Organizer:

UTHM | Faculty of Electrical and Electronic Engineering

Technical Co-Sponsors:

IEEE SMC Systems, Man, and Cybernetics Society

IEEE

Co-Organizer:

YAMAGUCHI UNIVERSITY 山口大学

Strategic Partner:

UniKL UNIVERSITI KUALA LUMPUR

Sponsor:

ViTrox

icftss@uthm.edu.my

UTHM
Universiti Tun Hussein Onn Malaysia

ICFTSS

International Conference on Future Technologies for Smart Society

7-8 August 2024

UniKL MIIT,
Kuala Lumpur, Malaysia

icftss@uthm.edu.my

TABLE of CONTENT

AUTHOR INDEX

(according to the flrst name and format as it appears in the conference paper)

TITLE PAGE

No.	Paper ID	Title	Authors	Page Start
1	1570987705	IoT-based Air Quality Monitoring System with Air Filter	Khukentthra Narayanasamy; Sharifah Saon; Abd Kadir Mahamad; Maisara Othman	1
2	1570989258	Development of Drowsiness & Drunk Detection Using YOLOv8	Wei How Ho; Abd Kadir Mahamad; Sharifah Saon; Maisara Othman; Herdawatie Abdul Kadir; Hakkun Elmunsyah	6
3	1570991729	L1-Norm-Minimized Variable-Center-Frequency Bandpass Filters with Widened-Range Tunability	Tian-Bo Deng	12
4	1570997039	Object Recognition using Yolov8 for Car Seat Identification	Leong Chee Sum; Siti Zarina Binti Mohd. Muji; Suhaila Sari	17
5	1571001534	Privacy-Preserving Decision-Making for Digital Transformation with Divide-and-Conquer Data Aggregation	Mohd Anuaruddin Bin Ahmadon; Alireza Jolfaei	23
6	1571002068	A Compact and Semi-Flexible Dual-Band Patch Antenna for IoT-Based WBAN Applications	Umar Musa; Shaharil Mohd Shah; Huda A. Majid; Muhammad Sani Yahya; Zuhairiah Zainal Abidin; Shamsul Mohamad	28
7	1571005856	Model To Predict Flight Arrival Delay In John F. Kennedy Airport	Justin Gilbert; Chong Jian How	32
8	1571006047	Smart Pet House	Pey Ching Goo; Kim Gaik Tay; Chang Choon Chew; Audrey Huong	37
9	1571006506	Smart Irrigation Monitoring and Controlling System with the Internet of Things (IoT)	Nur Mieza Syamimi Ahmad Shuhaimi; Chang Choon Chew; Kim Gaik Tay; Kian Sek Tee	43
10	1571009631	Monitoring Melon Manis Terengganu (MMT) by Implementation of IoT Technology	Adriana Balqis Ali; Mohd Kamir Yusof; Zahratul Amani Zakaria	48
11	1571010325	Fuzzy Logic Controller-based Efficient Li-ion Battery Charging System for Electrical Vehicles	Khairul Ikmal Ibrahim; Nabihah Ahmad; Muhammad Anas Razali	52
12	1571012551	A Hybrid Method of N-Gram and Bag-of-Words (BoW) Models on the Assessment of Adolescent Personality Traits Through the Application of the Naïve Bayes Algorithm	Irzal Arief Wisky, IAW; Sarjon Defit; Gunadi Widi Nurcahyo	57
13	1571012554	Enhancing Classification Performance: A Study on SMOTE and Ensemble Learning Techniques	Nadya Alinda Rahmi, NAR; Sarjon Defit; Okfalisa Okfalisa	63
14	1571012593	Monitoring and Reporting Unethical Actions in Higher Education using the BiLSTM and BiGRU algorithms	Lucky Lhaura Van FC; Sarjon Defit; Saiful Bukhori	69

No.	Paper ID	Title	Authors	Page Start
15	1571014992	Improvement of Machine Learning Algorithms with Hyperparameter Tuning on Various Datasets	Akmar Efendi; Iskandar Fitri; Gunadi Widi Nurcahyo	75
16	1571015442	IoT-Based Portable Health Monitoring System for Elderly Patients	Harrish Kumar Letchumanan; Chang Choon Chew; Kim Gaik Tay	80
17	1571016130	Proposal of Programming Education Content for Elementary School Teachers Using Drone	Jun Sakurai; Minoru Kobayashi	86
18	1571016397	Performance Improvement of Machine Learning Algorithm Using Ensemble Method on Text Mining	M Khairul Anam; Muhammad Bambang Firdaus; Fadli Suandi; Lathifah Lathifah; Torkis Nasution; Sofiansyah Fadli	90
19	1571016685	A Lifting-Transferring Assistive Device Between Wheelchair and Car Seat for Disabled	Toong Hai Sam; Kian Sek Tee; Whee Yen Wong; Chang Choon Chew; Chin Fhong Soon	96
20	1571019905	DDoS Anomaly Detection in Software-Defined Networks: An Evaluation of Machine Learning Techniques for Traffic Classification and Prediction	Abdinasir Hirsi; Lukman Audah; Adeeb Salh; Nan Sahar; Salman Ahmed; Mohammed A Alhartomi	100
21	1571020609	Enhanced Diagnostic of Pulmonary Embolism Detection using DenseNet and XGBoost	Zahoor Ahmad; Somaya Ali Al-maadeed; Muhammad Asif Khan	106
22	1571021374	An Enhanced Probabilistic Elastic Net Regression Model (EPERM) for Heart Disease Prediction	G. Vennira Selvi Selvi; T Ganesh Kumar; Shajilin Loret; K Sampath Kumar; Anitha Julian; Priti Rishi	112
23	1571024199	Enhancing Cooking Recipe Browsing Experience through Integration of Recipes and Music	Koki Urakawa; Yuanyuan Wang	117
24	1571025096	Designing Optimal Secure Hardware IPs for Consumer Electronics using Fingerprint Watermark and Evolutionary Algorithm Framework	Anirban Sengupta; Vishal Chourasia; Nilesh Kumar; Praveen Dawar	123
25	1571026947	A Retrieval-based Chatbot for Customer Service Using Large Language Model	Pei-Chi Huang; Mei-Lien Pan	128
26	1571027322	Generating Appreciation Contents of Artworks for Creativity with Paintings and Music	Takumi Miyamoto; Yuanyuan Wang	132
27	1571027605	Indoor Navigation Using Augmented Reality with A* Algorithm	Shazmin Sariman; Maisara Othman; Sharifah Saon; Abd Kadir Mahamad; Herdawatie Abdul Kadir; Vignan Ramamurthi	136
28	1571027615	Optimization Caching Performance in Named Data Networking using Cache Size Allocation Algorithm	Ridha Muldina Negara; Nana Rachmana Syambas; Eueung Mulyana; Alesandora Emanuella Pinto	141

