan Budiawan Zulfikar, Mohamad Irfan, Cecep Nurul Alam, Muhammad Indra</i>	50-54
11	Parallel Evolutionary Association Rule Mining for Efficient Summarization of Wireless Sensor Network Data Pattern <i>Wirarama Wedashwara, Shingo Mabu, Candra Ahmadi</i>	55-60
12	The Implementation of K-Nearest Neighbor Algorithm in Case-Based Reasoning Model for Forming Automatic Answer Identity and Searching Answer Similarity of Algorithm Case <i>Yana Aditia Gerhana, Aldy Rialdy Atmadja, Wildan Budiawan Zulfikar, Nurida Ashanti</i>	61-65
13	The Implementation of E-Learning into Mobile-Based Interactive Data Structure Subject <i>Rismayani, Andi Irmayana</i>	66-70
14	Prototype of Authentication System of Motorcycle Using RFID Implants <i>Marchel Thimoty Tombeng, Haryanto Samuel Laluyan</i>	71-75
15	Implementation of Principal Component Analysis Method for Detection of Chlorine and Bleach in Rice <i>Qadavi Muhammad Sofyan, Arini, Nurul Faizah Rozy</i>	76-80
16	Comparative Study for Better Result on Query Suggestion of Article Searching with MySQL Pattern Matching and Jaccard Similarity <i>Komang Rinatha, Wayan Suryasa</i>	81-84
17	Green Computing Survey Based on User Behavior: A Case Study in Board of Investment and Licensing of Bali Province <i>Luh Gede Surya Kartika, Gede Adhitya Bayu Pramana, I Putu Agus Aditya Satria Wibawa</i>	85-90
18	Explaining Acceptance of E-health Services: An Extension of TAM and Health Belief Model Approach <i>Rinda Wahyuni, Nurbojatmiko</i>	91-97
19	A Review: The Affair of Al-Qur'an and Green Computing <i>Arif Ridho Lubis, Ferry Fachrizal, Halim Maulana</i>	98-102

20	Characteristics Signal Spectrum Analyzer and AWGN with RF Filter Method to reduce interference on the UMTS System	103-108
	<i>Made Adi Surya Antara</i>	
21	Assessing Privacy and Readiness of Electronic Voting System in Indonesia	109-115
	<i>Muharman Lubis, Mira Kartiwi, Yusuf Durachman</i>	
22	Expert System for Predicting the Early Pregnancy with Disorders using Artificial Neural Network	116-121
	<i>Dian Sa'adillah Maylawati, Muhammad Ali Ramdhani, Wildan Budiawan Zulfikar, Ichsan Taufik, Wahyudin Darmalaksana</i>	
23	A Study of Information Technology Infrastructure Library (ITIL) Framework Implementation at the Various Business Field in Indonesia	122-125
	<i>Andrean Limanto, Azqa Fikri Khwarizma, Imelda, Reinert Yosua Rumagit, Victor Prasetya Pietono, Yohanes Halim, Suryadiputra Liawatimena</i>	
24	Boosted Classifier and Features Selection for Enhancing Chronic Kidney Disease Diagnose	126-131
	<i>Made Satria Wibawa, I Made Dendi Maysanjaya, I Made Agus Wirahadi Putra</i>	
25	Improving Information Performance of Schools in Higher Education through IT Service Management	132-137
	<i>Sandy Kosasi, Harjanto Prabowo, Dyah Budiastuti</i>	
26	The Application of Centroid Linkage Hierarchical Method and Hill Climbing Method in Comments Clustering Online Discussion Forum	138-143
	<i>Okfalisa, Joni Iskandar</i>	
27	Numerical Simulation to Design Single Mode Fiber Coupler with Fiber Bragg Grating Combination	144-147
	<i>Saktioto, Rosmeri, Okfalisa, Muhammad Hamdi</i>	
28	Development of Document Plagiarism Detection Software Using Levensthein Distance Algorithm on Android Smartphone	148-153
	<i>Nurhayati, Busman</i>	

29	Routing Protocol RIPng, OSPFv3, and EIGRP on IPv6 for Video Streaming Services <i>Nurhayati, Rahmat Fajar Al Farizky</i>	154-159
30	Framework of Sentiment Annotation for Document Specification in Indonesian Language Base on Topic Modeling and Machine Learning <i>Tata Sutabri, Miftah Ardiansyah</i>	160-165
31	Hybrid Method using HWT-DCT for Image Watermarking <i>Ajib Susanto, De Rosal Ignatius Moses Setiadi, Christy Atika Sari, Eko Hari Rachmawanto</i>	166-170
32	Improving IT Performance through IT Innovation: A Conceptual Model <i>David, Edi Abdurachman, Raymondus Raymond Kosala</i>	171-176
33	Inventory Model of Supply Chain Management 3-Echelon Multi-Tiers <i>Armin Lawi, Nur Ilmiyati Djalal, Aidawayati Rangkuti</i>	177-181
34	Adoption of Information Technology in Business Performance of Small and Medium Enterprises Woven Fabric <i>Susanti Margaretha Kuway, Raymondus Raymond Kosala, Ngatindriatus, Wendy</i>	182-185
35	Toward to Operationalization of Socio-Technical Ontology Engineering Methodology <i>Dana Indra Sensuse, Yudho Giri Sucahyo, Mesnan Silalahi, Ika Arthalia Wulandari, Izzah Fadhilah Akmaliah, Handrie Noprisson</i>	186-192
36	GIS Technology Selection for Visualization of Independent Economic Modeling of Former Woman Migrant Worker (WMW) <i>Kusrini, Muhamad Idris Purwanto, Kusuma Chandra Kirana, Arif Dwi Laksito</i>	193-197
37	Clustering and Profiling of Customers Using RFM for Customer Relationship Management Recommendations <i>Ina Maryani, Dwiza Riana</i>	198-203
38	Contribution of Information Technology through Consumer Engagement to Improve Market Growth of Credit Union <i>Gat, Edi Abdurahman, Stephanus Remond Waworuntu</i>	204-209
39	Delay Analysis of Dynamic Bandwidth Allocation for Triple-Play-Services in EPON	210-215

