ISBN: 978-1-7281-2379-0
IEEE Catalog Number: CFP19B09-USB

IEEE INDONESIA SECTION

ICSITech

PROCEEDING

2019 5th International Conference on Science in Information Technology (ICSITech)

October

23-24, 2019

Hotel Grand Inna Malioboro
Yogyakarta, Indonesia

Hosted by:

Jointly organized with:

2019 5th International Conference on Science in Information Technology (ICSITech)

October 23-24, 2019
Yogyakarta, Indonesia

COPYRIGHT AND REPRINT PERMISSION:

IEEE Catalog Number : CFP19B09-USB
ISBN : 978-1-7281-2379-0

Editor : Awang Hendrianto Pratomo, Andri Pranolo, Bagus Muhammad Akbar, Shoffan Saifullah, Ahmad Taufiq Akbar, Rochmat Husaini, Heriyanto, Andiko Putro Suryotomo, Vynska Amalia Permadi, Sylvert Prian Tahalea

Publisher : IEEE

Secretariat : Department of Informatics Engineering
Universitas Pembangunan Nasional Veteran Yogyakarta
Jln. Babarsari no. 2, Yogyakarta 55281
Special Region of Yogyakarta, Indonesia

PROCEEDING

2019 5th International Conference on Science in Information Technology (ICSITech)

"Embracing Industry 4.0 : Towards Innovation in Cyber Physical System"

Version: 2019-10-20

October 23-24, 2019
Yogyakarta, Indonesia

Introduction

Welcome to ICSITech 2019. Universitas Pembangunan Nasional "Veteran" Yogyakarta is honored to be the host of this year's International Conference on Science in Information Technology (ICSITech). The ICSITech is jointly organized with Universitas Ahmad Dahlan, Universitas Mulawarman, Universitas Pembangunan Nasional "Veteran" Yogyakarta, Universiti Teknologi Malaysia (UTM) Big Data Centre, Universitas Pendidikan Indonesia, Universitas Muhammadiyah Surakarta, Universiti Putra Malaysia, Universiti Malaysia Sabah, Universitas Budi Luhur, Politeknik Negeri Samarinda, Politeknik Negeri Padang, Universitas Negeri Malang, Universiti Teknikal Malaysia Melaka(UTeM), and ITSA University Colombia.

Since this is the Fifth conference, we wish to repeat the success of the four previous conferences. We do hope this annual conference will continue to be held in the next coming years (2020 in Malaysia, 2021 in Surakarta - Indonesia, etc.) with increasing quality. For this year's conference, we proudly present the theme of ICSITech 2019, "**Embracing Industry 4.0 : Towards Innovation in Cyber Physical System**". The theme is taken from our university identity as a university which consistently takes part in education and responds to the development of science, technology, art, society demands, and global change. It is also our university vision to be actively involved and acted as a leader and initiator in research and development in order to achieve academic excellence.

We are pleased to inform you that the ICSITech 2019 has been approved by IEEE for technical co-sponsorship; therefore, the papers which are accepted and presented will be further considered to be published in the IEEE Xplore Digital Library. I wish to extend a warm welcome to Prof. Dr. Wisnu Jatmiko, S.T., M.Kom., as IEEE Indonesia Section Chair. There are 118 papers from 17 countries submitted to the ICSITech 2019 with 48.28% acceptance ratio. Congratulations for all authors and presenters whose papers are accepted.

Thank you for choosing ICSITech 2019 and disseminating your research here.

Today, we are lucky to have two keynote speakers who will broaden our insights about Big Data Era in IT perspective. They will talk about their expertise and we do hope this event could bring many benefits, especially in the fields of education, industry, and society. We are honored for the presence of Prof. Dr. Leonel Hernandes (ITSA Colombia) and Assoc. Prof. Dr. Mohammad Shanuddin bin Zakaria (Universiti Kebangsaan Malaysia), thank you very much.

The previous conferences were held in Bandung-Indonesia and Melaka – Malaysia, respectively. This year, the ICSITech 2019 is taking place in Yogyakrta. Our city nickname is The City of Education. Yogyakarta is one of the favorite travel destinations, especially in Java, with many wonderful heritage tourism destinations and delightful culinary creations. Please enjoy your stay in Yogyakarta.

Finally, we would like to thank all of keynote speakers, participants, sponsors, associations, and partners for being a part of this conference. On behalf of the organizing committee, we wish to express our highest appreciation and sincere thanks to all of you who attend this event and we wish you have valuable discussion and networking. I also thank the committee for all efforts to make ICSITech 2019 successful.

Thank you.

General Chair
Dr. Awang Hendrianto Pratomo, S.T., M.T.
Department of Informatics Engineering
Universitas Pembangunan Nasional "Veteran" Yogyakarta, Indonesia

Welcome Message from Rector of Universitas Pembangunan Nasional Veteran Yogyakarta

Greetings to all distinguished guests, keynote speakers, and conference participants of the fifth (5th) International Conference on Science in Information Technology ICSITech 2019. It is our honor and pleasure to be the host of this year's ICSITech.

Ladies and gentlemen,
Universitas Pembangunan Nasional Veteran Yogyakarta has been taking part in organizing ICSITech since the first conference back in 2015. Since its inception ICSITech is jointly organized by Universitas Ahmad Dahlan, Universitas Mulawarman, Universitas Pembangunan Nasional "Veteran" Yogyakarta, Universiti Teknologi Malaysia (UTM) Big Data Centre, Universitas Pendidikan Indonesia, Universitas Muhammadiyah Surakarta, Universiti Putra Malaysia, Universiti Malaysia Sabah, Universitas Budi Luhur, Politeknik Negeri Samarinda, Politeknik Negeri Padang, Universitas Negeri Malang, Universiti Teknikal Malaysia Melaka(UTeM), and ITSA University Colombia. And this year, for the first time, our university got the opportunity to be the host of ICSITech 2019. In this occasion, I want to personally congratulate our Department of Informatics Engineering for their commitment and hard work to ensure the success of ICSITech 2019.

Ladies and gentlemen,
ICSITech 2019 was held to provide as an event for IT academic and IT expert to disseminate their knowledge on the development of computer science education and expand the network connection on the research activities. Furthermore we intend to make this conference as a motivation for researchers to publish their ideas about theory and application of IT for education, society, and industry in order to support the development and quality improvement of local, regional , and global researches, in line with Universitas Pembangunan Nasional Veteran Yogyakarta vision to be the pioneer of national development based on the spirit of patriotism.

Finally, on behalf of the organizing committee, we wish to express our highest appreciation and sincere thanks to all of our distinguished participants who attend this event and we wish you have valuable discussion and networking. Welcome to Yogyakarta and we hope all of our guests enjoying the services we provide.

Thank you.

Rector of Universitas Pembangunan Nasional Veteran Yogyakarta,
Dr. Mohamad Irhas Effendi, M.S.

Welcome Message from IEEE Indonesia Section

Prof. Dr.Eng Wisnu Jatmiko, SMIEEE
Chairman, IEEE Indonesia Section

Dr. Kurnianingsih, SMIEEE
Vice Chair, IEEE Indonesia Section

Dear Distinguished Guests, Colleagues, researchers, professionals, ladies and gentlemen, Good morning, a prosperous, warm, and spirited greeting.

On behalf of IEEE Indonesia Section, we would like to extend our warmest welcome to all keynote speakers, presenters, and participants to 2019 International Conference on Science in Information Technology (ICSITech). The conference theme is: "**Embracing Industry 4.0 : Towards Innovation in Cyber Physical System**".