No.	Paper ID	Title	Authors	Page Start
29	1571030247	IoT/WSN-based Security/Privacy Methods for Forest Monitoring	Asmarashid Ponniran; Ashok Vajravelu; Wan Suhaimizan Wan Zaki; Yamunarani Thanikachalam; Syed Riyaz Ahammed; Sivaranjani S	146
30	1571030391	Daylight Illuminance Levels IIoT Monitoring System	Mohd Faiz Adnan; Khairul Rijal Wagiman; Mohd Noor Abdullah; Nur Izzah Izzati Norli; Abd Rahim Kasiman	152
31	1571030756	Expert System Development Diagnosis of Rabbit Disease Using Case-Based Reasoning Method and Certainty Factor Method	Abdel Rahman Al Ali Al Ali; Reta Diana Pradani; Danial Md Nor; Fahmy Rinanda Saputri; Khairun Nidzam Ramli; Mohd Nazri Mahmud	157
32	1571034242	An Antibody-Mediated Immune Mechanism in Botnet Defense System	Shingo Yamaguchi	161
33	1571034399	Visual Deception: Demonstrating Spoofing Attacks on Autonomous Vehicle Cameras	Muhammad Asif Khan; Hamid Menouar; Ayah Nassar; Mohamed M. Abdallah	165
34	1571034712	Comparative Analysis of RAG, Fine-Tuning, and Prompt Engineering in Chatbot Development	Harshit Kumar Chaubey; Gaurav Tripathi; Rajnish Ranjan; Srinivasa k. Gopalaiyengar	169
35	1571034714	Utilizing Attention-based Vision Transformer (AtViT) for Brain Tumour Detection	Harshit Kumar Chaubey; Siri Arelli; Karan Agrawal; Srinivasa k. Gopalaiyengar	173
36	1571034844	Leveraging Trusted Input Framework to Correct and Forecast the Electricity Load in Smart City Zones Against Adversarial Attacks	Mohd Hafizuddin Bin Kamilin; Shingo Yamaguchi; Mohd Anuaruddin Bin Ahmadon	177
37	1571035163	Analyzing Classification Models: Random Forest vs. Neural Networks in Health Prediction	Nazirul Faris Md Zahidi; Nurul Huda Ismail; Zarina Tukiran	183
38	1571035446	Bridging Cybersecurity Education and Industry Demands: Mapping and Prioritizing Curriculum Guidelines	Muhammad Asfand Yar; Hock Guan Goh; Kiran Adnan; Ming-Lee Gan; Vasaki Ponnusamy	188

KEYNOTE 1:
Dr. Yuichi Nakamura

NEC Corporation, Japan

DX AND DATA LINKAGE PLATFORM USE CASES IN SMART CITY PROJECTS

Abstract - One major advantage of digital transformation (DX) is the ability to help create new services through the extraction of more detailed and essential information through the Integration across digitized data items of heterogeneous origins and of different kinds. Realizing this inevitably requires a mechanism for linking heterogeneous data items from different sources for efficient utilization. A smart city is a technologically modern urban area that uses different types of electronic methods and sensors to collect significant big data. Collected data and data analysis in smart cities could change efficiency of the city, remarkably. This talk introduces an open platform for linking data items of many different kinds and reports its usage in a couple of smart city projects. Additionally, the potential of quantum computers as an advanced component of DX in smart city projects is presented.

Speaker Bibliography - Dr. Yuichi Nakamura received his B.E. in information engineering and M.E. in electrical engineering from the Tokyo Institute of Technology in 1986 and 1988, respectively. He received his PhD. from the Graduate School of Information, Production and Systems, Waseda University, in 2007. He joined NEC Corp. in 1988 and he led NEC's research about embedded system design, HPC, communication systems and quantum technologies as a general manager and a vice president of NEC research and development. Currently, he is a senior executive professional of NEC Corp. He is also a guest professor at Kyushu University, Waseda University and Tokyo University and an invited professor at Osaka University. He has more than 30 years of professional experience in electronic design automation, signal processing, supercomputer design, combinational optimization and quantum computing.

KEYNOTE 2:
Assoc. Prof. Dr. Afandi Ahmad

Provost, UTHM Pagoh Branch Campus

FUTURE-PROOF ENGINEERING EDUCATION TOWARDS SMART SOCIETY

Abstract - Whilst the concept of a smart society is widely discussed, it continues to encounter numerous macro and micro challenges. Emerging trends and technologies, such as artificial intelligence, the Internet of Things (IoT), and smart infrastructure, are transforming industries and societies, presenting both opportunities and obstacles. To achieve the vision of a smart society, which requires a cross-disciplinary approach, higher education—particularly engineering education—plays a crucial role. This involves integrating the design, implementation, and evaluation of science and technology applications into educational frameworks.

This keynote session will explore five (5) core pillars essential for future-proofing engineering education in the context of a smart society:

Student-centered Learning: Fostering personalized and engaging learning experiences tailored to individual student needs.

Contextual Practice: Embedding real-world applications and industry practices within the curriculum to ensure relevance and applicability.

Professional Competency: Equipping graduates with the necessary skills and knowledge to excel in their professional careers.

Digital Learning: Utilising advanced digital technologies to enhance both teaching and learning processes.

Value-based Education: Promoting ethical and sustainable values to prepare engineers for responsible innovation.

Although this session focuses on engineering education, the proposed ideas and principles can also benefit and be applied to other disciplines. By adopting these core pillars, various fields of study can enhance their educational frameworks, ensuring that graduates are well-equipped to contribute to the development of a smart society. Attendees will gain valuable insights into how educational institutions can proactively adapt to the evolving demands of a smart society. The session aims to cultivate a new generation of engineers who are not only technically skilled but also visionary leaders capable of driving transformative change.

Speaker Bibliography - Afandi Ahmad received a PhD in computer engineering from Brunel University, London U.K., in 2010 with various awards, and specialised in reconfigurable computing architecture for medical diagnostic imaging. He is an associate professor in the computer engineering cluster at the Department of Electronic Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia. Afandi is currently the Provost at the UTHM Pagoh Campus. Previously, he contributed a significant role in institutional leadership as Assistant Vice Chancellor (Strategic Planning and Corporate Relations) and Deputy Vice Chancellor (Student Affairs and Alumni).

Afandi has published one book, 71 journal articles and conference proceedings, and 35 articles in the national mass media. He has also received 48 awards in leadership and services, teaching and learning, and research and development. He has also successfully supervised students and completed their research with high-impact industrial recognition. Afandi recently completed his sabbatical attachment as a visiting scholar in the discipline of institutional leadership and higher education policy at the Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, Aalborg-UNESCO Centre for Problem-Based Learning in Engineering Science & Sustainability, Aalborg University, Denmark and the Teaching & Learning Laboratory, Faculty of Engineering & IT, The University of Melbourne, Australia.

Invited Speaker 1:
Datuk Ir. Dr. Leong Siew Mun

SENA Traffic System Sdn Bhd

THE FUTURE AND RELIABLE INTELLIGENT TRANSPORT SYSTEM

Abstract -Intelligent Transport System or ITS has been developed for more than few decades. It started with simple computerized traffic control system and progressed into an integrated ICT system. With the advent of cloud computing technology, big data analytics, internet of things and mobile applications in this age of digital revolution, ITS underwent further with tremendous new transformation. This transformation process brings many positive changes to the realm of ITS and has changed the landscape and eco-system of ITS. This transformation process is still ongoing with new technologies emerging. What then could this lead to and what future ITS can we expect? Will the future ITS with all new emerging disruptive technologies be reliable? This paper attempts to address these issues by outlining the path towards future ITS and explore what is the likely focus of the future ITS. At the same time, this paper will also look at the factors affecting reliability how the viscous cycle of major critical factors needs to be addressed.

Speaker Bibliography - Datuk Leong Siew Mun is a Professional Engineer and Chartered Engineer. He obtained his Doctor of Philosophy from Renssellear Polytechnic Institute, USA, specializing in the field of intelligent transportation system (ITS) and its application in urban management. He was the former Executive Director of City Hall Kuala Lumpur and has accumulated more than 35 years of project management and road transport infrastructure planning and operation experience.

