2024 International Seminar on Intelligent Technology and Its Applications (ISITIA) | 979-8-3503-7857-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISITIA63062.2024.10668210

25th ISITIA 2024

25th International Seminar on Intelligent Technology and Its Applications (ISITIA) 2024

PROCEEDINGS

Collaborative Innovation:
A Bridging from Academia to Industry towards Sustainable Strategic Partnership

Lombok, Indonesia
10 - 12 July 2024

IEEE Conference Number: 63062
IEEE Catalog Number: CFP24TIA-ART
ISBN: 979-8-3503-7857-3
ISSN: 2769-5492

ITS Institut Teknologi Sepuluh Nopember
IEEE INDONESIA SECTION
MTT-S AP-S IEEE MTT/AP Joint Chapter Indonesia
ICS
Universitas Mataram
FORTEI Forum Pendidikan Tinggi Teknik Elektro Indonesia
PLN
Unitronic Jaya

PROGRAM BOOK

2024 International Seminar on Intelligent Technology and Its Application (ISITIA 2024)

Collaborative Innovation: A Bridging from Academia to Industry towards Sustainable Strategic Partnership

10 – 12 July 2024

HYBRID CONFERENCE

MATARAM – INDONESIA

Organized by:

Department of Electrical Engineering

Institut Teknologi Sepuluh Nopember, Indonesia

Technical Co-Sponsor

Sponsors

ISITIA 2024 Organizing Committee

GENERAL CHAIR

Dr. Ir. Puji Handayani, M.T.

VICE-CHAIR

Dr. Ir. Hendra Kusuma, M.Eng.Sc.

TECHNICAL PROGRAM

Dr.techn. Prasetiyono Hari Mukti, S.T., M.T. (Chair)

Yurid Eka Nugraha, S.T., M.Eng., Ph.D. (Co-chair)

Dr. Made Ginarsa, S.T., M.T. (Co-chair)

PUBLICATION

Dr. Devy Kuswidiastuti, S.T., M.Sc. (Chair)

Dr. Suwito, S.T., M.T. (Co-chair)

Dr. Rosmaliati, S.T., M.T. (Co-chair)

SECRETARIAT

Dr. I Gusti Ngurah Satriyadi Hernanda, ST, MT (Chair)

Ahmad Sirojuddin, S.T., M.Sc., Ph.D. (Co-chair)

TREASURER

Dr. Dimas Fajar Uman Putra, S.T., M.T. (Chair)

Dr. Daniar Fahmi, S.T., M.T. (Co-chair)

CONTACT DETAILS:

Department of Electrical Engineering
Gedung B, C dan AJ Kampus ITS Sukolilo Surabaya, Jawa Timur, Indonesia, 60111
Phone: (+62 31) 5947302; (+62 31) 593 1237
Email: isitia@its.ac.id
Website: http://isitia.its.ac.id/