ICSITech is an annual international conference technical co-sponsored by IEEE Indonesia Section and this year conference is hosted by Universiti Pembangunan Nasional "Veteran" Yogyakarta in collaboration with Universiti Teknologi Malaysia (UTM) Big Data Centre, Universitas Ahmad Dahlan, Universitas Mulawarman, Universitas Pendidikan Indonesi, Universitas Muhammadiyah Surakarta, Universiti Putra Malaysia, Universiti Malaysia Sabah, Universitas Budi Luhur, Politeknik Negeri Samarinda, Politeknik Negeri Padang, Universitas Negeri Malang, Universiti Teknikal Malaysia Melaka(UTeM), and ITSA University Columbia. The conference aims to bring together researchers and experts in information systems to share their ideas, experiences and insights.

IEEE Indonesia Section has conducted many activities over 32 years in Indonesia. In terms of collaboration, IEEE Indonesia section has a good and mutual relationship with ICT organizations, Industries, Government, Universities as well as the Community in Indonesia. IEEE Indonesia Section has contributed in about 58 different International conferences annually. As the fifth year of ICSITech, this conference shows its sustainability due to the hard work of the conference organizers, well organized conference and high-quality papers. We do hope in the near future some high-quality conferences will be continued and strengthened, so the result will give more benefit and positive impact to the human being, especially to Indonesian people.

In this occasion, I would also like to say welcome to Yogyakarta, which serves beautiful heritages, culture, with warm, polite and friendly people, a vibrant culture and lifestyle.
Finally, we do hope all of you will have enjoyable and valuable experience during this event. You may share your best knowledge in your area of research and professional activities.
Thank you.

Yogyakarta, 23rd October 2019
IEEE Indonesia Section

Organization

Steering Committee

- Munir (Universitas Pendidikan Indonesia)
- Dwi Hendratmo W. (Institut Teknologi Bandung)
- Satriyo Dharmanto (IEEE Indonesia Section)
- Shi-Jinn Horng (National Taiwan University Science and Technology, Taiwan)
- Hj. Siti Mariyam Shamsudin (Universiti Teknologi Malaysia)
- Amelia Ritahani Ismail (International Islamic University Malaysia)
- Rafal Drezewski (AGH University of Science and Technology)
- Rodziah Atan (Universiti Putra Malaysia)
- Rayner Alfred (Universiti Malaysia Sabah)
- Heui Seok Lim (Korea University)
- Goutam Chakraborty (Iwate Prefectural University)
- Didi Sukyadi (Universitas Pendidikan Indonesia)
- Tsukasa Hirashima (Hiroshima University)
- Halimah Badioze Zaman (Universiti Kebangsaan Malaysia)

Scientific Committee

- Agus Pratondo (Telkom University)
- Abderrafiaa Koukam (UTBM, France)
- Abdul Samad Shibghatullah (Universiti Teknikal Malaysia Melaka)
- Ade Gafar Abdullah (Universitas Pendidikan Indonesia)
- Adiwijaya (Telkom University)
- Ag Asri Ag Ibrahim (Universiti Malaysia Sabah)
- Agus Harjoko (Universitas Gadjah Mada)
- Agus Setiabudi (Universitas Pendidikan Indonesia)
- Agus Setiawan (Universitas Pendidikan Indonesia)
- Aji Wibawa (Universitas Negeri Malang)
- Alejandro Rosales (INAOE, Mexico)
- Amer Farea (Taiz University, Arab Saudi)
- Anca Ralescu (University of Cincinnati Ohio)
- Andria Arisal (Indonesia Institute of Science)
- Anhar Risnumawan (Politeknik Elektronika Negeri Surabaya)
- Anton Satria Prabuwono (King Abdulaziz University, Saudi Arabia)
- Anton Yudhana (Universitas Ahmad Dahlan)
- Arda Yunianta (King Abdulaziz University)
- Ari Barmawi (Telkom University)
- Armin Lawi (Hasanuddin University)
- Awang Hendrianto Pratomo (UPN "Veteran" Yogyakarta)
- Azhari SN (Universitas Gadjah Mada)
- Azizi Abdullah (Universiti Kebangsaan Malaysia)
- Azuraliza Abu Bakar (Universiti Kebangsaan Malaysia)
- Bana Handaga (Universitas Muhammadiyah Surakarta)
- Cepy Riana (Universitas Pendidikan Indonesia)
- Chin Kim On (Universiti Malaysia Sabah)
- Danial Hooshyar (Korea University)
- Dedi Rohendi (Universitas Pendidikan Indonesia)
- Deris Stiawan (Universitas Sriwijaya)
- Dewi Octaviani (Korea University)
- Didi Rosiyadi (Research Center for Informatics LIPI)
- Didin Wahyudin (Universitas Pendidikan Indonesia)
- Edi Kurniawan (Research Center for Informatics LIPI)
- Eko Aribowo (Universitas Ahmad Dahlan)

- Engkos Kosasih (Universitas Indonesia)
- Esa Prakasa (Research Center for Informatics LIPI)
- Faaizah Shahbodin (Universiti Teknikal Malaysia Melaka)
- Fajar Suryawan (Universitas Muhammadiyah Surakarta)
- Ford Lumban Gaol (Bina Nusantara University)
- Gede Pramudya Ananta (Universiti Teknikal Malaysia Melaka)
- Gunawan Ariyanto (Universitas Muhammadiyah Surakarta)
- Habib Shah (Universiti Tun Hussein Onn Malaysia)
- Halizah Basiron (Universiti Teknikal Malaysia Melaka)
- Hamzah bin Ahmad (Universiti Malaysia Pahang)
- Hanung Adi Nugroho (Universitas Gadjah Mada)
- Herlina Jayadianti (UPN Veteran Yogyakarta)
- Heru Supriyono (Universitas Muhammadiyah Surakarta)
- Hidayah Rahmalan (Universiti Teknikal Malaysia Melaka)
- Husni Thamrin (Universitas Muhammadiyah Surakarta)
- I Wayan Mustika (IEEE Inonesia Section)
- Ibrahim Ahmad (Universiti Teknikal Malaysia Melaka)
- Igi Ardiyanto (Universitas Gadjah Mada)
- Intan Ermahani A. Jalil (Universiti Teknikal Malaysia Melaka)
- Ito Wasito (Universitas Indonesia)
- Iwan Tri Riyadi Yanto (Universitas Ahmad Dahlan)
- Jazi Istiyanto (Universitas Gadjah Mada)
- Joey M. Suba (University of The Assumption)
- Kamarul Hawari bin Ghazali (Universiti Malaysia Pahang)
- Keng-Yap Ng (Universiti Putra Malaysia)
- Khabib Mustofa (Universitas Gadjah Mada)
- Lala Septem Riza (Universitas Pendidikan Indonesia)
- Lau Hui Keng (Universiti Malaysia Sabah)
- Leonel Hernandez (ITSA University)
- Lian Duan (New Jersey Institute of Technology, USA)
- Lili Wulandhari (Bina Nusantara University)
- Lizawati Salahuddin (Universiti Teknikal Malaysia Melaka)
- Maizatul Akmar Ismail (Universiti Malaya)
- Maman Abdurohman (Telkom University)
- Mar Yah Said (Universiti Putra Malaysia)
- Marzanah Abdul Jabar (Universiti Putra Malaysia)
- Masayu Leyla Khodra (Institut Teknologi Bandung)
- Masrah Azrifah Azmi Murad (Universiti Putra Malaysia)
- Mohammad Syafrullah (Universitas Budi Luhur, Jakarta)
- Mohd Hanafi Ahmad Hijazi (Universiti Malaysia Sabah)
- Mohd Sazali Khalid (Universiti Sains Malaysia)
- Mohd Shahizan Bin Othman (Universiti Teknologi Malaysia)
- Moslem Yousefi (UNITEN, Malaysia)
- Muh. Nursalman (Universitas Pendidikan Indonesia)
- Muhammad Hakim Othman (Universiti Teknikal Malaysia Melaka)
- Muhammad Said Hasibuan (IBI Darmajaya)
- Muhammad Zarlis (Universitas Sumatera Utara)
- Munir (Universitas Pendidikan Indonesia)
- Mustafa Kaiali (Mevlana University, Turkey)
- Nana Suryana Herman (Universiti Teknikal Malaysia Melaka)
- Nataniel Dengen (Universitas Mulawarman)
- Nazri Nawi (Universiti Tun Hussein Onn Malaysia)
- Ngo Hea Choon (Universiti Teknikal Malaysia Melaka)
- Nita Solehati (Chonbuk National University)
- Noel Lopes (Polytechnic of Guarda, Portugal)