He is currently the visiting Professor and chair holder of the University Tun Hussein Onn Malaysia - SENA Smart City Research & Development Chair as well as the Adjunct Professor for the Civil Engineering Department of University Malaya. He is also the group technical advisor for SENA group of companies. He is actively involved in Asia Pacific ITS activities locally or abroad. Besides as one of the board members of Asia Pacific ITS and ITS Malaysia, he also holds other portfolio in many professional institutions. As the Chairman of the National Committee on ISO/TC204 for Malaysia, he leads a committee responsible for the standardization of ITS standards for Malaysia.

Invited Speaker 2:
Mrs Karen Savarimuthu

Head of Academia Development
ViTrox Technologies Sdn Bhd

REVOLUTIONIZING INDUSTRIES, THE TRANSFORMATIVE POWER OF SMART SYSTEMS IN THE IOT ERA

Abstract - The convergence of smart systems and the Internet of Things (IoT) heralds a revolutionary epoch of technological advancement, poised to redefine industries and elevate daily living to unprecedented heights. This presentation will unravel the foundational principles and cutting-edge innovations within smart systems designed for IoT applications. We will navigate through the intricate architecture of IoT, emphasizing seamless connectivity and data symbiosis between devices. By showcasing transformative real-world applications, we will illustrate the profound impact of smart systems across diverse sectors such as agriculture, smart cities, and industrial automation. We will delve into pioneering topics, including the deployment of advanced sensors, the integration of edge computing, and the infusion of artificial intelligence to forge systems that are not only responsive but also adaptive, enabling real-time decision-making. Moreover, this presentation will offer a glimpse into future trajectories and the potential global impact of smart IoT systems on society and the economy. Embark on this journey with us to uncover the transformative power of smart systems in the IoT realm, driving a wave of innovation and efficiency in our interconnected world.

Speaker Bibliography - Karen graduated with honors in Information Technology from Harris University, United States, and earned an MBA from MMU. With 24 years of experience, she has developed expertise in information technology, customer experience management, data center management, and Industry 4.0 solutions. Her career began at Standard Chartered Bank, UK, and includes roles at MIMOS Berhad, JARING Communications Sdn Bhd, and ViTrox Corporations Berhad.

Karen's expertise covers big data, network management, cloud computing, smart factory solutions, and academic development. At V-ONE, she focuses on strategic academia development, forming cross-industry alliances, and preparing professionals for future technological advancements. She also has extensive experience in consulting, training, and developing technology solutions, with a strong commitment to advancing Industry 4.0 and addressing future business challenges.

ORGANIZING COMMITTEE

Patron	:	Prof. Dato' Ir. Ts. Dr. Ruzairi bin Abdul Rahim
Committee Advisors		Assoc. Prof. Ts. Dr. Asmarashid bin Ponniran
General Chair I	:	Assoc. Prof. Ir. Dr. Abd Kadir Mahamad
General Chair II	:	Prof. Dr. Shingo Yamaguchi (Japan)
Executive Conference Chair I (SMC Malaysia Chapter Vice Chair)	:	Mohd Helmy Abd Wahab
Executive Conference Chair II	:	Prof. Dr. Yoshinobu Tamura (Japan)
Technical Program Committee (TPC)	:	Dr. Mohd Anuaruddin Ahmadon (Japan)
		Assoc. Prof. Ir. Ts. Dr. Mohd Razali Md Tomari
Secretariat	:	Ts. Sharifah Saon
Publication	:	Dr. Yuanyuan Wang (Japan) - Chair
		Dr. Tasiransurini Ab Rahman - Chair
		Zarina Tukiran
		Assoc. Prof. Dr. Danial Md Nor
		Mohd Helmy Abd Wahab
Finance & Registration	:	Aizan Ubin – Chair
		Dr. Rohaida Mat Akir
		Dr Roziah Aziz
Publicity	:	Dr. Lei Zhou (Japan) - Chair
		Assoc. Prof. Dr. Maisara Othman - Chair
		Dr. Mohd Anuaruddin Ahmadon (Japan)
		Dr. Siti Hajar Aminah Ali
		Rosley Sawarno
		Aidi Basar
Relations	:	Assoc. Prof. Dr. Zuhairiah Zainal Abidin – Chair
		Dr. Shaharil Mohd Shah
		Assoc. Prof. Dr. Herdawatie Abdul Kadir
Technical	:	Assoc. Prof. Dr. Shamsul Aizam Zulkifli - Chair
		Dr. Muhammad Asif Khan (Qatar) – Co-Chair
		Dr. Ariffuddin Joret
Logistic	:	Dr. Nabihah @ Nornabihah Ahmad – Chair
		Dr. Jabbar al-Fattah Yahaya
Event & Protocol	:	Dr. Mariyam Jamilah Homam – Chair
		Dr. Shamsul Mohamad
		Assoc. Prof. Dr. Yee See Khee

eCF Paper Id	Paper / Article Title	Author Name	Author Email	Authorized Signatory Name	Authorized Signatory Email	Date of Signature	Copyright Type	A/V Recordings Permitted	OA Payment Received
1570987705	IoT-based Air Quality Monitoring System with Air Filter	Mr. Khukentthra Narayanasamy, Mrs. Sharifah Saon, Dr. Abd Kadir Mahamad and Ms. Maisara Othman	khukenn@gmail.com, sharifa@uthm.edu.my, kadir@uthm.edu.my, maisara@uthm.edu.my			2024-07-24	IEEE	Y	
1571035446	Bridging Cybersecurity Education and Industry Demands: Mapping and Prioritizing Curriculum Guidelines	Mr. Muhammad Asfand Yar, Dr. Hock Guan Goh, Dr. Kiran Adnan, Dr. Ming-Lee Gan and Dr. Vasaki Ponnusamy	asfandyar.yar@gmail.com, gohhg@utar.edu.my, kiranadnan@utar.edu.my, ganml@utar.edu.my, vasaki.ponnusamy@gmail.com			2024-07-10	IEEE	Y	
1570991729	L1-Norm-Minimized Variable-Center-Frequency Bandpass Filters with Widened-Range Tunability	Prof. Tian-Bo Deng	deng@is.sci.toho-u.ac.jp			2024-07-02	IEEE	Y	
1570997039	Object Recognition using Yolov8 fo Car Seat Identification	Leong Chee Sum, Dr. Siti Zarina Binti Mohd. muji and Mrs. Suhaila Sari	leongcheesum7788@gmail.com, szarina@uthm.edu.my, suhailas@uthm.edu.my			2024-07-10	IEEE	Y	
1571001534	Privacy-Preserving Decision-Making for Digital Transformation with Divide-and-Conquer Data Aggregation	Dr. Mohd Anuaruddin Bin Ahmadon and Dr. Alireza Jolfaei	anuar@yamaguchi-u.ac.jp, alireza.jolfaei@flinders.edu.au			2024-06-30	IEEE	Y	
1571002068	A Compact and Semi-Flexible Dual-Band Patch Antenna for IoT-Based WBAN Applications	Mr. Umar Musa, Dr. Shaharil Mohd Shah, Dr. Huda A. Majid, Dr. Zuhairiah Zainal Abidin, Dr. Shamsul Mohamad and Mr. Muhammad Sani Yahya	umusa.ele@buk.edu.ng, shaharil@uthm.edu.my, mhuda@uthm.edu.my, zuhairia@uthm.edu.my, shamsulm@uthm.edu.my, muhammad_22000359@utp.edu.my			2024-07-10	IEEE	Y	
1571005856	Model To Predict Flight Arrival Delay In John F. Kennedy Airport	Mr. Justin Gilbert and Mr. Chong Jian How	justin.gilbert@apu.edu.my, tp058176@mail.apu.edu.my			2024-07-10	IEEE	Y	
1571006047	Smart Pet House	Pey Ching Goo, Dr. Kim Gaik Tay, Dr. Chang Choon Chew and Dr. Audrey Huong	taykimgaik@gmail.com, tay@uthm.edu.my, chewcc@uthm.edu.my, audrey@uthm.edu.my			2024-07-03	IEEE	Y	
1571006506	Smart Irrigation Monitoring and Controlling System with the Internet of Things (IoT)	Ms. Nur Mieza Syamimi Ahmad Shuhaimi, Dr. Chang Choon Chew, Dr. Kim Gaik Tay and Dr. Kian Sek Tee	nurmieza43@gmail.com, chewcc@uthm.edu.my, tay@uthm.edu.my, tee@uthm.edu.my			2024-07-02	IEEE	Y	
1571009631	Monitoring Melon Manis Terengganu (MMT) by Implementation of IoT Technology	Ms. Adriana Balqis Ali, Dr. Mohd Kamir Yusof and Mrs. Zahratul Amani Zakaria	adrianabalqis0701@gmail.com, mohdkamir@unisza.edu.my, zahrahtulamani@unisza.edu.my			2024-08-14	IEEE	Y	
1571010325	Investigation of Vehicle Battery Charging Speed via Voltage Control	Dr. Nabihah Ahmad, Mr. Khairul Ikmal Ibrahim and Dr. Muhammad Anas Razali	nabihah@uthm.edu.my, ge230020@student.uthm.edu.my, anas@uthm.edu.my			2024-07-10	IEEE	Y	
1571012551	A Hybrid Method of N-Gram and Bag-of-Words (BoW) Models on the Assessment of Adolescent Personality Traits Through the Application of the Naïve Bayes Algorithm	Mr. Irzal Arief Wisky, IAW, Prof. Sarjon Defit and Mr. Gunadi Widi Nurcahyo	irzal_ariefwisky@upiyptk.ac.id, sarjon_defit@upiyptk.ac.id			2024-07-08	IEEE	Y	
1571012554	Enhancing Classification Performance: A Study on SMOTE and Ensemble Learning Techniques	Mrs. Nadya Alinda Rahmi, NAR, Prof. Sarjon Defit and Prof. Okfalisa Okfalisa	nadyaalindaa@upiyptk.ac.id, sarjon_defit@upiyptk.ac.id, okfalisa@gmail.com			2024-07-08	IEEE	Y	
1571012593	Monitoring and Reporting Unethical Actions in Higher Education using the BiLSTM and BiGRU algorithms	Mrs. Lucky Lhaura Van FC, Prof. Sarjon Defit and Mr. Saiful Bukhori	lucky@unilak.ac.id, sarjon_defit@upiyptk.ac.id, saiful.ilkom@unej.ac.id			2024-07-10	IEEE	Y	