- 40 Knowledge Management for Creativity Improvement: A Systematic Review 216-223
Pamela Kareen, Dana Indra Sensuse, Elin Cahyaningsih, Handrie Noprisson, Yudho Giri Sucahyo
- 41 Variety and Trends on Geographic Information Systems Research A Literature Study 224-230
Eri Rustamaji
- 42 Decision Support Systems Design on Sharia Financing using Yager's Fuzzy Decision Model 231-234
Aries Susanto, Lisa Latifah, Nuryasin, Aida Fitriyani
- 43 Combining Integrated Sampling Technique with Feature Selection for Software Defect Prediction 235-240
Sukmawati Anggraeni Putri, Frieyadie
- 44 Store Image of Organic Product: Social Responsibility and Trust's Mediator 241-244
Doni Purnama Alamsyah, Oda I. B. Hariyanto
- 45 An Empirical Investigations of User Acceptance of "SCALSA" E-Learning in STIKES Harapan Bangsa Purwokerto 245-250
Hadi Jayusman, Djoko Budiyo Setyohadi
- 46 Strategic Information System Plan for the Implementation of Information Technology at Polytechnic "API" Yogyakarta 251-256
Deny Budiyo, Djoko Budiyo Setyohadi
- 47 Hommons: Hydroponic Management and Monitoring System for an IOT Based NFT Farm Using Web Technology 257-262
Padma Nyoman Crisnapati, I Nyoman Kusuma Wardana, I Komang Agus Ady Aryanto, Agus Hermawan
- 48 'Unsafe' Nutraceuticals Products on the Internet: The Need for Stricter Regulation in Malaysia 263-267
Mahyuddin Daud, Juriah Abd. Jalil, Ida Madieha Abdul Ghani Azmi, Suzi Fadhilah Ismail, Sahida Safuan
- 49 Eye Tracking Analysis of Consumer's Attention to the Product Message of Web Advertisements and TV Commercials 268-272
Masao Okano, Masami Asakawa

50	A Multi-Study Program Recommender System Using ELECTRE Multicriteria Method <i>Linda Marlinda, Yusuf Durachman, Taufik Baidawi, Akmaludin</i>	273-277
51	Comparing RDP and PcolP Protocols for Desktop Virtualization in VMware Environment <i>Louis Casanova, Marcel, Edy Kristianto</i>	278-281
52	Mapping Requirements into E-commerce Adoption Level: A Case Study Indonesia SMEs <i>Evi Triandini, Arif Djunaidy, Daniel Siahaan</i>	282-286
53	Strategic Plan with Enterprise Architecture Planning For Applying Information System at PT. Bestonindo Central Lestari <i>Marianus Omba Riku, Djoko Budiyo Setyohadi</i>	287-292
54	Flow Measurement of Charges and Electricity Costs Monitoring System with Android Based IoT (Case Study: Boarding House Adelina) <i>Nenny Anggraini, Andrew Fiade, Miftahul Fauzan</i>	293-297
55	Concept and Data Model of AK/I Card Digitization as Employment Information Distribution Media <i>Irwan Oyong, Awaludin Abid, Hasnan Afif, Ema Utami</i>	298-303
56	Implementation of TOPSIS Method in the Selection Process of Scholarship Grantee (Case Study: BAZIS South Jakarta) <i>Meinarini Catur Utami, Yuni Sugiarti, Ahmad Melani, Yusuf Durachman, A'ang Subiyakto</i>	304-308
57	Feature Selection Based on Genetic Algorithm, Particle Swarm Optimization and Principal Component Analysis for Opinion Mining Cosmetic Product Review <i>Dinar Ajeng Kristiyanti, Mochamad Wahyudi</i>	309-314
58	Design Concepts Smartcoop with Implementing Financial Technology <i>Adji Sukmana, Mihuandayani, Yayak Kartika Sari, Fuad Hasan, Ahmad Sarid Ezra Fathin, Khoirun Nisa, Ema Utami</i>	315-319
59	Smart Data Centre Monitoring System Based On Internet of Things (IoT) (Study Case: Pustipanda UIN Jakarta) <i>Feri Fahrianto, Nenny Anggraini, Hendra Bayu Suseno, Almas Shabrina, Alfatta Reza</i>	320-328

60	Determining Evaluated Domain Process through Problem Identification using COBIT 5 Framework <i>Fitroh, Sahbani Siregar, Eri Rustamaji</i>	329-334
61	The Psychometric and Interpretative Analyses for Assessing the End-User Computing Satisfaction Questionnaire <i>A'ang Subiyakto, Rosalina, Meinarini Catur Utami, Nia Kumaladewi, Syopiansyah Jaya Putra</i>	335-340
62	Comparison of Characteristic of Two and Three Couplers Mach-Zehnder Interferometers <i>Fauzan Al Ayyubi, Ary Syahriar, Sasono Rahardjo, Faisal Ali</i>	341-345
63	A Novel System to Visualize Aerial Weapon Scoring System (AWSS) using 3D Mixed Reality <i>Andria Kusuma Wahyudi, Ardian Infantono</i>	346-350
64	Decision Making with AHP for Selection of Employee <i>Ria Eka Sari, Abdul Meizar, Dahriani Hakim Tanjung, Ahir Yugo Nugroho Harahap</i>	351-355
65	Applications of Artificial Intelligence to Identify Psychoanalysis Drug Addiction Patients and HIV / AIDS in Cognitive Science Modeling using Bayes Method <i>A. Hanifa Setianingrum, Bagus Sulistyo Budhi</i>	356-362
66	Application of Kalman Filter to Track Ship Maneuver <i>Amicytia Nadzilah, Danny M. Gandana, Jemie Muliadi, Yanto Daryanto</i>	363-367
67	Implementation of SDR for Video Transmission Using GNU Radio and USRP B200 <i>Octarina Nur Samijayani, Pramuditoruni Gitomojati, Dwi Astharini, Suci Rahmatia, Nurul Ihsan Hariz Pratama</i>	368-371
68	Strategic Planning For the Information Development of IPDC (Instituto Profissional De Canossa) Library Using TOGAF Method <i>Umbelina de Fatima Gusmao, Djoko Budiyanto Setyohadi</i>	372-377
69	A Fast and Accurate Detection of Schizont Plasmodium Falciparum Using Channel Color Space Segmentation Method <i>Edy Victor Haryanto S, M. Y. Mashor, A.S. Abdul Nasir, H. Jaafar</i>	378-381
70	Malaria Parasite Detection with Histogram Color Space Method in Giemsa-stained Blood Cell Images <i>Edy Victor Haryanto S, M. Y. Mashor, A.S. Abdul Nasir, H. Jaafar</i>	382-385

71	Automated Segmentation Procedure for Ziehl-Neelsen Stained Tissue Slide Images <i>Bob Subhan Riza, M. Y. Mashor, M. K. Osman, H. Jaafar</i>	386-390
72	Information Security Evaluation using KAMI Index for Security Improvement in BMKG <i>D. I. Sensuse, M. Syarif, H Suprpto, R. Wirawan, D. Satria, Y Normandia</i>	391-394
73	Classification of Maturity Level of Fuji Apple Fruit With Fuzzy Logic Method <i>Evi Dewi Sri Mulyani, Susanto, Jeni Poniman</i>	395-398
74	Exploring the Organizational Factor Contributing to Effective IT Implementation <i>Muhamamd Qomarul Huda, Nur Aeni Hidayah, Meinarini Catur Utami</i>	399-403
75	PeGI in Practice: The e-Government Assessment in National Library of Indonesia <i>Dana Indra Sensuse, Abrar Nasbey, Nordianto, Retno Dewiyanti, Rio Novira, M Fadhil Dzulfikar</i>	404-409
76	Comparative Analysis of Business Process Litigation Using Queue Theory and Simulation (Case Study: Religious Courts South Jakarta) <i>Rizal Broer Bahaweres, Anida Fitriyah, Nurul Faizah Rozy</i>	410-416
77	Design of E-Commerce Information Systems for Houseplants: the Case of Yasyifa Nursery Plantation <i>Ujang Maman, Yuni Sugiarti, Nia Kumaladewi</i>	417-421
78	Development of CCRP Algorithm Based On Departure Time to Support Disaster Evacuation Scheduling <i>Ida Ayu Gde Suwiprabayanti Putra</i>	422-426
79	Critical Success Factors of E-Government Implementation Based on Meta-Ethnography <i>Darmawan Napitupulu, Dana Indra Sensuse, Yudho Giri Sucahyo</i>	427-432
80	Supply Chain Model for University Al Azhar Indonesia in the Field of Education <i>Syarif Hidayat, Cinthia Amalia Martayodha</i>	433-438