ISITIA 2024 Technical Program Committee and Reviewer

Abdul Wachid Syamroni, Ph.D.	National Research and Innovation Agency (Indonesia)
Dr. Ir. Achmad Affandi, DEA	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Ir. Achmad Mauludiyanto, M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Ir. Adhi Dharma Wibawa, S.T., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Agung Nugroho Pramudhita, S.T., M.T.	Politeknik Negeri Malang (Indonesia)
Dr. Ir. Agung Wahyu Setiawan, M.T.	Institut Teknologi Bandung (Indonesia)
Dr. Agus Indra Gunawan, S.T., M.Sc	Politeknik Elektronika Negeri Surabaya (PENS) (Indonesia)
Ahmad Sirojuddin, S.T., M.Sc., Ph.D.	Institut Teknologi Sepuluh Nopember (Indonesia)
Ajay Lotan Thakur, BE.E., M.Eng,	Intel Corp Canada (Canada)
Aji Akbar Firdaus, S.T.,M.T.	Universitas Airlangga (Indonesia)
Aji Gautama Putrada, S.T., M.T.	Telkom University (Indonesia)
Prof. Dr. Ir. Andi Adriansyah, M.Eng	Universitas Mercu Buana (Indonesia)
Andy Rachman, S.T.,M.Kom	Institut Teknologi Adhi Tama Surabaya (Indonesia)
Angga Wahyu Aditya, S.ST., M.T.	Polytechnic State of Balikpapan (Indonesia)
Dr. Eng. Ardyono Priyadi, ST., M.Eng.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Arief Kurniawan, S.T., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Ir. Arwin Datumaya Wahyudi Sumari, S.T., M.T.,	State Polytechnic of Malang (Indonesia)
Ary Mazharuddin Shiddiqi, S.Kom., M.Comp.Sc., Ph.D	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Asokan Ramasamy	Kongunadu College of Engineering and Technology (India)
Astria Nur Irfansyah, S.T., M.Eng., Ph.D.	Institut Teknologi Sepuluh Nopember (Indonesia)
Atik Novianti, S.ST., M.T.	State Polytechnic of Malang (Indonesia)
Budi Artono, S.T., M.T.	Politeknik Negeri Madiun (Indonesia)
Dani Adhipta, S.Si., M.T.	Universitas Gadjah Mada (Indonesia)
Daniar Fahmi, S.T., M.T., Ph.D.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dedet Candra Riawan, S.T., M.Eng., Ph.D.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Devy Kuswidiastuti, S.T., M.Sc.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dewi H Budiarti, S.T., M.T., Ph.D.	National Research and Innovation Agency (Indonesia)
Dr. Dhany Arifianto, S.T., M.Eng.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Dia Ali	Ninevah University (Iraq)
Dian Puspita Hapsari, S.Kom.,M.Kom.	Institut Teknologi Adhi Tama Surabaya (Indonesia)
Dr. Diana Purwitasari, S.Kom., M.Sc.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dimas Anton Asfani, S.T, M.T.,Ph.D.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Dimas Fajar Uman Putra, S.T., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dion Hayu Fandiantoro, S.T., M.Eng.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dwi Ely Kurniawan, S.Pd., M.Kom.	Politeknik Negeri Batam (Indonesia)
Prof. Eduard Babulak D.Sc., Ph.D.	Liberty University (USA)

Prof. Dr. Elieser Tarigan	University of Surabaya (Indonesia)
Dr. Ir. Endang Widjiati, M.Eng.Sc.	National Research and Innovation Agency (BRIN) (Indonesia)
Fanhui Zeng	Google (USA)
Fardani Annisa Damastuti, S.ST., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Feby Agung Pamuji, ST., MT., PhD.	Institut Teknologi Sepuluh Nopember (Indonesia)
Ferdib Al Islam, B.Sc., M.Sc.	Northern University of Business and Technology Khulna (Bangladesh)
Gusti Eka Yuliastuti, S.Kom., M.Kom.	Institut Teknologi Adhi Tama Surabaya (Indonesia)
Prof. Ir. Hanny Hosiana Tumbelaka, M.Sc., Ph.D.	Petra Christian University (Indonesia)
Harits Ar Rosyid, S.T., M.T., Ph.D.	State University of Malang (Indonesia)
Assoc. Prof. Ir. Dr. Hazlee Azil Illias	University of Malaya (Malaysia)
Dr. Ir. Hendra Kusuma, M.Eng.Sc.	Institut Teknologi Sepuluh Nopember (Indonesia)
Henning Titi Ciptaningtyas, S.Kom, M.Kom	Institut Teknologi Sepuluh Nopember (Indonesia)
Prof. Heri Suryoatmojo, ST., MT., Ph.D	Institut Teknologi Sepuluh Nopember (Indonesia)
Hero Wintolo, S.T., M.Kom.	Institut Teknologi Dirgantara Adisutjipto (Indonesia)
Dr. Hindarto, S.Kom., MT	University Muhammadiyah Of Sidoarjo (Indonesia)
Hisyam Fahmi, S.Kom., M.Kom	Universitas Islam Negeri Maulana Malik Ibrahim Malang (Indonesia)
Dr. Houari Aoued	University of Sciences and Technolgies of Oran Mohammed Boudhiaf (Algeria)
Dr. I Gede Puja Astawa, ST., MT	Politeknik Elektronika Negeri Surabaya (Indonesia)
Dr. I Gusti Ngurah Satriyadi, ST., MT.	Institut Teknologi Sepuluh Nopember (Indonesia)
Prof. Dr.Eng. I Made Yulistya Negara, ST., M.Sc.	Institut Teknologi Sepuluh Nopember (Indonesia)
Prof. Ir. Ida Ayu Dwi Giriantari, M.EngSc, Ph.D	Udayana University, Bali (Indonesia)
Ima Kurniastuti, S.T., M. T.	Universitas Nahdlatul Ulama Surabaya (Indonesia)
Prof. Dr. Ir. Imam Robandi, M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Eng. Imam Wahyudi Farid, S.T., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Indar Chaerah Gunadin, S.T, M.T.	Hasanuddin University (Indonesia)
Dr. Indra Kharisma Raharjana, S.Kom., M.T.	Universitas Airlangga (Indonesia)
Dr. Ir. Ingrid Nurtanio, M.T.	Hasanuddin University (Indonesia)
Irwan Alnarus Kautsar, S.Kom. M.Kom. Ph.D	Universitas Muhammadiyah Sidoarjo (Indonesia)
Jega Anish Dev, B.E., M.Sc.	Microsoft (USA)
Prof. John Joshua Montañez	Bicol State College of Applied Sciences and Technology (Philippines)
Prof. Koichi Maru, Ph.D.	Kagawa University (Japan)
Prof.Dr.Layla H.Abood	University of Technology (Iraq)
Dr. Ir. Lie Jasa, MT	Udayana University (Indonesia)
Dr. Mahirah Binti Jahari	Universiti Putra Malaysia (Malaysia)
Prof. Makoto Ohki	Graduate School of Tottori University (Japan)
Dr. Mardhani Riasetiawan, SE Ak, MT	Universitas Gadjah Mada (Indonesia)