- Noor Aida Husaini (Universiti Tun Hussein Onn Malaysia)
- Nor Fazlida Mohd Sani (Universiti Putra Malaysia)
- Nor Haslinda binti Ismail (Universiti Teknikal Malaysia Melaka)
- Noraini Che Pa (Universiti Putra Malaysia)
- Noraini Ibrahim (Universiti Tun Hussein Onn Malaysia)
- Norashikin Ahmad (Universiti Teknikal Malaysia Melaka)
- Norazah Mohd Nordin (Universiti Kebangsaan Malaysia)
- Norhalina Senan (Universiti Tun Hussein Onn Malaysia)
- Norhayati Harum (Universiti Teknikal Malaysia Melaka)
- Norhayati Mohd Ali (Universiti Putra Malaysia)
- Nurgiyatna (Universitas Muhammadiyah Surakarta)
- Nurul Amelina Nashruddin (Universiti Putra Malaysia)
- Nurul Azma Zakaria (Universiti Teknikal Malaysia Melaka)
- Oki Wicaksono (Universitas Gadjah Mada)
- Omar Al Jadaan (MHSU, Uni Arab Emirates)
- Omid Motlagh (CSIRO, Australia)
- Othman Mohd (Universiti Teknikal Malaysia Melaka)
- Ouri Wolfson (University of Illinois, USA)
- Parman Sukarno (Politeknik Pos Indonesia)
- Per Johan Runeson (System Lund University, Sweden)
- Prima Vitasari Hj. Purwadi (Universiti Malaysia Pahang)
- Raden Bagus Fajriya Hakim (Universitas Islam Indonesia)
- Rafah Mohamed Almuttairi (University of Babylon, Iraq)
- Rafal Drezewski (AGH UST, Poland)
- Ramadiani (Universitas Mulawarman)
- Rani Megasari (Universitas Pendidikan Indonesia)
- Rasim (Universitas Pendidikan Indonesia)
- Rayner Alfred (Universiti Malaysia Sabah)
- Reza Firsandaya Malik (Universitas Sriwijaya)
- Reza Pulungan (Universitas Gadjah Mada)
- Rifki Sadikin (Research Center for Informatics LIPI)
- Rimba Widhiana Ciptasari (Telkom University)
- Rinaldi Munir (Institut Teknologi Bandung)
- Riyanarto Sarno (Institut Teknologi Sepuluh Nopember)
- Rodina binti Ahmad (Universiti Malaya)
- Rodziah Atan (Universiti Putra Malaysia)
- Ronny Mardiyanto (Institut Teknologi Sepuluh Nopember)
- Rusli Abdullah (Universiti Putra Malaysia)
- Rusydi Umar (Universitas Ahmad Dahlan)
- Sa'adah Hassan (Universiti Putra Malaysia)
- Sabrina Ahmad (Universiti Teknikal Malaysia Melaka)
- Salfarina Abdullah (Universiti Putra Malaysia)
- Sarina Sulaiman (Universiti Teknologi Malaysia)
- Sarni Suhaila Rahim (Universiti Teknikal Malaysia Melaka)
- Shafaatunnur Hasan (Universiti Teknologi Malaysia)
- Shah Nazir (University of Swabi, Pakistan)
- Shahril bin Parumo (Universiti Teknikal Malaysia Melaka)
- Shaik Shakeel Ahamad (K.G. RCET, Hyderabad, India)
- Sharifah Sakinah Syed Ahmad (Universiti Teknikal Malaysia Melaka)
- Shi-Jinn Horng (NUST, Taiwan)
- Sigit Yuwono (Telkom University)
- Siti Mariyam Shamsuddin (Universiti Teknologi Malaysia)
- Siti Nurul Mahfuzah Mohamad (Universiti Teknikal Malaysia Melaka)
- Siti Sophiayati Yuhaniz (Universiti Teknologi Malaysia)
- Slamet Riyadi (Universitas Muhammadiyah Yogyakarta)

- Songhoua Xu (New Jersey Institute of Technology, USA)
- Sri Kusumadewi (Universitas Islam Indonesia)
- Suhirman (Universitas Teknologi Yogyakarta)
- Sukrisno Mardiyanto (Institut Teknologi Bandung)
- Sultan Noman Qasem (Taiz University, Arab Saudi)
- Sunardi (Universitas Ahmad Dahlan)
- Sunu Wibirama (Universitas Gadjah Mada)
- Sutarman (Universiti Malaysia Pahang)
- Suyanto (Telkom University)
- Suyoto (Universitas Atma Jaya Yogyakarta)
- Teguh Bharata Adji (Universitas Gadjah Mada)
- Teo Susnjak (Massey University, New Zealand)
- Tony Dwi Susanto (Institut Teknologi Sepuluh Nopember)
- Tutut Herawan (Universiti Malaysia Pahang)
- Ummi Rabaah Hashim (Universiti Teknikal Malaysia Melaka)
- Vimala Balakrishnan (Universiti Malaya)
- Waleed Ali Ahmed Abdullah (King Abdulaziz University, Arab Saudi)
- Wan Nurhayati Wan Ab Rahman (Universiti Putra Malaysia)
- Warusia Yassin (Universiti Teknikal Malaysia Melaka)
- Wawan Setiawan (Universitas Pendidikan Indonesia)
- Wendi Usino (Universitas Budi Luhur, Jakarta)
- Younes Saadi (Universiti Tun Hussein Onn Malaysia)
- Yudi Wibisono (Universitas Pendidikan Indonesia)
- Yusmadi Yah Jusoh (Universiti Putra Malaysia)
- Zakiah Ayop (Universiti Teknikal Malaysia Melaka)
- Zuwairie bin Ibrahim (Universiti Malaysia Pahang)

Organizing Committee

Patron
Rector of Universitas Pembangunan Nasional “Veteran” Yogyakarta

Advisor
- Vice Rector of Universitas Pembangunan Nasional “Veteran” Yogyakarta
- Dean of Faculty of Industrial Engineering Universitas Pembangunan Nasional “Veteran” Yogyakarta
- Head of Informatics Engineering Department Universitas Pembangunan Nasional “Veteran” Yogyakarta

General Chair
Awang Hendrianto Pratomo (Universitas Pembangunan Nasional “Veteran” Yogyakarta)

General Co-chair
- Haviluddin (Universitas Mulawarman)
- Aji Prasetya Wibawa (Universitas Negeri Malang)
- Bambang Yuwono (Universitas Pembangunan Nasional “Veteran” Yogyakarta)
- Hidayatullah Himawan, Universitas Pembangunan Nasional “Veteran” Yogyakarta

Secretary
- Oliver Samual Simanjuntak (Universitas Pembangunan Nasional “Veteran” Yogyakarta)
- Hari Prapcoyo (Universitas Pembangunan Nasional “Veteran” Yogyakarta)
- Rifki Indra Perwira (Universitas Pembangunan Nasional “Veteran” Yogyakarta)
- Sylvert Prian Tahalea (Universitas Pembangunan Nasional “Veteran” Yogyakarta)