- 81 IT Security Governance Evaluation with use of COBIT 5 Framework: A Case Study on UIN Syarif Hidayatullah Library Information System 439-443
Yusuf Durachman, Yuliza Chairunnisa, Djoko Soetarno, Agus Setiawan, Fitri Mintarsih
- 82 Inventory Management Information System Development at BPRTIK Kemkominfo Jakarta 444-447
Elvi Fetrina, Eri Rustamaji, Tatat Nuraeni, Yusuf Durachman
- 83 Hadith Degree Classification for Shahih Hadith Identification Web Based 448-453
Ina Najiyah, Sari Susanti, Dwiza Riana, Mochamad Wahyudi
- 84 Mobile Tourism Application Using Augmented Reality 454-459
Riri Safitri, Deska Setiawan Yusra, Denny Hermawan, Endang Ripmiatin, Winangsari Pradani
- 85 Pilgrimage Organizers Monitoring System To Improve Umrah Services (Case Study: Sub Directorate of Umrah Development of the Ministry of Religious Affairs of the Republic of Indonesia) 460-463
Nia Kumaladewi, Muhammad Anas, Suci Ratnawati, M. Qomarul Huda, Yusuf Durachman
- 86 Spatial Data Management System for Spread of Diniyah Takmiliyah Awaliyah 464-468
Eva Khudzaeva, Zainul Arham, Sunarya
- 87 Conceptual Approach for Gathering SPL Requirement from Goal Model 469-473
Imam Marzuki Shofi, Ahmad Nurul Fajar
- 88 Improvement Accuracy of Oil Meal Packaging with Method ANP 474-479
Asbon Hendra Azhar, Ratih Adinda Destari, Linda Wahyuni, Fitriana Harahap
- 89 A Comparison of Mamdani and Sugeno Method for Optimization Prediction of Traffic Noise Levels 480-483
Alfa Saleh, Fujiati, Rika Rosnelly, Khairani Puspita, Andi Sanjaya
- 90 The Prototype of Zakat Management System in Indonesia by Using the Social Society Approach: A Case Study 484-487
Husni Teja Sukmana, Devi Lestiani, Nenny Anggraeni, Djoko Soetarno

91	Embryos Sorting Efficiency Identification of Eggs with Algorithms Using Gabor Wavelet <i>Adil Setiawan, Rika Rosnelly, Soeheri, Ratna Sri Hayati, Rita Novita Sari</i>	488-493
92	Enterprise Architecture Modeling for Oriental University in Timor Leste to Support the Strategic Plan of Integrated Information System <i>Sergio Soares, Djoko Budiyanto Setyohadi</i>	494-499
93	Optimization of Multiple Depot Vehicle Routing Problem (MDVRP) on Perishable Product Distribution by Using Genetic Algorithm and Fuzzy Logic Controller (FLC) <i>Elin Haerani, Luh Kesuma Wardhani, Dian Kumala Putri, Husni Teja Sukmana</i>	500-504
94	Application for Determining Mustahiq Based on the Priority using Weight Product Method (Case Study: BAZIS DKI Jakarta) <i>Harry Okta Maulana, Imam M. Shofi, Nurul Faizah Rozy, Fenty Eka Muzayyana Agustin</i>	505-508
95	Segmentation of Crack Area on Road Image Using Lacunarity Method <i>I Putu Gede Abdi Sudiatmika</i>	509-514
96	Context for the Intelligent Search of Information <i>Syopiansyah Jaya Putra, Ismail Khalil</i>	515-518
97	Quality Dimensions of Delone & Mclean Model to Measure Students' Accounting Computer Satisfaction: An Empirical Test on Accounting System Information <i>Robi Aziz Zuama, Jamal Maulana Hudin, Diah Puspitasari, Eni Heni Hermaliani, Dwiza Riana</i>	519-524
98	Designing Dipole Antenna for TV Application and Rectangular Microstrip Antenna Working at 3 GHz for Radar Application <i>Suci Rahmatia, Enggar Fransiska DW, Nurul Ihsan Hariz Pratama, Putri Wulandari, Octarina Nur Samijayani</i>	525-530
99	Integration of Bagging and Greedy Forward Selection on Image Pap Smear Classification using Naïve Bayes <i>Dwiza Riana, Achmad Nizar Hidayanto, Fitriyani</i>	531-537
100	Indonesian Teacher Engagement Index (ITEI): Decision Support System for Education <i>Sasmoko, Andi Muhammad Muqsith, Danu Widhyatmoko, Yasinta Indrianti, Aqeel Khan</i>	538-542

101	Evaluating the Accessibility of Provinces' E-Government Websites in Indonesia	543-548
	<i>I Gusti Bagus Ngurah Eka Darmaputra, Sony Surya Wijaya, Media Anugerah Ayu</i>	
102	Development of a Retrieval System for Al Hadith in Bahasa (Case Study: Hadith Bukhari)	549-553
	<i>Atqia Aulia, Dewi Khairani, Nashrul Hakiem</i>	
	Author Index	554-559

Combining Integrated Sampling Technique with Feature Selection for Software Defect Prediction

Sukmawati Anggraeni Putri

STMIK Nusa Mandiri, Information System Program
Jakarta, Indonesia
sukmawati@nusamandiri.ac.id

Frieyadie

AMIK BSI Jakarta, Management Informatic Program
Jakarta, Indonesia
frieyadie@bsi.ac.id

Abstract—Good quality software is a supporting factor that is important in any line of work in of society. But the software component defective or damaged resulting in reduced performance of the work, and can increase the cost of development and maintenance. An accurate prediction on software module prone defects as part of efforts to reduce the increasing cost of development and maintenance of software. From the results of these studies are known, there are two problems that can decrease performance prediction of classifiers such imbalances in the distribution of the class and irrelevant of the attributes that exist in the dataset. So as to handle both of these issues, we conducted this research using integrated a sample technique with feature selection method. Based on research done previously, there are two methods of samples including random under sampling and SMOTE for random over sampling. While on feature selection method such as chi square, information gain and relief methods. After doing the research process, integration SMOTE technique with relief method used on Naïve Bayes classifiers, the result of the predicted value better than any other method that is 82%.