Marzuki Pilliang, M.Kom	Bina Insani University (Indonesia)
Prof.Dr.Ir. Mauridhi Hery Purnomo, M.Eng.	Institut Teknologi Sepuluh Nopember (Indonesia)
Md Shakhawat Hossen, B.Sc., M.Eng.	Universiti Kebangsaan Malaysia (Malaysia)
Dr. Mochammad Sahal, S.T., M.Sc.	Institut Teknologi Sepuluh Nopember (Indonesia)
Mochammad Wahyudi, S.T., M.T.	Universitas Gadjah Mada (Indonesia)
Moh Hasbi Assidiqi, S.T, M.T.	Politeknik Elektronika Negeri Surabaya (Indonesia)
Mohammad Hidayat, Ph.D.	State Polytechnic of Malang (Indonesia)
Dr. Mohammad Teduh Uliniansyah	National Research and Innovation Agency (Indonesia)
Dr. Mohd Ashraf Ahmad, B.Eng, M.Eng	Universiti Malaysia Pahang (Malaysia)
Dr. Eng. Muhammad Abdillah, S.T., M.T.	Universitas Pertamina (Indonesia)
Muhammad Afif Hendrawan, S.Kom., M.T.	Politeknik Negeri Malang (Indonesia)
Muhammad Arief Nugroho, S.T., M.T.	Telkom University (Indonesia)
Muhammad Attamimi, B.Eng., M.Eng., Ph.D.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Eng. M. Fikry, S.Kom., M.Kom	Universitas Malikussaleh (Indonesia)
Muhammad Hilman Fatoni, S.T., M.T.,	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Muhammad Irwan Padli Nasution, S.T., M.M., M.Kom	Universitas Islam Negeri Sumatera Utara (Indonesia)
Dr. Muhammad Rivai, S.T., MT.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Muladi , ST, MT	State University of Malang (Indonesia)
Nada Fitrieyatul Hikmah, S.T., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Prof. Dr. Ir. Nasaruddin, S.T.,M.Eng.	Syiah Kuala University (Indonesia)
TS. DR. Nazrulazhar Bahaman	Univeristi Teknikal Malaysia Melaka (Malaysia)
Dr. Ir. Ni Ketut Aryani, M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Nor Asrina Ramlee	University of Technology Sarawak (Malaysia)
Dr. Norma Hermawan, S.T., M.Sc	Institut Teknologi Sepuluh Nopember (Indonesia)
Novendra Setyawan, ST., MT	University of Muhammadiyah Malang (Indonesia)
Dr. Okto Dinaryanto, Dr., S.T., M.M, M.Eng	Institut Teknologi Dirgantara Adisutjipto (Indonesia)
Prof. Ir. Ontoseno Penangsang, MSc., PhD.	Institut Teknologi Sepuluh Nopember (Indonesia)
Prof. Ouadoudi Zytoune, Ph.D.	Ibn Tofail University (Morocco)
Pallav Dutta, B. Tech., M. Tech	Aliah University (India)
Prof. Paolo Crippa, Ph.D.	Marche Polytechnic University (Italy)
Pramana Yoga Saputra, S.Kom., M.MT.	State Polytechnic of Malang (Indonesia)
Dr.techn. Prasetiyono Hari Mukti, ST, MSc	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Ir. Puji Handayani, M.T.	Institut Teknologi sepuluh Nopember (Indonesia)
Dr. Eng. Puput Dani Prasetyo Adi, S.Kom., M.T	BRIN (Indonesia)
Dr.techn. Ir. Raden Venantius Hari Ginardi, M.Sc	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Radi, STP., M.Eng.	Universitas Gadjah Mada (Indonesia)
Rahmat Yasirandi, S.T., M.T., Ph.D.	King Mongkuts Institute of Technology Ladkrabang (Thailand)
Dr. Ratheesh Kumar Meleppat	University of California Davis (USA)
Ratih Kartika Dewi, S.T., M.Kom.	Brawijaya University (Indonesia)