Technical Program Chair
- Lala Septem Riza (Universitas Pendidikan Indonesia)
- Andri Pranolo (Universitas Ahmad Dahlan)
- Ramadiani (Universitas Mulawarman)
- Muhammad Syafrullah (Universitas Budi Luhur)
- Gunawan Ariyanto (Universitas Muhammadiyah Surakarta)
- Aji Prasetya Wibawa (Universitas Negeri Malang)
- Ummi Raba'ah Hashim (Universiti Teknikal Malaysia)
- Awang Hendrianto Pratomo (Universitas Pembangunan Nasional "Veteran" Yogyakarta)

Layout TPC
- Haviluddin Sukirno (Universitas Mulawarman)
- Yudi Wibisono (Universitas Pendidikan Indonesia)
- Hamdani (Universitas Mulawarman)
- Iwan Tri Riyadi Yanto (Universitas Ahmad Dahlan)
- Krisna Adiyarta (Universitas Budi Luhur)
- Adhi Prahara (Universitas Ahmad Dahlan)
- Oki Wicaksono (Universitas Mulawarman)
- Indra Riyanto (Universitas Budi Luhur)

Publication
- Simon Pulung Nugroho (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Budi Santoso (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Herlina Jayadianti (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Shoffan Saifullah (Universitas Pembangunan Nasional "Veteran" Yogyakarta)

Treasury, Financial, & Sponsorship
- Wilis Kaswidjanti (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Yuli Fauziah (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Vynska Amalia Permadi (Universitas Pembangunan Nasional "Veteran" Yogyakarta)

Publicity & Website
- Dessyanto Boedi Prasetyo (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Bagus Muhammad Akbar (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Frans Richard Kodong (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Ahmad Taufiq Akbar (Universitas Pembangunan Nasional "Veteran" Yogyakarta)

Registration, Certificate & Gift
- Herry Sofyan (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Yuli Fauziah (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Rochmat Husaini (Universitas Pembangunan Nasional "Veteran" Yogyakarta)

Technical, Facility & Logistic
- Agus Sasmito A. (Universitas Pembangunan Nasional "Veteran" Yogyakarta)\
- Andiko Putro Suryotomo (Universitas Pembangunan Nasional "Veteran" Yogyakarta)
- Riza Prapascatama Agusdin (Universitas Pembangunan Nasional "Veteran" Yogyakarta)

Event Management (Program & Protocol)
Herry Sofyan (Universitas Pembangunan Nasional "Veteran" Yogyakarta)

Keynote Speakers Biography

Leonel Eduardo Hernandez Collantes receive both his undergraduate degree in System Engineering and his postgraduate degree with Computer Networks specialization from Universidad del Norte, Barranquilla. He holds a masters degree in Strategic Management in Telecommunications from Universidad Internacional Iberoamericana. He has been a reviewer for scientific journal IJAIN since 2017 and an active member of IEEE Colombia section. He is an active researcher of the Telematics Engineering program, Faculty of Engineering, University Institution ITSA.

Assoc. Prof. Mohamad Shanudin Zakaria received his B.Sc. and M.Sc. titles from Northrop University, California. He holds a Ph.D. degree from University of Reading, United Kingdom. He is an Associate Professor in the Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia. He is part of a team taking a systemic approach and appllying ICT to the conservation of Lake Chini biodiversity and the improvement of its ecosystem. His research interest are application of System Thinking in ICT, architecting IT solutions, Business-IT alignment and ICT Governance.

Conference Schedule

Wednesday, October 23rd 2019

Time	Event
06:30 – 07:30	Hospitality & Registration Desk Open
07:30 – 09:00	Opening Ceremony
	1. National Anthem – Indonesia Raya
	2. Bela Negara Anthem
	3. Cultural Performance
	4. Welcome Address by ICSITech Chairperson
	5. Supporting Address by IEEE Indonesia Section
	6. Welcome Address by Rector of Universitas Pembangunan Nasional Veteran Yogyakarta
	7. Opening Ceremonial by Rector of Universitas Pembangunan Nasional Veteran Yogyakarta
	8. Photo Session
09:00 – 09:30	Coffee Break
09:30 – 12:00	Keynote Speech
	Keynote Speaker 1
	Keynote Speaker 2
	Moderator: Dr. Herlina Jayadianti, S.T., M.T.
12:00 – 13:00	Lunch and Prayer Time
13:00 – 15:00	Parallel Session 1
15:00 – 15:20	Coffee Break
15:20 – 17:20	Parallel Session 2
17:20 – 18.30	Break
18.30 – 21.00	Gala Dinner Session
	Best Paper Announcement
	Closing Ceremony
	Miscellaneous Information

Thursday, October 24th 2019

Yogyakarta Tour

Parallel Session Schedule

SAMAS Room – Artificial Intelligence Track

Time	Paper Title
13:00 – 13:20	**(#1570562661) Problem Event Extraction to Develop Causal Loop Representation from Texts** *Chaveevan Pechsiri, Intaka Piriyakul, Narongdech Keeratipranon*
13:20 – 13:40	**(#1570570061) An Analytical Study on Email Classification Using 10-Fold Cross-Validation** *Takorn Prexawanprasut, Piyanuch Chaipornkaew*
13:40 – 14:00	**(#1570572339) K-Nearest Neighbor (K-NN) based Missing Data Imputation** *Della Murbarani Prawidya Murti, Aji P Wibawa, Muhammad Akbar, Utomo Pujianto*
14:00 – 14:20	**(#1570570793) Seizure Detection Based on EEG Signals Using Katz Fractal and SVM Classifiers** *Inung Wijayanto, Achmad Rizal, Annisa Humairani*
14:20 – 14:40	**(#1570573510) Comparison of Performance Support Vector Machine Algorithm and Naïve Bayes for Diabetes Diagnosis** *Dominikus Boli Watomakin, Andi Wahju Rahardjo Emanuel*
14:40 – 15:00	**(#1570573730) Hybrid approach redefinition with SMOTE-CSELM in handling class imbalance problem** *Hartono Hartono, Erianto Ongko, Dahlan Adbullah*
15:00 – 15:20	**Coffee Break**
15:20 – 15:40	**(#1570574409) Adaptive Localization and Segmentation of Optic Disc using K-means and Active Contour** *Hanung Adi Nugroho, Augustine Herin Hutami, Eka Frannita, Rizki Nurfauzi*
15:40 – 16:00	**(#1570574423) A Scalable Approach for Improving Implementation of a Frequent Pattern Mining Algorithm using MapReduce Programming** *Md. Abed Hasan, Naima Hassan, Md. Hasibuzzaman, Mohammad Rezwanul Huq*
16:00 – 16:20	**(#1570583082) The Multi Layer Auto Encoder Neural Network (ML- AENN) for Encryption and Decryption of Text Message** *Achmad Gaffar, Arief Bramanto Wicaksono Putra, Rheo Malani*
16:20 – 16:40	**(#1570586194) An Effective Quantum Inspired Genetic Algorithm for Continuous Multiobjective Optimization** *Julisa Bana Abraham, Oyas Wahyunggoro, Noor Akhmad Setiawan*
16:40 – 17:00	**(#1570586348) Categorical Data Classification based on Fuzzy K-Nearest Neighbor Approach** *Heru C Rustamaji, Shalfa F Luhrie, Oliver Simanjuntak, Juwairiah Juwairiah, Bambang Yuwono*

PARANG TRITIS ROOM – Artificial Intelligence Track

Time	Paper Title
13:00 – 13:20	**(#1570568748) General Formula for Finding Optimize Hologram Reflector Dimension to Project Images on Screen** *Nur Ezha Vidawati, Syaiful Adala, Nugroho Pratama, Simon Nugroho*
13:20 – 13:40	**(#1570568879) Image-based processing for ripeness classification of oil palm fruit** *Anindita Septiarini, Hamdani Hamdani, Anita Ahmad Kasim*
13:40 – 14:00	**(#1570569209) Accurate and Low-cost Fingerprint Classification via Transfer Learning**