Keywords—*imbalance class, feature selection, software defect prediction*

I. INTRODUCTION

In the development of the use of software to support the activities and the work increases, certainly the quality of the software must be considered. But the software component defective or damaged resulting in a decrease in customer satisfaction, as well as an increase in the cost of development and maintenance [1].

An accurate prediction on software module software defects as part of effort to reduce the increasing cost of development and maintenance of software that has been done by previous researchers [2]. In this study focuses on 1) estimating the amount of defect in the software, 2) find the relationship of software defects, 3) classifying defect software components, which defect module and non defect module [3].

While the software defect prediction research that has been done by previous research such as Naïve Bayes classifier [4] produce a good performance with an average probability of

71%. Naïve Bayesian is a simple classification [5] with a time of learning process is faster than any other machine learning [4]. Additionally it has a good reputation on the accuracy of prediction [6]. However, this method is not optimal in the case of having an unbalanced dataset [7].

The predicted performance of this method gets worse when the dataset has an irrelevant attribute [8]. While NASA MPD dataset [9] which have been used by previous researchers on software predictions have unbalanced defect datasets with attributes that are not all usable. To deal with unbalanced datasets there are three approaches that can be used, including data level (sample technique), algorithm level and ensemble method [10].

In general, the sample technique is divided into two types, including over sampling method is Random Over Sampling [11]. While the under sampling method is Random Under Sampling [12] and Resample method [13].

As for solving the problem of attributes that are irrelevant using attribute selection methods such as Information Gain, Chi Square, and Relief [14].

In this study, we propose to integrate the sample technique with feature selection method to handle imbalance class and attribute irrelevant to the Naïve Bayesian classification to produce a better accuracy in the software defect prediction.

There are several steps done in this study, First, sample technique to handle the imbalance class. Then, approaching the selection attributes thrown clasifying for software defect prediction. Then calculating the validation and evaluation technique to determine the proposed method is wheter it better or not with the existing methods.

II. RELATED WORK

Research on the software defect prediction are one of the research that has been done by previous researches. From these studies it is known state of the art about software defect prediction research that discusses the imbalance class.

As research done by Chawla [15] who proposed the use of Synthetic Minority Oversampling Technique to handle the class imbalance by using a Naïve Bayesian classifier,

implemented in eight different dataset from the UCI repository. The results showed for all the data using SMOTE technique on balancing process has a greater potential to improve the performance of Naïve Bayesian and C.45 classifier use in the classification process.

While the research done by Riquelme [13] which states that the dataset in software engineering is very unbalanced. Therefore to balance using two technique, including SMOTE and Weka Randomly Resampling using J48 and Naïve Bayesian classifier is applied to the five datasets form PROMISE repository. The results show the approach SMOTE able to increase the average AUC value of 11.6%. Based on these results, balancing techniques can better classify minority classes.

While the research done by Putri, Wahono [16] states that NASA MDP dataset has unbalanced classes and attribute not relevant. Using the balancing class SMOTE and feature selection information gain, can improve the prediction results are better than Riquelme research only to rebalance the dataset class.

Furthermore the study done by Gao using one of the feature selection algorithm which Relief that have been used in the research done by Kira. The research Gao shows that the Relief method as well as the Information Gain [17].

Therefore in this study implement an approach sample techniques which Synthetic Minority Over-sampling Technique (SMOTE) to reduce the influence of class imbalance and improve the ability to predict the minority class. Relief algorithm as well as for selection the relevant attribute. It also use Naïve Bayes algorithm used in the classification process.

III. METHODE

3.1. Sample Technique

The sample approach is one approach to solve the problem of class imbalance in a dataset. The commonly used sample approaches are over-sampling and under-sampling techniques [10].

a. Over-Sampling Technique

Over-sampling causes excessive duplication in the positive class cause over-fitting. Moreover, over-sampling can increase the number of training dataset, thus causing excessive computational costs [15].

Nevertheless, in research carried out by Chawla [15] found Synthetic Minority Over-sampling Technique (SMOTE) which produces artificially interpolated data on the over-sampling in the minority. The algorithm is simulated by finding k nearest to each minority sample, and then for each neighbor, randomly pick a point on the line connecting neighbors and sample itself. Finally, the data at that point is entered as an example of the new minority. By adding new

minority sample into training data, is expected to over-fitting can be resolved [15].

b. Under-sampling Technique

Under-sampling approaches have been reported to outperform over-sampling approaches in previous literatures. However, the under-sampling approach reduces the majority class, perhaps losing useful information. This results in less accurate predictions [11].

Sampling is done randomly, so the majority of the sample is as large as the number of minority samples. Meanwhile, the sample used in the under-sampled approach is the majority sample that is under the sample [18].

We implemented our proposed Random Under-Sampling and SMOTE in the WEKA tool.

3.2. Feature Selection

At dataset software defects, attributes represent software metrics taken from the source code of the software used in the learning process. However, some attributes that are not relevant to require the removal to improve the accuracy of software defects prediction.

There are two algorithms used in the selection of attributes that wrapper and a filter [17]. In the wrapper algorithm using feedback from learning algorithm. While on the filter algorithm, the training data are analyzed using methods that do not require learning algorithms to determine the most relevant attributes [17]. In this study only uses algorithms to filter the selection attribute. Such as, chi-square (CS), information gain (IG), and Relief algorithm (RLF) [17].

a. Chi Square (CS)

CS can evaluate attribute values by calculating the statistical value related to the class. Statistical CS (also symbolized as χ^2) is a nonparametric statistical techniques by using nominal data (category) with the test frequency.

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

where χ^2 is the test statistic is asymptotically approaching the χ^2 distribution, O_i is the observed frequencies, and E_i is the expected frequency. n is the number of possible outcomes of each event.

b. Information Gain (IG)

In the IG is able to assess the importance attribute by measuring the information gain associated with the class. Generally IG estimates that the change in entropy of information before the state took some information.

$$IG(\text{Class}, \text{Attribute}) = H(\text{Class}) - H(\text{Class} | \text{Attribute}) \quad (2)$$

where H determine entropy. More specifically, suppose that A is the set of all attributes and class attributes being dependent of all the training examples, the value of (a, y) with $y \in \text{Class}$ defines the value of specific examples to attribute $a \in A$, V is the set of attribute values, namely $V = \{\text{value}(a, y) \mid a \in A \cap y \in \text{Class}\}$ and $|s|$ is the number of elements in the set s. G to attribute $a \in A$ defined as follows:

$$IG(\text{Class}, a) = H(\text{class}) - \sum_{v \in V} \frac{| \{y \in \text{class} \mid \text{value}(a, y) = v\} |}{|\text{class}|} \times H(\{y \in \text{class} \mid \text{value}(a, y) = v\}) \quad (3)$$

c. Relief (RLF)

For a given sample R, Relief find the nearest neighbor of the same class or different, which is called the 'nearest hit H' and 'nearest miss M'. It will be updated estimate of the quality of W [A] for all attributes A depending on their values for R, F, and H. The process is repeated in accordance with the value of m, where m is determined by the user. Diff function (Attribute, Instance1, Instance2) clearly defined in accordance with the type attribute. For discrete attributes are defined as follows:

$$\text{diff}(A, I_1, I_2) = \frac{|\text{value}(A, I_1) - \text{value}(A, I_2)|}{\max(A) - \min(A)} \quad (4)$$

The underlying hypothesis is that the relevant attributes are able to distinguish between things of different classes and showed no difference between instances of the same class.