Rezki El Arif, S.T., M.T., Ph.D	Institut Teknologi Sepuluh Nopember (Indonesia)
Ridha Muldina Negara, S.T., M.T.	Telkom University (Indonesia)
Rieke Adriati Wijayanti, S.T., M.T.	Politeknik Negeri Malang (Indonesia)
Dr. Rizka Wakhidatus Sholikah, S. Kom	Institut Teknologi Sepuluh Nopember (Indonesia)
Ronny Mardiyanto, ST., MT., Ph.D	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Eng. Rony Seto Wibowo, S.T., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Rony Teguh, Ph.D.	University of Palangkaraya (Indonesia)
Prof. Rosaura Palma-Orozco, Ph.D.	Instituto Politécnico Nacional (Mexico)
Dr. Rudy Dikairono, S.T., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Dr. Salem Sati, M.Sc.	Misurata University (Libya)
Septian Enggar Sukmana, S.Pd., M.T.	Politeknik Negeri Malang (Indonesia)
Prof. Dr. Shanmugasundaram Hariharan	Vardhaman College of Engineering (India)
Prof. Dr. Shashikant Patil	Atlas SkillTech University Kurla West (India)
Sheikh Tanzim Meraj, B.Sc., M.Eng.	Universiti Teknologi PETRONAS (Malaysia)
Dr. Smita Nirkhi	Nagpur University (India)
Dr. Ir. Sri Wahjuni, M.T.	IPB University (Indonesia)
Dr. Suwito, S.T., M.T.	Institut Teknologi Sepuluh Nopember (Indonesia)
Taufiq Rizaldi, S.ST., MT	Politeknik Negeri Jember (Indonesia)
Teguh Prakoso, S.T., M.T., Ph.D	Diponegoro University (Indonesia)
Dr. Ir. Totok Mujiono, MIKom	Institut Teknologi Sepuluh Nopember (Indonesia)

Table of Contents

ISITIA 2024 General Program Schedule

General Schedule, Day 1
Wednesday, 10 July 2024

Time*	Agenda	Venue
08:00 – 09:00	Registration	Selaparan Ballroom (Hybrid)
09:00 – 09:30	Opening Ceremony: • **Welcome speech and official opening**	
09:30 – 09:40	Photo Session	
09:40 – 10:00	Coffee Break	
10:00 – 11:00	Keynote Speaker 1: Prof. Hazlie Mokhlis *(Universiti Malaya, Malaysia)* Title: Building Resilient Power Distribution System: Research	
11:00 – 12:00	Keynote Speaker 2: Dr. Wirawan *(Institut Teknologi Sepuluh Nopember, Indonesia)* Title: Building National Competence in Quantum Information Science and Technology: Yes, We Can	
12:00 – 13:00	Lunch Break	
13:30 – 15:30	Technical Session 1: Oral Presentation	Technical Session Room** (Hybrid)
15:30 – 16:00	Coffee Break	Selaparan Ballroom (Hybrid)