Time	Paper Title
	Latifah Listyalina, Ikhwan Mustiadi
14:00 – 14:20	**(#1570570191) Morphological Feature Extraction from Low Quality of Thick blood Smear Microphotographs** *Umi Salamah, Anto Nugroho, Ismail Eko Prayitno Rozi, Puji Budi Setia Asih*
14:20 – 14:40	**(#1570570214) Friendly and Progressive Visual Secret Sharing with Lossless Reconstruction** *Heri Prasetyo, Jing-Ming Guo, Hsia Chih-Hsien*
14:40 – 15:00	**(#1570570262) Blind Color Image Denoising Using Low Rank Matrix Minimization** *Heri Prasetyo, Umi Salamah, Wiranto Wiranto, Sari Sihwi, Winarno Winarno*
15:00 – 15:20	**Coffee Break**
15:20 – 15:40	**(#1570574888) Impact of Different Degree of Smoothing on Non-Local Means based Filter for Retinal Vessel Modeling** *Kartika Firdausy, Oyas Wahyunggoro, Hanung Adi Nugroho, Muhammad Bayu Sasongko, Risanuri Hidayat*
15:40 – 16:00	**(#1570584981) An Efficient Ear Recognition Based on Colour Texture and Edges Geometry** *Raad Ahmed Hadi*
16:00 – 16:20	**(#1570585570) Comparison of Egg Fertility Identification based on GLCM Feature Extraction using Backpropagation and K-means Clustering Algorithms** *Shoffan Saifullah, Vynska Amalia Permadi*
16:20 – 16:40	**(#1570585900) Plasmodium Parasites Detection in Digital Microscopic Images Containing of Luminance Noise** *Hanung Adi Nugroho, Rizki Nurfauzi, Eka Frannita*
16:40 – 17:00	**(#1570586492) Image Graph Matching Based on Region Adjacency Graph** *Akmal Akmal, Rinaldi Munir, Judhi Santoso*

PARANG KUSUMO ROOM – IoT and Big Data Track

Time	Paper Title
13:00 – 13:20	**(#1570562826) Implementation MQTT-SN Protocol on Smart City Application based Wireless Sensor Network** *M. Udin Harun Al Rasyid, Ferry Astika Saputra, Fadlul Fikri*
13:20 – 13:40	**(#1570586137) Smart Cane: Public Transportation Code Detection and Identification System for Visually Impaired** *Giva Andriana Mutiara, Gita Indah Hapsari, Kevin Suhandi*
13:40 – 14:00	**(#1570586147) Preventing Traffic Road Accident: Prototype of Monitoring Tires Failure Detection** *Giva Andriana Mutiara, Periyadi Periyadi, Mochamad Ghani Rachmatul Firdaus, Janfery Sinaga*
14:00 – 14:20	**(#1570586268) Feature Extraction of ECG Signals using Discrete Wavelet Transform and MFCC** *Siti Agrippina Alodia, Yusuf Risanuri Hidayat*
14:20 – 14:40	**(#1570586588) Anomaly-Based Intrusion Detection and Prevention Using Adaptive Boosting in Software-defined Network** *Rifki Indra Perwira, Yuli Fauziah, I Retya Mahendra, Dessyanto Boedi Prasetyo, Oliver Simanjuntak*
15:00 – 15:20	**Coffee Break**
15:20 – 15:40	**(#1570586677) Intrusion Detection System as Audit in IoT Infrastructure using Ensemble Learning and SMOTE Method**

Time	Paper Title
	Aldhi Ari Kurniawan, Heru Agus Santoso, Moch Arief Soeleman, Ahmad Zainul Fanani
15:40 – 16:00	**(#1570587190) Rule-Based Pronunciation Models to Handle OOV Words for Indonesian Automatic Speech Recognition System** *Fanda Yuliana Putri, Devin Hoesen, Dessi Lestari*
16:00 – 16:20	**(#1570567752) Performance of SVM in Classifying the Quartile of Computer Science Journals** *Sandika Maulana Putra, Aji P Wibawa, Triyanna Widiyaningtyas, Ilham A. E. Zaeni*
16:20 – 16:40	**(#1570570016) A Prediction Model for Human Happiness Using Machine Learning Techniques** *Piyanuch Chaipornkaew, Takorn Prexawanprasut*
16:40 – 17:00	**(#1570570293) Analysis Accuracy of Random Forest Model for Big Data - A Case Study of Claim Severity Prediction in Car Insurance** *Kartika Chandra Dewi, Hendri Murfi, Sarini Abdullah*
17:00 – 17:20	**(#1570573491) Performance Evaluation of Supervised Machine Learning Algorithms Using Different Data Set Sizes for Diabetes Prediction** *Melky Radja, Andi Wahju Rahardjo Emanuel*

BARON ROOM – Artificial Intelligence Track

Time	Paper Title
13:00 – 13:20	**(#1570586558) Weakening Feature Independence of Naïve Bayes Using Feature Weighting and Selection on Imbalanced Customer Review Data** *Reiza Adi Cahya, Fitra A. Bachtiar*
13:20 – 13:40	**(#1570586752) A Semiautomatic Sperm Cell Data Annotator for Convolutional Neural Network** *Priyanto Hidayatullah, Tati Erawati Rajab, Rinaldi Munir, Anggraini Barlian*
13:40 – 14:00	**(#1570586813) Improving Imbalanced Dataset Classification Using Oversampling and Gradient Boosting** *Nur Heri Cahyana, Siti Khomsah, Agus Sasmito Aribowo*
14:00 – 14:20	**(#1570574779) Comparison Performance of C4.5, Naïve Bayes and K- Nearest Neighbor in Determination Drug Rehabilitation** *Islamiyah Islamiyah, Nataniel Dengen, Anisa Afiyah, Medi Taruk*
14:20 – 14:40	**(#1570586627) Eyeball Movement Detection Using Sector Line Distance Approach and Learning Vector Quantization** *Gusti Pangestu, Fitra A. Bachtiar*
14:40 – 15:00	**(#1570586955) Enumeration of Plasmodium Parasites on Thin Blood Smear Digital Microscopic Images** *Alifia Prananda, Hanung Adi Nugroho, Igi Ardiyanto*
15:00 – 15:20	**Coffee Break**
15:20 – 15:40	**(#1570587166) Arnold Transformed Position Power First Mapping (AT- PPFM) for Secret Digital Image** *Wilis Kaswidjanti, Hidayatulah Himawan, Afra Oryza Mursita Dewi, Mangaras Yanu Florestiyanto, Rifki Indra Perwira*
15:40 – 16:00	**(#1570586958) Wound Classifications Of Breast Tissues with Electrical Impedance Spectroscopy (EIS): Comparison of LVQ and GA-LVQ** *Yoke K Arbawa, R. Adhitya Pisefty, Fitra A. Bachtiar*
16:00 – 16:20	**(#1570570315) Vowel Sound Analysis in the Indonesian Language using Multilevel Wavelet Entropy** *Risanuri Hidayat, Achmad Rizal, Agus Bejo, Sujoko Sumaryono*

Time	Paper Title
16:20 – 16:40	**(#1570570348) Analysis Accuracy of XGBoost Model for Multiclass Classification - A Case Study of Applicant Level Risk Prediction for Life Insurance** *Widya Fajar Mustika*
16:40 – 17:00	**(#1570583031) Cosine Similarity for Title and Abstract of Economic Journal Classification** *Putri Yuni Ristanti, Aji P Wibawa, Utomo Pujianto*
17:00 – 17:20	**(#1570586600) Handling Imbalance Issue in Hate Speech Classification using Sampling-based Methods** *Rathpisey Heng, Teguh Bharata Adji*

Table of Contents

Performance Evaluation of Supervised Machine Learning Algorithms Using Different Data Set Sizes for Diabetes Prediction

1st Melky Radja
Universitas Atma Jaya Yogyakarta
Yogyakarta, Indonesia 55281
Melky.radjastudy@gmail.com