3.3. Naive Bayesian (NB) Classifier

Naïve Bayes assumes that the impact of a certain class attribute value is independent of the values of other attributes. This assumption is called the independent class conditional. This is done to simplify the calculation involved, and in this sense it is considered naive. Naïve Bayes allows representation of dependencies among a subset of attributes [19]. By mathematical calculation as follows:

$$P(X + C_i) = \sum_{k=1}^n P(X_k | C_j) \quad (5)$$

The probability $P(X_1|C_1)$, $P(X_2|C_j)$, ..., $P(X_n|C_i)$ can be easily estimated from the training set. Given that X_k refers to the attribute values for the sample X.

a. If A_k is a category, then $P(X_k|C_j)$ is number of tuples in D class C_j has a value X_k to attribute A_k , divided from $|C_{1,D}|$, number of class C_j tuples in D.

b. If A_k is a continuous value, it is usually assumed that the values have a Gaussian distribution with mean (μ) and standard deviation (σ), can be defined as follows:

$$P(X + C_i) = \sum_{k=1}^n P(X_k | C_j) \quad (6)$$

While

$$P(X_k | C_j) = g(X_k, \mu_{ci}, \sigma_{ci}) \quad (7)$$

We need to calculate μ_{ci} and σ_{ci} , where the mean and standard deviation of the value attribute A_k for training samples of class C_j .

3.4. Validation Technique

In this study using validation techniques 10 fold cross validation, with resulting confusion matrix [20] which are described in Table 1. In the confusion matrix, TN is true negative results are classified (true negative). FN is a positive result that is not properly classified as negative. TP is a positive result correctly classified (true positive). FP is the negative results are not correctly classified as positive (false positive).

TABLE 1.
CONFUSION MATRIX

Class		Initial Value	
		True	False
Prediction	True	TP	FP
Value	False	FN	TN

Confusion matrix of values will produce the ROC curve (Receive Operating Characteristics) whose task is to evaluate the performance of the classifier algorithm. Then the Area Under the ROC as a reference for evaluating which provides a summary of the performance of the classifier algorithm [20]. Area Under the ROC (Receive Operating Characteristic) (AUC) is a single value measurements are derived from signal detection. AUC values range from 0 to 1. The ROC curve is used to characterize the trade-offs between true positive rate (TPR) and false positive rate (FPR). A classifier that provides a large area under the curve is more of a classifier with a smaller area under the curve [21].

3.5. Evaluation Technique

In the statistical evaluation consisted of testing parametric and non-parametric test. As for testing the significant difference of the classifier algorithm performance using the non-parametric tests, such as tests friedman [22].

Friedman test is a non-parametric test that is equivalent to the ANOVA parametric test. In the Friedman test ranking algorithm for each data set separately, the algorithm performance is good to be ranked first, while for the second-best given. Friedman test carried out by the appropriate post hoc test for comparison of more than one classifier with multiple datasets [22].

Below will be shown on the test friedman formula:

$$X_r^2 = \left[\frac{12}{(n \times k)(k+1)} \times \sum_{j=1}^k (R_j)^2 \right] - [(3n)(k+1)] \quad (8)$$

Where

X_r^2 = khai value - the level of two-way squares friedman

N = amount of sample

K = the number of groups samples

1, 3, 12 = constanta

IV. EXPERIMENT RESULT

4.1. Dataset

In this study, using a dataset of software metric (National Aeronautics and Space Administration) MDP repository. They are public datasets used by previous researchers in the field of software defects prediction. NASA dataset MDP can be obtained via the official website Wikispaces (<http://nasa-softwaredefectdatasets.wikispaces.com/>). Datasets used in this study consisted of CM1, MW1, PC1 and PC4 are described in Table 2.

TABLE 2.
NASA MDP DATASET

		NASA MDP Dataset			
		CM1	MW1	PC1	PC4
LOC Count	LOC total	X	X	X	X
	LOC blank	X	X	X	X
	LOC code_and_comment	X	X	X	X
	LOC comment	X	X	X	X
	LOC executable	X	X	X	X
	Number of lines	X	X	X	X
Halstead Attributes	Content	X	X	X	X
	Difficulty	X	X	X	X
	Effort	X	X	X	X
	Error est	X	X	X	X
	Length	X	X	X	X
	Level	X	X	X	X
	Prog_time	X	X	X	X
	Volume	X	X	X	X
	Num_operands	X	X	X	X
	Num_operators	X	X	X	X
	Num_unique_operands	X	X	X	X
	Num_unique_operators	X	X	X	X
	Cyclomatic_complexity	X	X	X	X
	Cyclomatic_density	X	X	X	X
McCabe Attributes	Design_complexity	X	X	X	X
	Essential_complexity	X	X	X	X
	Branch_count	X	X	X	X
	Call_pairs	X	X	X	X
Miscellaneous Attributes (another)	Condition_count	X	X	X	X
	Decision_count	X	X	X	X
	Decision_density	X	X	X	X
	Design_density	X	X	X	X
	Edge_count	X	X	X	X
	Essential_density	X	X	X	X
	Parameter_count	X	X	X	X
	Maintenance_severity	X	X	X	X
	Modified_condition_count	X	X	X	X
	Multiple_condition_count	X	X	X	X
	Global_data_complexity				
	Global_data_density				
	Normalized_cyclomatic_complexity	X	X	X	X
	Percent_comments	X	X	X	X
	Node_count	X	X	X	X
Number of code attribute		37	37	37	37
Number of Modul		342	266	759	1399
Number of defect modul		41	28	61	178

As shown in Table 2, that each dataset consists of several software modules, along with the number of errors and attributes characteristic code. NASA dataset preprocessing MDP has 38 attributes plus one attribute disabled or not

disabled (defective?). The attribute consists of an attribute type Halstead, McCabe, Line of Code (LOC) and miscellaneous attributes [23].

The dataset was obtained from NASA MDP software matrices which are described in Table 3, as follows:

TABLE 3.
SPECIFICATIONS AND ATTRIBUTES NASA MDP

System	Language Program	Dataset	LOC
Instruments a spacecraft	C	CM1	17K
Database	C	MW1	8K
Flight software for satellites orbiting the Earth	C	PC1	26K
		PC4	30K

4.2. Implementation and Experiment Results

In this study using Naive Bayesian classifier algorithm at 4 datasets NASA MDP (CM1, MW1, PC1 and PC4). Classifier algorithm will be applied on the integration sample technique with a selection attribute method. Like, NB classifier with SMOTE and CS, NB classifier with SMOTE and IG, NB classifier with SMOTE and RLF, NB classifier with RUS + CS, NB classifier with RUS and IG, and NB classifier with RUS and RLF.