*** All time mentioned in the schedule are in Middle Indonesian Time (WITA or UTC+8)**

**** Technical Session Rooms**:
Room A : Gili Air
Room B : Gili Meno
Room C : Gili Trawangan
Room D : Mandalika
Room E : Sangkareang
Room F : Selaparan

General Schedule, Day 2
Thursday, 11 July 2024

Time*	Agenda	Venue
07:30 – 08:00	Registration	Selaparan Ballroom (Hybrid)
08:00 – 10:00	Technical Session 2: Oral Presentation	Technical Session Room** (Hybrid)
10:00 – 10:15	Coffee Break	HYBRID
10:15 – 11:15	Keynote Speaker 3: Prof. Wan-Jen Huang *(National Sun Yat-sen University, Taiwan)* Title: Applications of AI in Communication Systems and Industrial Collaborations	Selaparan Ballroom (Hybrid)
11:15 – 12:15	Keynote Speaker 4: Dr. rer. nat. Teti Zubaidah, ST.MT. *(Universitas Mataram, Indonesia)* Title: Mitigation of Catastrophic Technology Failures due to Geomagnetic Storms	
12:15 – 13:00	Lunch Break	
13:00 – 14:00	Workshop on Community Development Chanakya Kumar, Ph.D *(Siddhant Group of Institutions, Pune , India)* Title: Funding opportunities through Humanitarian Activities under the SDG of United Nations	
14:00 – 16:00	Technical Session 3: Oral Presentation	Technical Session Room** (Hybrid)

*** All time mentioned in the schedule are in Middle Indonesian Time (WITA or UTC+8)**

**** Technical Session Rooms**:
Room A : Gili Air
Room B : Gili Meno
Room C : Gili Trawangan
Room D : Mandalika
Room E : Sangkareang
Room F : Selaparan

General Schedule, Day 3
Friday, 12 July 2024

Time	Agenda	Venue
08:00 – 16:00	One-day Tour*	Mataram, Mandalika, Tanjung Ann Beach, Sade

“ The detail will be announced later.

ISITIA 2024 Technical Session Schedule

Technical Session 1 (10 July 2024, 13:30 – 15:30)

Time*	Rooms & Track					
	A	**B**	**C**	**D**	**E**	**F**
	CE	PES	TSP	CE	CE	PES
13:30 – 13:45	A101	B101	C101	D101	E101	F101
13:45 – 14:00	A102	B102	C102	D102	E102	F102
14:00 – 14:15	A103	B103	C103	D103	E103	F103
14:15 – 14:30	A104	B104	C104	D104	E104	F104
14:30 – 14:45	A105	B105	C105	D105	E105	F105
14:45 – 15:00	A106	B106	C106	D106	E106	F106
15:00 – 15:15	A107	B107	C107	D107	E107	F107
15:15 – 15:30	A108	B108	C108	D108	E108	F108
15:30 – 15:45		B109				

* Central Indonesian Time (UTC+8)

Code:

BME : Biomedical Engineering
CE : Computer Engineering & Information Technology
CSE : Control Systems Engineering
EL : Electronics
PES : Power and Energy Systems
TSP : Telecommunication & Signal Processing