eCF Paper Id	Paper / Article Title	Author Name	Author Email	Authorized Signatory Name	Authorized Signatory Email	Date of Signature	Copyright Type	A/V Recordings Permitted	OA Payment Received
1571014992	Classification Techniques: Evaluating Performance on Diverse Datasets	Mr. Akmar Efendi, Iskandar Fitri and Mr. Gunadi Widi Nurcahyo	akmarefendi@eng.uir.ac.id, if@upiyptk.ac.id			2024-07-08	IEEE	Y	
1571015442	IoT-Based Portable Health Monitoring System for Elderly Patients	Mr. Harrish Kumar Letchumanan, Dr. Chang Choon Chew and Dr. Kim Gaik Tay	haarishkumar24@gmail.com, chewcc@uthm.edu.my, tay@uthm.edu.my			2024-07-14	IEEE	Y	
1571016130	Proposal of Programming Education Content for Elementary School Teachers Using Drone	Dr. Jun Sakurai and Prof. Minoru Kobayashi	sakuraij@bunkyo.ac.jp, mkoba98@bunkyo.ac.jp			2024-08-11	IEEE	Y	
1571016397	Performance Improvement of machine Learning Algorithm Using Ensemble Method on Text Mining	Mr. M Khairul Anam, Mr. Muhammad Bambang Firdaus, Mr. Fadli Suandi, Mrs. Lathifah Lathifah, Mr. Torkis Nasution and Mr. Sofiansyah Fadli	khairulanam@sar.ac.id, bambangf@unmul.ac.id, fadli.suandi@polibatam.ac.id, lathifah@teknokrat.ac.id, torkisnasution@sar.ac.id, sofiansyah182@gmail.com			2024-07-04	IEEE	Y	
1571016685	A Lifting-Transferring Assistive Device Between Wheelchair and Car Seat for Disabled	Dr. Kian Sek Tee, Dr. Toong Hai Sam, Dr. Chang Choon Chew, Dr. Chin Fhong Soon and Dr. Whee Yen Wong	tee@uthm.edu.my, toonghai.sam@newinti.edu.my, chewcc@uthm.edu.my, soon@uthm.edu.my, winnie-wwy@hotmail.com			2024-07-10	IEEE	Y	
1571019905	DDoS Anomaly Detection in Software-Defined Networks: An Evaluation of Machine Learning Techniques for Traffic Classification and Prediction	Mr. Abdinasir Hirsi Abdi and Dr. Lukman Audah	engr.abdinasir@gmail.com, hanif@uthm.edu.my			2024-07-09	IEEE	Y	