TABLE 4.
AUC VALUE

Classification	CM1	MW1	PC1	PC4
NB	0,694	0,727	0,768	0,825
NB with SMOTE and CS	0,766	0,759	0,734	0,856
NB with RUS and CS	0,752	0,722	0,79	0,859
NB with SMOTE and IG	0,751	0,767	0,817	0,856
NB with RUS and IG	0,753	0,722	0,79	0,859
NB with SMOTE and RLF	0,761	0,779	0,821	0,86
NB with RUS and RLF	0,755	0,747	0,793	0,878

In Table 4 shows the results AUC values were well on the use of models NB with SMOTE and RLF on two datasets (MW1, PC1). As for the CM1 dataset shows AUC good value on NB with SMOTE and CS models. And for PC4 dataset shows AUC good value on NB with RUS and RLF model.

4.3. Comparison Between Previous Models

To know that the proposed model has increased the accuracy after the optimized use of integration between sampling technique and feature selection algorithm, then do a comparison between the proposed model and a model that has been proposed by Menzies [4], Requilme [13] and Putri [24].

TABLE 5.
AUC OF COMPARISON BETWEEN PREVIOUS MODELS

Model	CM1	MW1	PC1	PC4
Menzies (2011), NB	0,694	0,727	0,768	0,825
Requielme (2008), NB with SMOTE	0,739	0,751	0,793	0,858
Putri, Wahono (2015), NB with SMOTE and IG	0,751	0,767	0,817	0,856
Propose Model, NB with SMOTE and RLF	0,761	0,779	0,821	0,86

Results of the experiments are shown in Table 5 to produce the best classification model in the dataset displayed in bold. Shows the proposed model produces increased AUC values compared to other models.

While in Figure 1 describes the a comparison chart of AUC values for the four models of the four datasets NASA MDP.

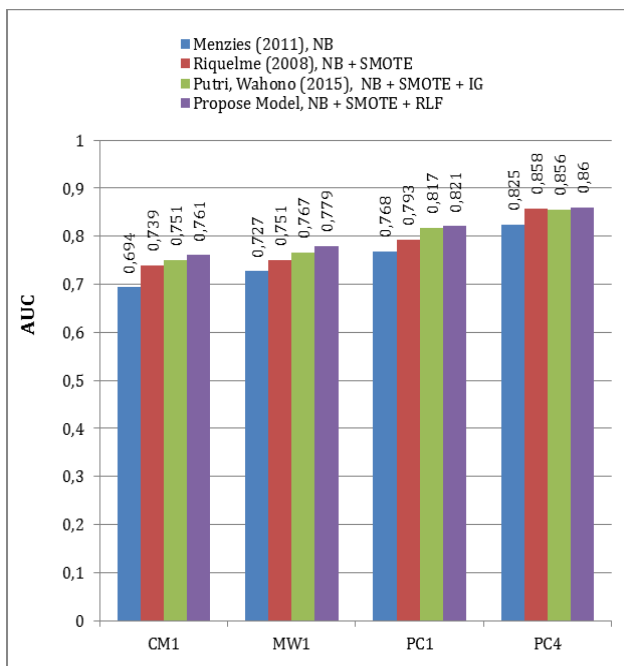


Figure 1. Chart of Comparison AUC values between Previous Model

To know the difference any proposed model, then do a comparison using the non-parametric statistical calculations used for the computation of the classifier algorithm. Like, friedman test.

AUC values model of NB, NB with SMOTE, NB with SMOTE and IG, and NB with SMOTE and RLF compared using friedman test described in Table 6.

TABLE 6.
THE P VALUE OF AUC COMPARISON FRIEDMAN TEST

	NB	NB with SMOTE	NB with SMOTE and IG	NB with SMOTE and RLF
NB (Menzies, 2011)	1	0,046 (Sig)	0,046 (Sig)	0,046 (Sig)
NB with SMOTE (Riquelme, 2008)	0,046 (Sig)	1	0,317 (No Sig)	0,046 (Sig)
NB with SMOTE and IG (Putri, 2015)	0,046 (Sig)	0,317 (No Sig)	1	0,046 (Sig)
NB with SMOTE and RLF (Proposed Model)	0,046 (Sig)	0,046 (Sig)	0,046 (Sig)	1

As shown in Table 6 shows the proposed model of model NB with SMOTE and RLF having P value 0.046, then $P < \alpha$ (0.05). So the NB model with SMOTE and RLF has significant differences with the pure NB model. The model of his study also has significant differences with NB, with each P value for NB with SMOTE is 0.046, while P value NB with SMOTE and IG is 0.046.

From these results, the SMOTE and RLF model applied to the Naive Bayesian classification has better calculation performance than the model that has been proposed with previous researchers.

V. CONCLUSION

From the results of calculations on research application of integration of sample method with the selection attribute of SMOTE and RLF in Naive Bayes classification yields better AUC value compared to the other model. SMOTE and RLF model is superior to the two datasets of the four datasets used, with a value of 78% in the MW1 and 82% datasets on the PC1 dataset.

Whereas when compared with models that have been proposed by previous researchers, such as Naive Bayesian, SMOTE on Naive Bayesian, SMOTE and IG on Naive Bayesian. From the results of research the value of AUC SMOTE and RLF on Naive Bayesian better performance than the model in all dataset used in the other research.

This result can be concluded from comparison result using friedman test, where P value is 0,046, which means $P < \alpha$ (0,05).

But from these results, the use of sample techniques and attribute selection algorithms in software defect prediction research can be done in the next research development, including:

1. For the selection of attributes in future studies may use techniques wrapper on attribute selection methods.
2. In further research can use a combination of sample technique with ensemble algorithm to improve the performance of the classifier.
3. In further research can use other classifiers, such as Logistic Regression, Neural Networks and SVM.

ACKNOWLEDGMENT

We should like to express our gratitude to RSW (Romi Satria Wahono) Intelligent Research Group for warm discussion about this research. Also for PPPM STMIK Nusa Mandiri Jakarta and PPPM AMIK BSI Jakarta, which has supported us to do this research.