2nd Andi Wahju Rahardjo Emanuel
Universitas Atma Jaya Yogyakarta
Yogyakarta, Indonesia 55281
andi.emanuel@uajy.ac.id

***Abstract*—** Data classification algorithm in machine learning is very helpful in analyzing a number of medical data with a large size and helps in making decisions to diagnose a disease. Not all supervised classification algorithms get accurate results in analyzing data sets. For this reason, testing the accuracy of each supervised classification algorithm is necessary, this can be used as a comparison in determining which types of algorithms are most accurate in measuring small amounts of data, and which algorithms are the most accurate in measuring large amounts of data. In this paper we will examine several classification algorithms including Naïve Bayes algorithms, functions (Support Vector Classifier algorithms), rules (decision table algorithms), trees (J48) by looking at the results of measurements made by each algorithm with measurement variables, which are Correctly Classified, incorrect classifieds, Precision, and Recall. The purpose of the study was to find the weaknesses and strengths of the supervised classification algorithm based on the measurement variables that have been determined against the testing of predictive databases of diabetes. Based on the results in this study, the best algorithm that can be used to help make a decision to diagnose a disease is the SVM algorithm with an accuracy value of 77.3%.

Keywords: Classification Algorithms; Machine Learning; Supervised Learning; Diabetes prediction; Data mining

I. INTRODUCTION

Machine learning is one of the fastest growing fields of computer science. The more amount of data available allows the use of machine learning to analyze each data that is quite complex. Understanding machine learning itself is a machine or program that is able to change the experience into expertise or knowledge by learning from input given to machines or programs [1]. Learning machines themselves can be used to assist decision-making such as in the field of health, and others [2]. The most commonly used machine learning application is data mining. There are several previous studies that tried to see the usefulness of data mining and machine learning processes in the world of health [3]. The reason for the application is due to the large size of the data in the medical world making it difficult for medical experts to process the data. For that the application of data mining in medical can help process data starting from classifying types of diseases, as well as helping to make decisions or diagnose diseases. Some of the data mining functions that are used are the process of finding interesting patterns from large amounts of stored data and the extraction of useful information from a collection of data that is not yet known. From every machine learning application in processing data mining, the important thing that makes a machine able to learn to see patterns in data is the learning technique. In data mining there are two types of learning techniques, namely supervised learning and unsupervised learning [4].

In the implementation of the supervised learning computational algorithm is needed to do the fastest search from a similar data set or classifier. The classifier algorithm in machine learning includes K-Nearest Neighbor, Naïve Bayes, Decision Tree, Regression, SVM and others. Each algorithm has a different level of accuracy in the process of classifying data. Not all algorithms will have good accuracy in performing classification with different data sizes [5]. In implementing the classifier algorithm in the medical field, especially in helping to make decisions about the diagnosis of disease. The type of classification that is used must be precise and the accuracy produced by the algorithm must be good, in order to avoid misdiagnosis of the disease. Based on these problems in this paper, we will try to compare each classifier algorithm. Comparisons made are expected to provide input regarding the weaknesses and strengths of each

algorithm in classifying data of different sizes in the classification data of diabetes.

II. LITERATURE REVIEW

Machine learning is an analytical method that can help process a data set using an algorithmic approach. In Machine Learning itself there are two methods of data analysis, namely unsupervised learning [6] and supervised learning. In this paper we will discuss one of the supervised learning techniques, namely classification techniques. In a previous journal reviewing the algorithm of classification [7] explained the accuracy of classification algorithms that had been previously developed by other researchers. In his journal It is said that the process of testing accuracy can be done through several stages such as split the training set, cross-validation, Leave-one-out validation. In other studies [8] explain the steps required in order to solve supervised learning are: determination of the type of training data, Gathering a training set, determination of the input feature representation of the learned function, determination of the various structures of the learned function and then devising the learning algorithm accordingly, completion of the design and evaluation of the accuracy of the learned function.

Classification is one of the tasks of decision making for human activities that are most often encountered [7]. Classification problems occur when an object needs to be assigned to a group or class that has been determined based on a number of observed attributes related to that object. For example medical diagnosis of certain diseases [9]. Some previous studies also compared several algorithms to test one dataset [10]. Research on breast cancer disease [11], grouping cancer based on the nature of the tumor by testing classification for decision making using three Bayes classification algorithms including, Bayes Belief Network (BBN), Tree Augmented Naive Bayes (TAN), and Boosted Augmented Naive Bayes (TIRE). The results obtained show the best classification accuracy for predicting cancer is by using the Tree Augmented Naive Bayes (TAN) algorithm. Another study [12] compared Naïve Bayes and J48 in a dataset to see the level of accuracy of sensitivity and specificity, where the results of the study stated that the measurement results of J48 were more efficient and accurate compared to Naïve Bayes. A related study on the machine learning classification algorithm to help make decisions in cases of diabetes, some of them have performed performance comparisons of the types of classification algorithms from, Naive Bayes, SVM, and Decision Tree using a single dataset [13]. The highest accuracy of the three classification algorithms tested is Naïve Bayes with an accuracy rate of 76.30% compared to SVM and Decision Tree. Most researchers comparing the results of each algorithm is to see how accurate the calculations are when using different algorithms in one case study.

In this paper we will discuss the results of research using different methods to see the performance of various types of classifications. The suggestion is to split the diabetes dataset into three parts with each section having a different size. The purpose of the dataset size divisions is to be able to see if there is a change in classification performance if the dataset is tested with different amounts of data.

III. THEORETICAL FRAMEWORK

A. Naïve Bayes algorithm

Naïve Bayes is a classification method based on the Bayes theorem that uses probability methods and statistics to predict future opportunities based on previous experience [14]. The naïve Bayes algorithm can be used for real-time prediction, text classification / spam filtering, and recommendation systems. The stage of the naïve Bayes algorithm process is counting the number of classes, calculating the number of class cases, multiplying all class variables and comparing the class results.

B. Support Vector Machine (SVM)

Support Vector Machine (SVM) classifier is a machine learning method that was first introduced by Vapnik [15]. SVM works based on the principles of Structural Risk Minimization with the aim of finding the best hyper plane that separates two classes at the input. SVM is very effective in solving linear classification problems [16]. The standard SVM takes a set of input data, and predicts, for each input given, the possibility of input is a member of one of the classes of the two existing classes, which makes an SVM a binary linear nonprobability classifier. Because SVM is a classifier, then given a training set, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether new data falls into a category or another.

C. Decision Table Algorithm

The Decision Table is also another description of the Decision Tree, which is visualized in a table where conditions and actions appear in decision columns

[17]. The decision table is a method to explain and describe the data flow logically that is used to solve a problem. Decision table works by combining all the conditions in the decision table, namely maximax (decisions that have the highest payoff regardless of the ground conditions), Maximin (Choosing decisions that have the maximum value and minimum payoffs), minimax regret (comparing each existing value in the table to find the best value), hurricane (determining the best value with the help of alpha values), and equal likelihood (determining the best value by entering the alpha value as needed).

D. Trees (J48)

J48 is a data classification method in machine learning. J48 is also a decision tree that has a hierarchical prediction mode by converting data into decision trees and decision rules. The slowness of the decision table is able to eliminate data that is considered not necessary to speed up the search process. The main idea of J48 is to divide data into range based on attribute values for items found in the training dataset. J48 classifies data through decision trees or through rules that result from classifier formation [18].