eCF Paper Id	Paper / Article Title	Author Name	Author Email	Authorized Signatory Name	Authorized Signatory Email	Date of Signature	Copyright Type	A/V Recordings Permitted	OA Payment Received
1571020609	Enhanced Diagnostic of Pulmonary Embolism Detection using DenseNet and XGBoost	Mr. Zahoor Ahmad, Dr. Somaya Ali Al-maadeed and Dr. Muhammad Asif Khan	za2211162@qu.edu.qa, s_alali@qu.edu.qa, asifk@ieee.org			2024-07-15	IEEE	Y	
1571021374	An Enhanced Probabilistic Elastic Net Regression Model (EPERM) for Heart Disease Prediction	Mrs. G. Vennira selvi Selvi, Dr. T Ganesh Kumar, Dr. Shajilin Loret, Dr. K Sampath Kumar, Dr. Anitha Julian and Dr. Priti Rishi	vennira@yahoo.co.in, t.ganesh@galgotiasuniversity.edu.in, shajilinphd@gmail.com, ksampathkumara@gmail.com, cse.anithajulian@gmail.com, drpritirishiphd@gmail.com			2024-08-14	IEEE	Y	
1571024199	Enhancing Cooking Recipe Browsing Experience through Integration of Recipes and Music	Mr. Koki Urakawa and Dr. Yuanyuan Wang	e087vgv@yamaguchi-u.ac.jp, circl.wang@gmail.com			2024-07-02	IEEE	Y	
1571025096	Designing Optimal Secure Hardware IPs for Consumer Electronics using Fingerprint Watermark and Evolutionary Algorithm Framework	Prof. Anirban Sengupta, Mr. Vishal Chourasia, Mr. Nilesh Kumar and Mr. Praveen Dawar	asengupt@iiti.ac.in, vishalchourasia@iiti.ac.in, cse210001043@iiti.ac.in, cse210001054@iiti.ac.in			2024-06-30	IEEE	Y	
1571026947	A Retrieval-based Chatbot for Customer Service Using Large Language Model	Ms. Pei-Chi Huang and Dr. Mei-Lien Pan	peggy1210.huang@gmail.com, mlpan66@nycu.edu.tw			2024-07-12	IEEE	Y	
1571027322	Generating Appreciation Contents of Artworks for Creativity with Paintings and Music	Mr. Takumi Miyamoto and Dr. Yuanyuan Wang	d109vgu@yamaguchi-u.ac.jp, circl.wang@gmail.com			2024-07-10	IEEE	Y	
1571027605	Indoor Navigation with Augmented Reality with A* algorithm	Mr. Shazmin Sariman, Ms. Maisara Othman, Mrs. Sharifah Saon, Dr. Herdawatie Abdul Kadir, Dr. Abd Kadir Mahamad and Mr. Vignan Ramamurthi	ge220034@student.uthm.edu.my, maisara@uthm.edu.my, sharifa@uthm.edu.my, watie@uthm.edu.my, kadir@uthm.edu.my, de190011@student.uthm.edu.my			2024-07-11	IEEE	Y	
1571027615	Optimization Caching Performance in Named Data Networking using Cache Size Allocation Algorithm	Mrs. Ridha Muldina Negara, Dr. Nana Rachmana Syambas, Mr. Eueung Mulyana and Ms. Alesandora Emanuella Pinto	ridhanegara@telkomuniversity.ac.id, nana@stei.itb.ac.id, eueung@stei.itb.ac.id, alesandoraemanuella@student.telkom university.ac.id			2024-07-10	IEEE	Y	
1571030247	IoT/WSN-based security/privacy methods for forest monitoring	Dr. Ashok Vajravelu, Dr. Asmarashid Bin Ponniran, Dr. Wan Suhaimizan Bin Wan Zaki, Mrs. Yamunarani Thanikachalam, Dr. Syed Riyaz Ahammed and Mrs. Sivaranjani S	ashok@uthm.edu.my, asmar@uthm.edu.my, mizan@uthm.edu.my, yamunabme@gmail.com, syedriyaz.global@gmail.com, swame12@gmail.com			2024-07-14	IEEE	Y	
1571030391	Daylight Illuminance Levels IIoT Monitoring System	Mr. Mohd Faiz Adnan, Dr. Khairul Rijal Wagiman, Dr. Mohd Noor Abdullah, Ms. Nur Izzah Izzati Norli and Mr. Abd Rahim Kasiman	faizma@uthm.edu.my, khairulrw@uthm.edu.my, mnoor@uthm.edu.my, cb200065@student.uthm.edu.my, arahim@uthm.edu.my			2024-07-01	IEEE	Y	
1571030756	Expert System Development Diagnosis of Rabbit Disease Using Case Base Reasoning Method and Certainty Factor Method	Abdel Rahman Al Ali Al Ali, Reta Diana Pradani, Dr. Danial Md Nor, Mrs. Fahmy Rinanda Saputri, Dr. Khairun Nidzam Ramli and Mr. Mohd Nazri Mahmud	he170096@student.uthm.edu.my, retadiana99@gmail.com, danial@uthm.edu.my, fahmy.rinanda@umn.ac.id, khairun@uthm.edu.my, nazriee@usm.my			2024-08-11	IEEE	Y	

eCF Paper Id	Paper / Article Title	Author Name	Author Email	Authorized Signatory Name	Authorized Signatory Email	Date of Signature	Copyright Type	A/V Recordings Permitted	OA Payment Received
1571034242	An Antibody-Mediated Immune Mechanism in Botnet Defense System	Prof. Shingo Yamaguchi	shingo@yamaguchi-u.ac.jp			2024-07-07	IEEE	Y	
1571034399	Visual Deception: Demonstrating Spoofing Attacks on Autonomous Vehicle Cameras	Dr. Muhammad Asif Khan, Dr. Hamid Menouar, Ms. Ayah Nassar and Prof. Mohamed M. Abdallah	asifk@ieee.org, hamidm@qmic.com, ayanassar2777@gmail.com, moabdallah@hbku.edu.qa			2024-07-15	IEEE	Y	
1571034712	Comparative Analysis of RAG, Fine-Tuning, and Prompt Engineering in Chatbot Development	Mr. Harshit Kumar Chaubey, Mr. Gaurav Tripathi, Mr. Rajnish Ranjan and Dr. Srinivasa k. Gopalaiyengar	harshit21100@iiitnr.edu.in, gaurav21101@iiitnr.edu.in, 21f3001109@ds.study.iitm.ac.in, srinivasa@iiitnr.edu.in			2024-07-15	IEEE	Y	
1571034714	Utilizing Attention-based Vision Transformer(AtVit) for Brain Tumour Detection	Mr. Harshit Kumar Chaubey, Mr. Karan Agrawal, Ms. Siri Arelli and Dr. Srinivasa k. Gopalaiyengar	harshit21100@iiitnr.edu.in, karan21100@iiitnr.edu.in, siri21102@iiitnr.edu.in, srinivasa@iiitnr.edu.in			2024-07-10	IEEE	Y	
1571034844	Leveraging Trusted Input Framework to Correct and Forecast the Electricity Load in Smart City Zones Against Adversarial Attacks	Mr. Mohd Hafizuddin Bin Kamilin, Prof. Shingo Yamaguchi and Dr. Mohd Anuaruddin Bin Ahmadon	d010wcu@yamaguchi-u.ac.jp, shingo@yamaguchi-u.ac.jp, anuar@yamaguchi-u.ac.jp			2024-07-04	IEEE	Y	
1571035163	Analyzing Classification Models: Random Forest vs. Neural Networks in Health Prediction	Mr. Nazirul Faris Md Zahidi, Mr. Nurul Huda Ismail and Mrs. Zarina Tukiran	nazirulfaris0323@gmail.com, nhuda@uthm.edu.my, zarin@uthm.edu.my			2024-08-15	IEEE	Y	
1570989258	Development of Drowsiness & Drunk Detection Using YOLOv8	Mr. Wei How Ho, Dr. Abd Kadir Mahamad, Mrs. Sharifah Saon, Ms. Maisara Othman, Dr. Herdawatie Abdul Kadir and Mr. Hakkun Elmunsyah	ce200127@student.uthm.edu.my, kadir@uthm.edu.my, sharifa@uthm.edu.my, maisara@uthm.edu.my, watie@uthm.edu.my, hakkun@um.ac.id			2024-07-23	IEEE	Y	

Improvement of Machine Learning Algorithms with Hyperparameter Tuning on Various Datasets

1st Akmar Efendi
department of Informatics Engineering
Universitas Islam Riau
Pekanbaru, Indonesia
akmarefendi@eng.uir.ac.id

2nd Iskandar Fitri
Department of Information Technology
Universitas Putra Indonesia YPTK
Padang, Indonesia
if@upiyptk.ac.id

3rd Gunadi Widi Nurcahyo
Department of Information Technology
Universitas Putra Indonesia YPTK
Padang, Indonesia
gunadiwidi@yahoo.co.id

Abstract—In the digital era with a data explosion, classification techniques have become a primary aspect of machine learning, especially in Supervised Learning methods. These techniques allow computers to learn from existing data and apply their knowledge to classify new data based on patterns found in the training data. Although algorithms such as Support Vector Machine (SVM) and Naïve Bayes are reliable in many cases, they are not always optimal due to data complexity. This study evaluates the performance of various models and applies optimization techniques to enhance model performance across different datasets. The study uses three different datasets: academic data from the Faculty of Engineering at Universitas Islam Riau (UIR), tweet data from the social media platform X, and diabetes disease data from Kaggle. Each model is tested with a 70:30 data split, employing techniques such as SMOTE, Hyperparameter Optimization with Optuna, and XGBoost to improve model performance. The combination of SMOTE with SVM or GNB shows significant improvement in accuracy, precision, recall, and F1-Score when optimization techniques are applied. For instance, the use of SMOTE, SVM, and Optuna achieves 100% accuracy on academic data, 97% on Twitter data, and 80% on diabetes data. Similarly, the combination of SMOTE, GNB, and XGBoost provides significant improvement. This study concludes that the application of optimization techniques like Optuna and the integration with algorithms such as XGBoost significantly enhance the performance of classification models across various datasets. This opens up opportunities for the development of more advanced and effective classification models in the future and makes a significant contribution to understanding the use of classification algorithms in various practical applications.