REFERENCES

- [1] A. B. de Carvalho, A. Pozo, and S. R. Vergilio, "A symbolic fault-prediction model based on multiobjective particle swarm optimization," *J. Syst. Softw.*, vol. 83, no. 5, pp. 868–882, May 2010.
- [2] C. Catal, "Software fault prediction: A literature review and current trends," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4626–4636, Apr. 2011.
- [3] Q. Song, Z. Jia, M. Shepperd, S. Ying, and J. Liu, "A General Software Defect-Proneness Prediction Framework," *IEEE Trans. Softw. Eng.*, vol. 37, no. 3, pp. 356–370, May 2011.
- [4] T. Menzies, J. Greenwald, and A. Frank, "Data Mining Static Code Attributes to Learn Defect Predictors," *IEEE Trans. Softw. Eng.*, vol. 33, no. 1, pp. 2–13, Jan. 2007.
- [5] P. Domingos, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," *Mach. Learn.*, vol. 29, no. 2–3, pp. 103–130, 1997.
- [6] B. Turhan and A. Bener, "Analysis of Naive Bayes' assumptions on software fault data: An empirical study," *Data Knowl. Eng.*, vol. 68, no. 2, pp. 278–290, Feb. 2009.
- [7] C. Andersson, "A replicated empirical study of a selection method for software reliability growth models," *Empir. Softw. Eng.*, vol. 12, no. 2, pp. 161–182, Oct. 2006.
- [8] T. M. Khoshgoftaar and K. Gao, "Feature Selection with Imbalanced Data for Software Defect Prediction," *2009 Int. Conf. Mach. Learn. Appl.*, pp. 235–240, Dec. 2009.
- [9] M. Shepperd, Q. Song, Z. Sun, and C. Mair, "Data Quality : Some Comments on the NASA Software Defect Data Sets," *Softw. Eng. IEEE Trans.*, vol. 39, no. 9, pp. 1–13, 2013.
- [10] B. W. Yap, K. A. Rani, H. Aryani, A. Rahman, S. Fong, Z. Khairudin, and N. N. Abdullah, "An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets," *Proc. First Int. Conf. Adv. Data Inf. Eng.*, vol. 285, pp. 13–23, 2014.
- [11] Y. Liu, X. Yu, J. X. Huang, and A. An, "Combining integrated sampling with SVM ensembles for learning from imbalanced datasets," *Inf. Process. Manag.*, vol. 47, no. 4, pp. 617–631, Jul. 2011.
- [12] K. Gao and T. M. Khoshgoftaar, "Software Defect Prediction for High-Dimensional and Class-Imbalanced Data," *Proc. 23rd Int. Conf. Softw. Eng. Knowl. Eng.*, no. 2, 2011.
- [13] J. C. Riquelme, R. Ruiz, and J. Moreno, "Finding Defective Modules from Highly Unbalanced Datasets," *Engineering*, vol. 2, no. 1, pp. 67–74, 2008.
- [14] K. Gao, T. M. Khoshgoftaar, H. Wang, and N. Seliya, "Choosing software metrics for defect prediction : an investigation on feature selection techniques," *Softw. Pract. Exp.*, vol. 41, no. 5, pp. 579–606, 2011.
- [15] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE : Synthetic Minority Over-sampling Technique," *J. Artif. Intell.*, vol. 16, pp. 321–357, 2002.
- [16] S. A. Putri and R. S. Wahono, "Integrasi SMOTE dan Information Gain pada Naive Bayes untuk Prediksi Cacat Software," *J. Softw. Eng.*, vol. 1, no. 2, pp. 86–91, 2015.
- [17] K. Gao and T. M. Khoshgoftaar, "Software Defect Prediction for High-Dimensional and Class-Imbalanced Data," *Conf. Proc. 23rd Int. Conf. Softw. Eng. Knowl. Eng.*, no. 2, 2011.
- [18] N. Japkowicz, "The Class Imbalance Problem : Significance and Strategies," *Proc. the 2000 Int. Conf. Artif. Intell. Spec. Track Inductive Learn. Vegas*, 2000.
- [19] M. Jain and V. Richariya, "An Improved Techniques Based on Naive Bayesian for Attack Detection," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 1, pp. 324–331, 2012.
- [20] C. X. Ling, "Using AUC and Accuracy in Evaluating Learning Algorithms," pp. 1–31, 2003.
- [21] C. X. Ling and H. Zhang, "AUC: a statistically consistent and more discriminating measure than accuracy," *Proc. 18th Int. Jt. Conf. Artif. Intell.*, 2003.
- [22] J. Demsar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.
- [23] S. Lessmann, S. Member, B. Baesens, C. Mues, and S. Pietsch, "Benchmarking Classification Models for Software Defect Prediction : A Proposed Framework and Novel Findings," *IEEE Trans. Softw. Eng.*, vol. 34, no. 4, pp. 485–496, 2008.
- [24] S. A. Putri and Wahono, "Integrasi SMOTE dan Information Gain pada Naive Bayes untuk Prediksi Cacat Software," *J. Softw. Eng.*, vol. 1, no. 2, pp. 86–91, 2015.

AUTHOR INDEX

Name	Page
A. Hanifa Setianingrum	356
A.S. Abdul Nasir	378, 382
A'ang Subiyakto	304, 335
Abdul Meizar	351
Abrar Nasbey	404
Achmad Nizar Hidayanto	531
Adil Setiawan	488
Adji Sukmana	315
Adyanata Lubis	47
Agus Hermawan	257
Agus Setiawan	439
Ahir Yugo Nugroho Harahap	351
Ahmad F. Misman	30
Ahmad Melani	304
Ahmad Nurul Fajar	469
Ahmad Sarid Ezra Fathin	315
Aida Fitriyani	231
Aidawayati Rangkuti	177
Ajib Susanto	166
Akmaludin	273
Aldy Rialdy Atmadja	61
Alfa Saleh	480
Alfatta Reza	320
Ali Akbar Nikoukar	210
Ali Rahman	12
Almas Shabrina	320
Amicytia Nadzilah	363
Andi Irmayana	66
Andi Muhammad Muqsith	538
Andi Sanjaya	480
Andrean Limanto	122
Andrew Fiade	293
Andrew Tanny Liem	210
Andria Kusuma Wahyudi	346
Anida Fitriyah	410
Aqeel Khan	538
Ardian Infantono	346
Aries Susanto	231
Arif Djunaidy	282
Arif Dwi Laksito	193
Arif Ridho Lubis	98
Arini	1, 23
Armin Lawi	177
Ary Syahriar	341
Asbon Hendra Azhar	474

Atqia Aulia	549
Awaludin Abid	298
Azqa Fikri Khwarizma	122
B. Herawan Hayadi	47
Bagus Sulistyo Budhi	356
Bob Subhan Riza	386
Busman	148
Candra Ahmadi	55
Cecep Nurul Alam	50
Christy Atika Sari	42, 166
Cinthia Amalia Martayodha	433
D. Satria	391
Dahriani Hakim Tanjung	351
Dana Indra Sensuse	186, 216, 391, 404, 427
Daniel Siahaan	282
Danny M. Gandana	363
Danu Widhyatmoko	538
Darmawan Napitupulu	427
David	171
De Rosal Ignatius Moses Setiadi	42, 166
Denny Hermawan	454
Deny Budiyanto	251
Deska Setiawan Yusra	454
Devi Lestiani	484
Dewi Khairani	549
Diah Puspitasari	519
Dian Kumala Putri	500
Dian Sa'adillah Maylawati	116
Dinar Ajeng Kristiyanti	309
Diyah Puspitaningrum	36
Djoko Budiyanto Setyohadi	245, 251, 287, 372, 494
Djoko Soetarno	439, 484
Doni Purnama Alamsyah	241
Dwi Astharini	368
Dwiza Riana	198, 448, 519, 531
Dyah Budiastuti	132
Edi Abdurachman	171
Edy Kristianto	278
Edy Victor Haryanto S	378, 382
Eko Hari Rachmawanto	42, 166
Elin Cahyaningsih	216
Elin Haerani	500
Elvi Fetrina	444
Ema Utami	298, 315
Endang Ripmiatin	545
Enggar Fransiska DW	252
Eni Heni Hermaliani	519
Eri Rustamaji	224, 444