E. Measuring Classification Algorithms

In determining the right type of algorithm model the important thing to do is to look at the results of the calculation of confusion matrix which functions as a benchmark in seeing the accuracy obtained is good or not. Table one shows an example of the resulting confusion matrix:

Table 1: Confusion matrix

	Classified Positive	**classified Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

In this study several variables measuring the confusion matrix of the machine learning classification algorithms used are:

1. Accuracy

Accuracy is explained as the level of closeness between predictive value and actual value, or in other description the number of classifications performed correctly is done from the total sample test tested. The formula for calculating accuracy values is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision

Precision is the frequency or level of accuracy between the information requested by the user and the answer given by the system. In other words precision is the quality of data produced in the classification process. The following is the formula used to calculate precision:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

3. Recall

Recall is the system's success rate in rediscovering information. In other words Recall is a relevant result that results from the classification process. Following is the formula for calculating recall:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

4. F-Measure

F-Measure is a harmonic value or an average value of precision and recall values with a range value of 0 to 1. The following is an example of the calculation formula for F-measure:

$$F - Measure = 2.\frac{precision\ .recall}{precision + recall}$$

IV. RESEARCH METHODOLOGY

To perform an accuracy analysis of several supervised machine learning algorithms. The diabetes dataset is used for testing the accuracy of algorithms in large datasets where the dataset consists of 2 attribute classes, namely "1 = positive diabetes" and "0 negative diabetes" and with Number of Instances: 768 (500 = positive diabetes, 268 = negative diabetes). The comparison variable to be used is Correctly Classified Instances; incorrectly classified Instances; Weighted Avg Precision; Weighted Avg Recall; Weighted Avg F-Measure; ROC Area; Confusion Matrix. Figure 1 is a description of the research methodology of supervised machine learning classifier test:

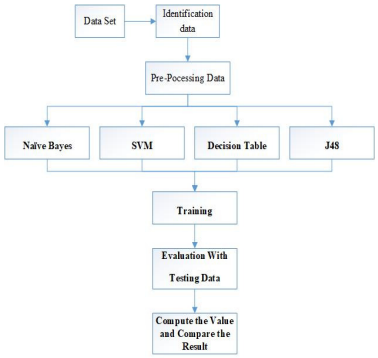

Figure 1: Methodology.

Supervised machine learning algorithms will be tested to make comparisons, to see the accuracy and precision of the algorithm using a diabetes dataset. On.

In figure 1 the pre-processing of the dataset will be divided into four. Among these datasets are 768 number of instances, 538 number of instances, 384 number of instances and 230 numbers of instances. Testing will be carried out four times using the different size of the diabetes dataset using four types of algorithmic classifications. The measurement objectives with the use of different sizes of training data are expected to be able to show the level of consistency in the calculation of algorithms at different dataset sizes. The classifier test in the WEKA application that can be used is K Folds Cross-validation. K Folds Cross-validation is a way to determine the best parameters of a model by testing the error size in testing data. In K Folds Cross-validation, data is divided into K parts with the same sample size. Then K-1 is used as training data and the remaining data as testing data. Process 10 Cross-validation will be tested 10 times with the same number of folds. In other words there is 10 subsets of data to evaluate algorithm performance. Figure 2 shows an example of how the 10 Cross-validation Classifier works.

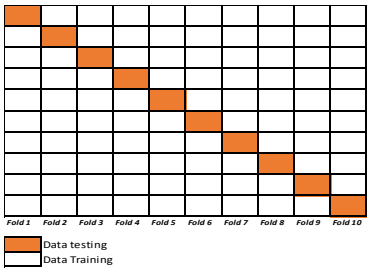

Figure 2: K Folds Cross-validation Method

V. RESULTS AND DISCUSSION

In this paper WEKA will be used as a tool to classify and compare machine learning algorithms. Table 2 shows the size of the diabetes dataset used to measure the performance / accuracy ratio of the four types of supervised classification algorithms.

Table 2: Dataset

DATA SET	Number of Instances	Number of Classifications	
		Tested Negative	Tested Positive
1	768	500	268
2	538	357	181
3	384	261	123
4	230	151	79

Table 3 shows the results of the comparison of the supervised classification algorithm with 768 Number of Instances.

Table 3: Comparison of classification algorithms with 768 Number of Instances

Algorithm	Correctly Classified	incorrectly classified	Precision of No	Precision of Yes	Recall of No	Recall of Yes	F-Measure	Classify Test	classification
Naïve Bayes	76.30% (586)	23.70% (182)	0.80	0.68	0.84	0.61	0.76	10 Folds cross-validation	**Bayes**
SVM (SMO)	**77.34 % (594)**	22.66 % (174)	0.79	0.74	0.90	0.54	0.76	10 Folds cross-validation	**functions**
decision table	71.22 % (547)	28.78 % (221)	0.76	0.60	0.81	0.53	0.71	10 Folds cross-validation	**Rule**
J48	73.83 % (567)	26.17 % (201)	0.79	0.63	0.81	0.60	0.74	10 Folds cross-validation	**Trees**

The results of the comparison of machine learning algorithms for datasets with large size or 768 number of instances are shown in table 3. The type of testing calcification used in WEKA is 10 Folds Cross-validation Method for testing diabetic datasets with 768 number of instances. In table 5, we can see the results of the dataset classification of 2 classes, namely the class "No (Tested-Negative)" and "Yes (Tested-Positive)" showing that, the Correctly Classified percentage or the accuracy of the proximity between the predicted values and the actual value of the SVM classification algorithm better than other algorithms, namely 77.34%. In Table 4 shows the results of the comparison of the supervised classification algorithm with 538 Number of Instances.

Table 4: Comparison of the classification algorithm with 538 Number of Instances

Algorithm	Correctly Classified	incorrectly classified	Precision of No	Precision of Yes	Recall of No	Recall of Yes	F-Measure	Classify Test	classification
Naïve Bayes	75.84 % (408)	24.16 % (130)	0.80	0.66	0.80	0.59	0.76	10 Folds cross-validation	**Bayes**
SVM (SMO)	**78.25 % (421)**	21.75 % (117)	0.79	0.76	0.91	0.52	**0.77**	10 Folds cross-validation	**functions**
decision table	74.35 % (400)	25.65 % (138)	0.78	0.64	0.85	0.54	0.74	10 Folds cross-validation	**Rule**
J48	72.30 % (389)	27.70 % (149)	0.78	0.59	0.81	0.56	0.72	10 Folds cross-validation	**Trees**

The results of the comparison of machine learning algorithms for datasets with 538 number of instances shown in table 4, have significant changes in the calculation of the four classification algorithms. The results of the accuracy of the SVM classification algorithm and the Decision Table have increased when testing datasets with 538 number of instances. Different from the Naïve Bayes and J48 classification algorithms which have decreased accuracy in testing using 538 number of instances. Overall the results of the comparison of accuracy in table 4, the best type of classification algorithm are SVM which successfully classifies 421 number of instances with an accuracy value of 78.25%. In Table 5 shows the results of a comparison of the supervised classification algorithm with 384 Number of Instances.

Table 5: Comparison of the classification algorithm with 384 Number of Instance

Algorithm	Correctly Classified	incorrectly classified	Precision of No	Precision of Yes	Recall of No	Recall of Yes	F-Measure	Classify Test	classification
naïve bayes	77.34 % (297)	22.66 % (87)	0.81	0.67	0.87	0.57	0.77	10 Folds cross-validation	**Bayes**
SVM (SMO)	**78.65 % (302)**	21.35 % (82)	0.80	0.73	0.91	0.53	**0.78**	10 Folds cross-validation	**functions**
decision table	77.08 % (296)	22.92 % (88)	0.81	0.67	0.87	0.56	0.77	10 Folds cross-validation	**Rule**
J48	74.22 % (285)	25.78 % (99)	0.80	0.60	0.83	0.56	0.74	10 Folds cross-validation	Trees

In the comparison results in table 6, the accuracy of the SVM classification algorithm is still the best with an accuracy of 78.65%. What distinguishes the results of the classification algorithm in table 5 compared to table 3 and table 4, the increase in the accuracy of the calculation of the SVM classification algorithm using 384 number of instances is not so significant, namely from 78.25% to 78.65%. Whereas in the Naïve Bayes and J48 classification algorithms which in the previous calculation results using datasets with 768 and 538 number of instances decrease, in the calculation using a dataset with 384 number of instances, the accuracy of the Naïve Bayes and J48 classification algorithms has increased. Table 6 shows the results of a comparison of the supervised classification algorithm with 230 Number of Instances.