Keywords—Machine Learning, SVM, GNB, Optuna, XGBoost

I. INTRODUCTION

In the digital era filled with an explosion of data, classification techniques have become one of the main aspects of Machine Learning, especially in Supervised Learning methods [1] These techniques allow computers to learn from existing data and then apply their knowledge to distinguish and classify new data based on patterns found in the training data [2]. The main goal is to empower programs to distinguish and classify new observations based on patterns hidden in the existing training data [3]. In the classification process, algorithms learn from past experiences and then apply their insights to classify new data into predefined categories [4]

Although classification algorithms such as Support Vector Machine (SVM) and Naïve Bayes have proven their reliability in many cases, they do not always provide optimal performance in all contexts.

Previous research, such as that conducted by [5], [6], highlights the variations in the performance of SVM and Naïve Bayes, depending on the type of data used. For example, when faced with the PeduliLindungi application dataset on the Google Play Store, SVM provided an accuracy of 80.5%, while Naïve Bayes achieved 76.9%. Then, when these algorithms were applied to the Twitter dataset, SVM accuracy increased to 87% and Naïve Bayes reached 83% [7]. Additionally, both have been used to predict cable TV payments, where Naïve Bayes achieved an impressive accuracy of 96%, while SVM only reached 66% [8].

However, fluctuations in this accuracy are often caused by the complexity of the data faced [9]. Therefore, this research aims to explore various models of Naïve Bayes and SVM algorithms, as well as apply techniques such as SMOTE [10], Hyperparameter Optuna [11], and XGBoost [12] to increase model performance. This study utilizes three different datasets: academic data from the Faculty of Engineering, Islamic University of Riau (UIR), tweet data from social media X, and diabetes disease data from Kaggle. Each model is tested using a 70:30 data split.

The main goal of this research is to develop effective classification models to improve accuracy across various types of datasets. By combining careful experiments with the application of state-of-the-art techniques in Machine Learning, this study aims to provide a deeper understanding of the performance of classification algorithms in diverse data contexts. Through this research, it is also expected to pave the way for the development of more sophisticated and effective classification models in the future, and also make a significant contribution to the understanding of the use of classification algorithms in various practical applications.

II. METHOD

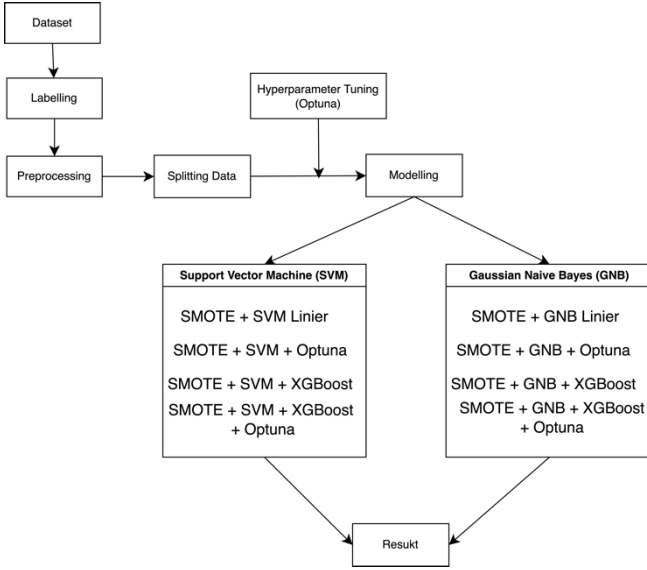In picture 1, the developed model for testing can be seen.

Fig. 1. Model Development

A. Dataset

The dataset in this study consists of 2 datasets. The first dataset is academic data from the Faculty of Engineering, Islamic University of Riau in 2016. Dataset 1 comprises 1282 records of students who have completed their thesis courses. Dataset 2 is sourced from social media X and includes 10,001 records. Dataset 3 is obtained from diabetes disease data on Kaggle, containing 768 records.

B. Labelling

Labeling data in dataset 1 involves 4 categories: DropOut (DO), Graduated On Time, Graduated Late, and Stopped. The labeling process follows these rules:

1. A student is labeled as DropOut (DO) if they exceed the specified time limit of 14 semesters or if they commit a serious violation.
2. A student is labeled as Graduated On Time if they graduate in semester 7, 8, or 9.
3. A student is labeled as Graduated Late if they graduate after 9 semesters.
4. A student is labeled as Stopped if they do not register for a thesis or if they discontinue their studies in a specific semester.

Dataset 2 consists of Twitter data categorized into 3 labels: positive, negative, and neutral sentiments. Dataset 3 is divided into 2 labels: diabetes and non-diabetes. In the labeling process of this study, each label is not balance, so data balancing is performed using the SMOTE method. SMOTE (Synthetic Minority Over-sampling Technique) is an oversampling technique where new minority class instances are generated to match the number of majority class instances [13].

C. Preprocessing

Dataset 1 and 3, after being labeled, the data standardization is taking place using the Standard Scaler.

The Standard Scaler is applied to ensure data has a consistent scale and range. Additionally, the Standard Scaler has another benefit which is more stable handling of outliers compared to some other normalization methods. Outliers can heavily influence other normalization methods, but the Standard Scaler relies on mean and standard deviation, which are less affected by extreme values. Meanwhile, Dataset 2 uses data cleaning, case folding, text normalization, tokenizin, filtering, and stemming [14].

D. Model

Then this research conducted several test with various models, table 1 represent the model that constructed in this research

TABLE I.   EXPERIMENTS CONDUCTED IN THIS RESEARCH

No	Researcher	Trial	Accuracy
1	[15]	Support Vector Machine (SVM) Algorithm	85.40%
2	[16]	SVM + Hyperparameter (Random Search)	63.81%
3	[17]	SVM + XGBoost	79%
4	-	SVM + XGBoost + Hyperparameter	-
5	[18]	Gaussian Naive Bayes (GNB) Algorithm	96%
6	[19]	GNB + Hyperparameter (Genetic Algorithm)	93.2%
7	[12]	GNB + XGBoost	81.55%
8	-	GNB + XGBoost + Hyperparameter	-

Table 1 presents the experimental results from some studies applying various combinations of classification algorithms and optimization techniques on different datasets. From the table, it can be seen that research using the Gaussian Naive Bayes (GNB) algorithm achieved the highest accuracy at 96%. However, there is variation in accuracy results among other studies. For instance, using the Support Vector Machine (SVM) algorithm with Hyperparameter optimization (Random Search) only achieved an accuracy of 63.81%, which is significantly lower compared to GNB. This finding highlights the importance of selecting appropriate classification algorithms and optimization techniques to achieve optimal accuracy in data classification tasks. However, there are 2 models that have not been found from previous research that combine this, namely SVM + XGBoost + Hyperparameter and GNB + XGBoost + Hyperparameter. This research will use Optuna as the hyperparameter used in this research

Optuna is a library that can automate the process of parameter tuning. Not only that automate this process, but Optuna also includes effective search algorithms that make the search process more efficient [11].