Eva Khudzaeva	464
Evi Dewi Sri Mulyani	395
Evi Triandini	282
Faisal Ali	341
Fauzan Al Ayyubi	341
Fenty Eka Muzayyana Agustin	505
Feri Fahrianto	320
Ferry Fachrizal	98
Fitri Mintarsih	439
Fitriana Harahap	474
Fitriyani	531
Fitroh	329
Frieyadie	235
Fuad Hasan	315
Fujiati	480
Gat	204
Gede Adhitya Bayu Pramana	85
Green Arther Sandag	210
Gries Yulianti	36
H Suprpto	391
H. Jaafar	378, 382, 386
Hadi Jayusman	245
Halim Maulana	98
Handrie Noprisson	186, 216
Harjanto Prabowo	132
Harry Okta Maulana	505
Haryanto Samuel Laluyan	71
Hasnan Afif	298
Hendra Bayu Suseno	320
Herlino Nanang	30
Husni Teja Sukmana	484, 500
I Gede Surya Rahayuda	18
I Gusti Bagus Ngurah Eka Darmaputra	543
I Komang Agus Ady Aryanto	257
I Made Agus Wirahadi Putra	126
I Made Dendi Maysanjaya	126
I Nyoman Kusuma Wardana	257
I Putu Agus Aditya Satria Wibawa	85
I Putu Gede Abdi Sudiarmika	509
I. Intan	6
I.S.W.B. Prasetya	36
Ichsan Taufik	116
Ida Ayu Gde Suwiprabayanti Putra	422
Ida Madieha Abdul Ghani Azmi	25, 263
Ika Arthalia Wulandari	186
Imam Marzuki Shofi	469, 505
Imelda	122
Ina Maryani	198

Ina Najiyah	448
Irwan Oyong	298
I-Shyan Hwang	210
Ismail Khalil	515
Isra Mouludi	47
Izzah Fadhilah Akmaliah	186
Jamal Maulana Hudin	519
Jemie Muliadi	363
Jeni Poniman	395
Joni Iskandar	138
Jufri	47
Juriah Abd. Jalil	263
Kasman Rukun	47
Khairani Puspita	480
Khoirun Nisa	315
Komang Rinatha	81
Kusrini	193
Kusuma Chandra Kirana	193
Linda Marlinda	439
Linda Wahyuni	474
Lisa Latifah	231
Louis Casanova	278
Luh Gede Surya Kartika	85
Luh Kesuma Wardhani	500
M Fadhil Dzulfikar	404
M. K. Osman	386
M. Syarif	391
M. Y. Mashor	378, 382
Made Adi Surya Antara	103
Made Satria Wibawa	126
Mahyuddin Daud	25, 263
Marcel	278
Marchel Thimoty Tombeng	71
Marianus Omba Riku	287
Masami Asakawa	268
Masao Okano	268
Media Anugerah Ayu	543
Meinarini Catur Utami	304, 335, 399
Mesnan Silalahi	186
Miftah Ardiansyah	160
Miftahul Fauzan	293
Mihuandayani	315
Mira Kartiwi	109
Mochamad Wahyudi	309
Mohamad Irfan	50
Muhamad Idris Purwanto	193
Muhamad Rizka Akbar	1
Muhamamd Qomarul Huda	399, 460

Muhammad Ali Ramdhani	12
Muhammad Anas	460
Muhammad Hamdi	144
Muhammad Indra	50
Muharman Lubis	109
Nashrul Hakiem	549
Nenny Anggraini	293, 320
Ngatindriatus	182
Ni Putu Linda Santiari	18
Nia Kumaladewi	335, 417, 460
Nordianto	404
Nur Aeni Hidayah	399
Nur Ilmiyati Djalal	177
Nurbojatmiko	91
Nurhayati	148, 154
Nurida Ashanti	61
Nurul Faizah Rozy	410, 505
Nurul Ihsan Hariz Pratama	368
Nuryasin	231
Octarina Nur Samijayani	368, 525
Oda I. B. Hariyanto	241
Okfalisa	138, 144
Padma Nyoman Crisnapati	257
Pamela Kareen	216
Pramuditoruni Gitomojati	368
Putri Wulandari	525
Qadavi Muhammad Sofyan	76
R. Wirawan	391
Rahmat Fajar Al Farizky	154
Ratih Adinda Destari	474
Ratna Sri Hayati	488
Raymondus Raymond Kosala	171, 182
Reinert Yosua Rumagit	122
Retno Dewiyanti	404
Ria Eka Sari	351
Rika Rosnelly	480, 488
Rinda Wahyuni	91
Rio Novira	404
Riri Safitri	454
Rismayani	66
Rita Novita Sari	488
Rizal Broer Bahaweres	1, 410
Robi Aziz Zuama	519
Rosalina	335
Rosmeri	144
Safril	47
Sahbani Siregar	329
Sahida Safuan	263

Saktioto	144
Sandy Kosasi	132
Sari Susanti	448
Sasmoko	538
Sasono Rahardjo	341
Sergio Soares	494
Shingo Mabu	55
Soeheri	488
Sony Surya Wijaya	543
Stephanus Remond Waworuntu	204
Suci Rahmatia	368
Suci Ratnawati	460
Sukmawati Anggraeni Putri	235
Sunarya	464
Suryadiputra Liawatimena	122
Susanti Margaretha Kuway	182
Susanto	395
Suzi Fadhilah Ismail	25
Syarif Hidayat	433
Syopiansyah Jaya Putra	335, 515
T. Sutojo	42
Tata Sutabri	160
Tatat Nuraeni	444
Taufik Baidawi	273
Ujang Maman	417
Umbelina de Fatima Gusmao	372
Victor Prasetya Pietono	122
Wahyudin Darmalaksana	12
Wayan Suryasa	81
Wendy	182
Wildan Budiawan Zulfikar	61
Winangsari Pradani	454
Wirarama Wedashwara	55
Y. Normandia	391
Yana Aditia Gerhana	61
Yanto Daryanto	363
Yasinta Indrianti	538
Yayak Kartika Sari	315
Yohanes Halim	122
Yudho Giri Sucahyo	186, 216, 427
Yuliza Chairunnisa	439
Yuni Sugiarti	304.417
Yusuf Durachman	109, 273, 304, 439, 444
Zahidah Zulkifli	30
Zainul Arham	464