Table 6: Comparison of the classification algorithm with 230 Number of Instances

Algorithm	Correctly Classified	incorrectly classified	Precision of No	Precision of Yes	Recall of No	Recall of Yes	F-Measure	Classify Test	classification
naïve bayes	**77.39 % (178)**	22.61 % (52)	0.83	0.67	0.83	0.67	0.77	10 Folds cross-validation	**Bayes**
SVM (SMO)	**77.39 % (178)**	22.61 % (52)	0.80	0.71	0.87	0.58	0.77	10 Folds cross-validation	**functions**
decision table	**77.39 % (178)**	22.61 % (52)	0.81	0.70	0.86	0.60	0.77	10 Folds cross-validation	**Rule**
J48	73.04 % (168)	26.96 % (62)	0.807	0.60	0.78	0.65	0.73	10 Folds cross-validation	Trees

Accuracy calculation results on datasets with 230 instances, Naïve Bayes, SVM, and decision table types have the same accuracy results of 77.39% or successfully classifying 178 instances. The difference is in precision, and recall, where the harmonic value or the highest average value of precision and recall value is in the Naïve Bayes algorithm, which is 0.77 higher compared to SVM or Decision Table. With the results of this calculation we can see that Naïve Bayes is more comparable than SVM in the process of classification of data with a small amount.

VI. ANALYSIS

Based on the results of a comparison of the four supervised classification algorithms on the diabetes dataset, the SVM algorithm has the best level of accuracy in testing with data sets of 768, 538, and 384 number of instance funds. While in testing the use of a dataset size of 230 number of instances, the J48 algorithm has the lowest accuracy compared to the other three algorithms. In testing the use of a dataset size of 230 numbers of instances the three algorithms other than J48 have the same accuracy results. Overall, in measuring accuracy in Figure 2, it can be concluded

that the naïve Bayes algorithm always shows good accuracy performance on small size data. This is shown in the results of the study, wherein each test the algorithm experienced an increase in accuracy.

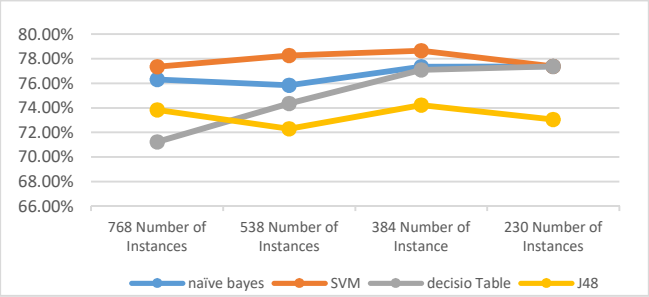

Figure 3: Calculation Accuracy

In figure 4: it can be seen that the measurement of F-Measure, which according to the function of F-Measure itself, which is the average value of precision and recall data, shows that in every test, accuracy cannot always be used to show the performance of an algorithm. Rather, it must require a measurement of the average value of precision and recall to support it. The statement is supported by the results of research where the same accuracy value in the naïve Bayes algorithm, SVM, and Decision table, in Figure 3 is different from the F-Measure value that is referred to in Figure 4. Each has a different F-Measure value although with the same accuracy value.

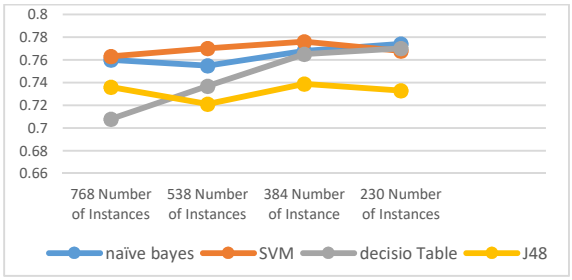

Figure 4: Calculation F-measure

VII. CONCLUSION

Based on the overall research that has been done with the use of measurement variables including accuracy, precision, recall, and F-Measure and experiments on one data type set with different sizes it was concluded that, the algorithm that can be used to help make a decision to the diagnosis of diabetes is the SVM algorithm. The reason for using four measurement variables and different types of data sets is that if you only use the level of accuracy in seeing the performance, it will be difficult to determine which performance is the best. Another reason is that there is a change in the ability of the classification algorithm in the data for each size of the data set being tested. Although overall the accuracy of the SVM algorithm is the best, in the case of the smallest size data experiment, the results of precision, recall, and F-Measure measurements on SVM are not very good compared to naïve Bayes and decision tables.

REFERENCES

[1] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*, vol. 9781107057. 2013.

[2] J. H. Phan, A. N. Young, and M. D. Wang, "OmniBiomarker: A web-based application for knowledge-driven biomarker identification," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3364–3367, 2013.

[3] H. C. Koh and G. Tan, "Data Mining Applications in Healthcare," vol. 19, no. 2, pp. 64–72.

[4] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," vol. 349, no. 6245, 2015.

[5] D. J. Hand, "Measuring classifier performance: A coherent alternative to the area under the ROC curve," *Mach. Learn.*, vol. 77, no. 1, pp. 103–123, 2009.

[6] R. Sathya and A. Abraham, "Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification," *Int. J. Adv. Res. Artif. Intell.*, vol. 2, no. 2, pp. 34–38, 2013.

[7] T. Oladipupo Ayodele, "Classification of Machine Learning Algorithms," *New Adv. Mach. Learn.*, vol. 3, no. 03, p. 30, 2016.

[8] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Supervised Machine Learning: A Review of Classification Techniques," *Artif. Intell. Rev.*, vol. 26, pp. 159–190, 2006.

[9] M. Fatima and M. Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic," *J. Intell. Learn. Syst. Appl.*, vol. 09, no. 01, pp. 1–16, 2017.

[10] B. F. Z. Al-Bayaty and S. Joshi, "Comparative Analysis between Naïve Bayes Algorithm and Decision Tree to Solve WSD Using Empirical Approach," *Lect. Notes Softw. Eng.*, vol. 4, no. 1, pp. 82–86, 2015.

[11] B. Banu A and P. Thirumalaikolundusubramanian, "Comparison of Bayes Classifiers for Breast Cancer Classification," *Asian Pac. J. Cancer Prev.*, vol. 19, no. 10, pp. 2917–2920, 2018.

[12] T. R. Patil and M. . Sherekar, "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification," *Int. J. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 256–261, 2013.

[13] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1578–1585, 2018.

[14] H. Zhang, "The Optimality of Naive Bayes," 2004.

[15] C. CORTES and V. VAPNIK, "Support-Vector Networks," *Mach. Leaming*, vol. 20, no. 3, pp. 273–297, 1995.

[16] Y. Yuan *et al.*, "SVM-based detection in visible light communications," *Optik (Stuttg).*, vol. 151, pp. 55–64, 2017.

[17] A. YK and B. I. Khodanpur, "Decision Table Based Testing," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 3, no. 3, pp. 1298–1301, 2015.

[18] G. Kaur and A. Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes," *Int. J. Comput. Appl.*, vol. 98, no. 22, pp. 13–17, 2014.

Certificate of Appreciation

This is certify that

Andi Wahju Rahardjo Emanuel

has participated as

Presenter

In 2019 5th International Conference
on Science in Information Technology (ICSITech)

Yogyakarta - Indonesia, 23-24 October 2019

General Chair of ICSITech 2019

Dr. Awang Hendrianto Pratomo, S.T., M.T.

Jointly organized with:

Telkom University · UTM · UPM · UTeM · ITSA