E. Performance Evaluation

The performance evaluation in this study involves assessing various classification models such as SVM and GNB, as well as combinations of optimization techniques like Optuna and XGBoost, across three distinct datasets. Metrics including accuracy, precision, recall, and F1-Score are employed to comprehensively evaluate model effectiveness. Accuracy measures the overall correctness of predictions relative to the total number made. Precision

gauges the accuracy of positive predictions, indicating how many correctly predicted positives exist among all positive predictions. Recall assesses the completeness of positive predictions by determining how many actual positive instances were correctly identified by the model. The F1-Score, being the harmonic mean of precision and recall, provides a balanced measure that considers both metrics. This evaluation framework aims to understand each model's capability to handle diverse datasets and identifies optimal combinations of optimization techniques to enhance performance. Additionally, analyzing prediction errors helps uncover scenarios where models struggle and identifies underlying data patterns contributing to these errors. Such insights contribute to refining classification models for future applications across varied data contexts.

III. RESULT AND DISCUSSION

The results of the research are as follows:

TABLE II. SVM DATA AKADEMIK

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + SVM Linier	91%	91%	91%	91%
SMOTE + SVM + Optuna	100%	100%	100%	100%
SMOTE + SVM + XGBoost	95%	96%	95%	95%
SMOTE + SVM + Optuna + XGBoost	100%	100%	100%	100%

Table 2 presents the evaluation results of various classification models applied to academic data using the Support Vector Machine (SVM) algorithm. The finding shows that using the SMOTE technique alongside linear SVM achieves an accuracy of 91%, with corresponding precision, recall, and F1-Score all at the same level. Furthermore, combining SMOTE with SVM and using the optimization algorithm Optuna results in perfect accuracy of 100%, along with precision, recall, and F1-Score also reaching 100%. Moreover, employing SMOTE with SVM and XGBoost yields an accuracy of 95%, with precision at 96%, recall at 95%, and F1-Score at 95%. Finally, utilizing a combination of SMOTE, SVM, Optuna, and XGBoost also achieves perfect accuracy of 100%, with optimal precision, recall, and F1-Score rates. From the table, it can be concluded that employing optimization techniques such as Optuna and combining them with the XGBoost algorithm significantly enhances the performance of classification models on this academic dataset.

TABLE III. SVM DATA TWITTER

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + SVM Linier	94%	94%	94%	94%
SMOTE + SVM + Optuna	96%	96%	96%	96%
SMOTE + SVM + XGBoost	93%	93%	93%	93%
SMOTE + SVM + Optuna + XGBoost	97%	97%	97%	97%

Table 3 presents the evaluation results of various classification models applied to data from the Twitter

platform using the Support Vector Machine (SVM) algorithm. The results show that using the SMOTE technique alongside linear SVM achieves an accuracy of 94%, with precision, recall, and F1-Score all at the same level. Furthermore, combining SMOTE with SVM and employing the Optuna optimization algorithm increases the accuracy to 96%, with optimal precision, recall, and F1-Score. Next, employing SMOTE with SVM and the XGBoost algorithm results in an accuracy of 93%, with consistent precision, recall, and F1-Score rates. Finally, using a combination of SMOTE, SVM, Optuna, and XGBoost achieves the highest accuracy of 97%, with optimal precision, recall, and F1-Score rates. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm significantly enhances the performance of classification models on Twitter data.

TABLE IV. SVM DATA DIABETES

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + SVM Linier	78%	78%	78%	78%
SMOTE + SVM + Optuna	80%	80%	80%	80%
SMOTE + SVM + XGBoost	73%	73%	73%	73%
SMOTE + SVM + Optuna + XGBoost	77%	78%	77%	77%

Table 4 presents the evaluation results of various classification models applied to diabetes-related data using the Support Vector Machine (SVM) algorithm. The results indicate that using the SMOTE technique alongside linear SVM achieves an accuracy of 78%, with precision, recall, and F1-Score all at the same level. Furthermore, combining SMOTE with SVM and employing the Optuna optimization algorithm increases the accuracy to 80%, with optimal precision, recall, and F1-Score. However, using SMOTE with SVM and the XGBoost algorithm results in slightly lower accuracy at 73%, with consistent precision, recall, and F1-Score rates. Finally, using a combination of SMOTE, SVM, Optuna, and XGBoost achieves an accuracy of 77%, with relatively stable precision, recall, and F1-Score rates. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm improves the performance of classification models on diabetes data, although the accuracy levels vary depending on the combination used.

TABLE V. GNB DATA ACADEMIC

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + GNB	90%	90%	90%	90%
SMOTE + GNB + Optuna	91%	91%	91%	91%
SMOTE + GNB + XGBoost	95%	96%	95%	95%
SMOTE + GNB + Optuna + XGBoost	91%	91%	91%	91%

Table 5 shows the evaluation results of various classification models applied to academic data using the

Gaussian Naive Bayes (GNB) algorithm. From the table, it is observed that using the SMOTE technique alongside GNB achieves an accuracy of 90%, with precision, recall, and F1-Score all at the same level. When combined with the Optuna optimization algorithm, the accuracy slightly increases to 91%, maintaining optimal precision, recall, and F1-Score. Furthermore, employing SMOTE with GNB and the XGBoost algorithm yields the highest accuracy of 95%, with precision at 96% and recall and F1-Score at 95%. However, using a combination of SMOTE, GNB, Optuna, and XGBoost maintains accuracy at 91%, with consistent precision, recall, and F1-Score rates. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm significantly enhances the performance of classification models on academic data. The accuracy levels vary depending on the combination used, highlighting the importance of selecting appropriate techniques for specific datasets.

TABLE VI.   GNB DATA TWITTER

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + GNB	68%	74%	69%	67%
SMOTE + GNB + Optuna	50%	77%	50%	44%
SMOTE + GNB + XGBoost	81%	85%	81%	80%
SMOTE + GNB + Optuna + XGBoost	54%	78%	54%	50%

Table 6 depicts the evaluation results of several classification models applied to Twitter data using the Gaussian Naive Bayes (GNB) algorithm. The results show that using the SMOTE technique alongside GNB achieves an accuracy of 68%, with precision at 74%, recall at 69%, and F1-Score at 67%. However, when this combination is supplemented with the Optuna optimization algorithm, the accuracy decreases to 50%, with precision at 77%, recall at 50%, and F1-Score at 44%. Furthermore, employing SMOTE with GNB and the XGBoost algorithm improves accuracy to 81%, with precision at 85%, recall at 81%, and F1-Score at 80%. However, using a combination of SMOTE, GNB, Optuna, and XGBoost results in lower accuracy at 54%, with precision at 78%, recall at 54%, and F1-Score at 50%. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm can enhance the performance of classification models on Twitter data, although there is variation in accuracy levels depending on the combination used.

TABLE VII.   GNB DATA DIABETES

Model	Accuracy	Precision	Recall	F1-Score
SMOTE + GNB	75%	75%	75%	75%
SMOTE + GNB + Optuna	73%	74%	73%	73%
SMOTE + GNB + XGBoost	73%	75%	73%	73%
SMOTE + GNB + Optuna + XGBoost	80%	81%	80%	80%

Table 7 presents the evaluation results of several classification models applied to diabetes-related data using the Gaussian Naive Bayes (GNB) algorithm. From the table, it can be seen that using the SMOTE technique alongside GNB achieves an accuracy of 75%, with precision, recall, and F1-Score all at the same level. However, when this combination is supplemented with the Optuna optimization algorithm, the accuracy slightly decreases to 73%, with precision, recall, and F1-Score remaining stable. Similarly, using SMOTE with GNB and the XGBoost algorithm yields similar results, with accuracy and other evaluation metrics remaining relatively stable around 73%. However, the combination of SMOTE, GNB, Optuna, and XGBoost improves accuracy to 80%, with optimal precision, recall, and F1-Score. From the table, it can be concluded that utilizing optimization techniques such as Optuna and integrating them with the XGBoost algorithm can enhance the performance of classification models on diabetes data, although there is variation in accuracy levels depending on the combination used.

Based on the evaluation results from those tables, it can be concluded that the use of optimization techniques such as Optuna and integration with the XGBoost algorithm consistently improves the performance of the classification model, regardless of the nature and type of data used.

To give more comprehensive analysis, this study include several key aspects. Firstly, performance comparison evaluations of various classification models such as SVM and GNB, as well as combinations of optimization techniques like Optuna and XGBoost, were conducted on three different datasets. Model performance was measured using metrics such as accuracy, precision, recall, and F1-Score to identify which models performed best in the context of each dataset. The findings indicate that the model combining SMOTE, SVM, and Optuna achieved the best results with 100% accuracy on the academic dataset. On the Twitter dataset, the combination of SMOTE, SVM, and Optuna+XGBoost achieved the highest accuracy of 97%, while on the diabetes dataset, the combination of SMOTE, GNB, and Optuna+XGBoost achieved 80% accuracy.

The influence of optimization techniques was also deeply analyzed. Optuna was utilized for hyperparameter optimization, significantly increase the performance of the classification models by helping to find the best parameter sets that maximize accuracy and other evaluation metrics. Additionally, the integration of XGBoost, known for its ability to handle imbalanced data and improve model performance, also showed significant performance improvements in several datasets.

Error analysis was conducted to understand prediction errors in the model, including false positives and false negatives. This analysis helps in comprehending specific conditions where the model fails and identifying patterns or data characteristics that lead to these errors. It is crucial for improving models in the future.

Performance variation based on datasets was also analyzed. The classification models showed the best performance on the academic dataset, achieving the highest accuracy of 100%. Despite the Twitter dataset being more complex and diverse, the optimized models also demonstrated very good performance, indicating that the optimization techniques used are effective in handling

varied social media text data. On the diabetes dataset, model performance showed greater variation, with the best model achieving 80% accuracy. This suggests that medical data may require specific approaches and more optimization to achieve optimal performance.

With this much deeper approach, this research not only provides performance evaluation results but also a comprehensive analysis of how and why certain models outperform others. It offers further insights into research directions in the future as well.

IV. CONCLUSION

Overall, the evaluation results from the tables shows the relative performance of various classification models in processing different datasets. Meanwhile, the conclusions provide an overview of the use of optimization techniques and recommendations for future research.

The evaluation results provide concrete information about the accuracy, precision, recall, and F1-Score from each classification model on the datasets used. This helps assess the relative performance of each model and algorithm combination with clear evaluation metrics.

On the other hand, the conclusions provide a summary of these findings and explore further implications. This includes recommendations for future research, such as the development of more advanced optimization techniques or exploring models on more diverse datasets. Conclusions may also include a brief reflection on the key findings and their relevance in wider research contexts.

REFERENCES

[1] R. Pugliese, S. Regondi, and R. Marini, "Machine learning-based approach: Global trends, research directions, and regulatory standpoints," *Data Science and Management*, vol. 4. KeAi Communications Co., pp. 19–29, Dec. 01, 2021. doi: 10.1016/j.dsm.2021.12.002.

[2] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, pp. 685–695, 2021, doi: 10.1007/s12525-021-00475-2.

[3] I. Nurcahyani and J. W. Lee, "Role of machine learning in resource allocation strategy over vehicular networks: A survey," *Sensors*, vol. 21, no. 19. MDPI, Oct. 01, 2021. doi: 10.3390/s21196542.

[4] J. S. Harrison, M. A. Josefy, M. Kalm, R. Krause, and C. S. Joseph Harrison, "Using supervised machine learning to scale human-coded data: A method and dataset in the board leadership context," *Strat Mgmt J*, vol. 44, pp. 1780–1802, 2023, doi: 10.5281/zenodo.7304697.

[5] A. Salma and W. Silfianti, "Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors," *International Research Journal of Advanced Engineering and Science*, vol. 6, no. 4, pp. 158–162, 2021.

[6] P. T. Putra, A. Anggrawan, and H. Hairani, "Comparison of Machine Learning Methods for Classifying User Satisfaction Opinions of the PeduliLindungi Application," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 3, pp. 431–442, Jun. 2023, doi: 10.30812/matrik.v22i3.2860.

[7] L. D. Utami and S. Masripah, "Comparation Of Classification Algorithm On Sentiment Analysis Of Online Learning Reviews And Distance Education," *Jurnal Techno Nusa Mandiri*, vol. 18, no. 2, pp. 101–110, 2021, doi: 10.33480/techno.v18i2.2715.

[8] M. E. Lasulika, "Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Accuracy Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel," *ILKOM Jurnal Ilmiah*, vol. 11, no. 1, pp. 11–16, 2019, doi: 10.33096/ilkom.v11i1.408.11-16.

[9] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 3, pp. 1–21, May 2021, doi: 10.1007/s42979-021-00592-x.

[10] M. K. Anam *et al.*, "Sentiment Analysis for Online Learning using The Lexicon-Based Method and The Support Vector Machine Algorithm," *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, pp. 290–302, 2023, doi: 10.33096/ilkom.v15i2.1590.290-302.

[11] W. A. G. Kodri and S. Hadianti, "Optimization of The Machine Learning Approach using Optuna in Heart Disease Prediction," *Journal Medical Informatics Technology*, vol. 1, no. 3, pp. 59–64, Sep. 2023, doi: 10.37034/medinftech.v1i3.15.

[12] T. R. Mahesh, V. Vinoth Kumar, V. Muthukumaran, H. K. Shashikala, B. Swapna, and S. Guluwadi, "Performance Analysis of XGBoost Ensemble Methods for Survivability with the Classification of Breast Cancer," *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/4649510.

[13] N. Matondang and N. Surantha, "Effects of oversampling SMOTE in the classification of hypertensive dataset," *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 4, pp. 432–437, 2020, doi: 10.25046/AJ050451.

[14] A. P. Joshi and B. V. Patel, "Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process," *Oriental journal of computer science and technology*, vol. 13, no. 0203, pp. 78–81, Jan. 2021, doi: 10.13005/ojcst13.0203.03.

[15] R. Haque, N. Islam, M. Tasneem, and A. K. Das, "Multi-class sentiment classification on Bengali social media comments using machine learning," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 21–35, Jun. 2023, doi: 10.1016/j.ijcce.2023.01.001.

[16] H. S. Kang, Y. S. Choi, J. S. Yu, S. W. Jin, J. M. Lee, and Y. J. Kim, "Hyperparameter Tuning of OC-SVM for Industrial Gas Turbine Anomaly Detection," *Energies (Basel)*, vol. 15, no. 22, pp. 1–12, Nov. 2022, doi: 10.3390/en15228757.

[17] R. Rahmaddeni, M. K. Anam, Y. Irawan, S. Susanti, and M. Jamaris, "Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, pp. 32–38, Apr. 2022, doi: 10.33096/ilkom.v14i1.1090.32-38.

[18] M. Vedaraj, C. S. Anita, A. Muralidhar, V. Lavanya, K. Balasaranya, and P. Jagadeesan, "Early Prediction of Lung Cancer Using Gaussian Naive Bayes Classification Algorithm," *International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 11, no. 6s, pp. 838–848, 2023, [Online]. Available: www.ijisae.org

[19] I. Firman Ashari, M. Cahyo Untoro, E. Maharani Sutowo, D. Salsabila, and D. Athiyah Zhabiyan, "Hyperparameter Tuning Feature Selection with Genetic Algorithm and Gaussian Naïve Bayes for Diabetes Disease Prediction," *Jurnal Telematika*, vol. 17, no. 1, pp. 17–23, 2022, doi: 10.61769/jurtel.v17i1.488.