Session 1 – Track: Computer Engineering and Information Technology, Room E

Paper code	Paper ID	Title	Authors
E101	1571027594	Analysis of Consistency and Structure of Scholarly Papers Using Natural Language Processing	Sitti Mawaddah Umar, Ingrid Nurtanio, Zahir Zainuddin
E102	1571016862	Topic Modeling for Online Health Consultation on Low-Risk Diseases	Safitri Juanita, Diana Purwitasari, I Ketut Purnama, Abid Abdillah, Mauridhi Hery Purnomo
E103	1571022007	Burglar Detection System Using Wi-Fi Based Secure Indoor Positioning System	Hatma Suryotrisongko, Naufal Ramadhan, R. V. Hari Ginardi
E104	1571027642	Migration of Monolithic to Microservices With an Extraction Design Pattern in Single Sign on (SSO) Module Using Graph Neural Network (GNN)	Laeila Mardhatillah, Siti Rochimah
E105	1571016993	Automating School Timetabling: An Intelligent System Application Using Simulated Annealing	Tio Arya Dewa Prakasa, Ahmad Muklason, I Gusti Agung Premananda
E106	1571022229	Indonesian Sign Language (SIBI) Recognition and Extraction Using Convolutional Neural Networks - Symmetric Deletion Spelling Correction	Maulina Safitri, Eko Mulyanto Yuniarno, Reza Fuad Rachmadi
E107	1571027326	Ship Evacuation Simulation Based on Reinforcement Learning: A Case Study on NPCs Behavior	Ilham Jalu Prakosa, Supeno Mardi Susiki Nugroho, Diah Wulandari
E108	1571019325	Novel Ranking Indexes for Many-Objective Optimization Evolutionary Algorithms	Leo Hagihara, Makoto Ohki

E102 (Computer Engineering & Information Technology)

Paper ID : 1571016862

Paper Title : **Topic Modeling for Online Health Consultation on Low-Risk Diseases**

Authors : Safitri Juanita (Institut Teknologi Sepuluh Nopember (ITS) & Universitas Budi Luhur, Indonesia); Diana Purwitasari and I Ketut Purnama (Institut Teknologi Sepuluh Nopember, Indonesia); Abid Abdillah (Politeknik Elektronika Negeri Surabaya, Indonesia); Mau

Abstract : Online Health Consultation (OHC) can improve healthcare quality, efficiency, and equity by offering medical services online and overcoming geographical barriers for those far from health institutions. However, this huge dataset of Q&As from OHCs has yet to reveal Indonesians' online health information needs for low-risk diseases. Therefore, the distribution of topics should be analyzed and identified to find trends in low-risk diseases and group the topics into the most important clusters. The proposed approach for this research is topic modeling using LDA. This study aims to identify topics derived from textual analysis of physicians' answers on low-risk diseases in OHC and investigate the trends of low-risk disease topics based on topic modeling results in order to provide personalized information service recommendations and improve OHC service quality, especially for information seekers or users of OHC services on low-risk diseases. This research contributes to identifying trends in low-risk diseases based on OHC history from Indonesia and describing topic clusters from topic modeling based on physician responses. The results show that based on topic modeling with LDA, physicians identified 33 topics and two trends in online health consultations for low-risk diseases (12 diseases) and non-diseases, such as healthy lifestyle, pregnancy, and medical procedures.

C304 (Computer Engineering & Information Technology)

Paper ID : 1571016866

Paper Title : **Relevance Detection Using Text Entailment for Health-Related Question-Answer Texts With Imbalanced Data**

Authors : Marina Irdayanti, Diana Purwitasari and Daniel Oranova Siahaan (Institut Teknologi Sepuluh Nopember, Indonesia)

Abstract : The health consultation website has questions repeated with similar topics. Thus hence high volume the questions were referred to by the users and given answers by the doctor. Moreover, the answers obtained by doctors can give different answers to similar questions. Otherwise, users find it difficult to judge whether the given answers including relevant or not despite the questions referred to being similar. The research goal is to use the finetuning IndoBERT-QA model for entailment detection and then yield the prediction entailment, contradiction, and neutral to judge the answer based on referred to the question. Strength finetuning the IndoBERT-QA model can manage the word out-of-vocabulary in the Indonesian language and can learn the context from word sequence or the

Topic Modeling for Online Health Consultation on Low-Risk Diseases

Safitri Juanita[1], Diana Purwitasari[2], I Ketut Eddy Purnama[3], Abid Famasya Abdillah[4], Mauridhi Hery Purnomo[5]
[1,5]*Departement of Electrical Engineering,* Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2,4]*Departement of Informatics*, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[3,4]*Departement of Computer Engineerings,* Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2,3,5]*University Center of Excellence on Artificial Intelligent for Healthcare and Society (UCE AIHeS)*, Surabaya, Indonesia
[1]*Departement of Information System,* Universitas Budi Luhur, Jakarta, Indonesia
7022201018@student.its.ac.id, diana@if.its.ac.id, ketut@te.its.ac.id, abid.famasya@gmail.com, hery@ee.its.ac.id

Abstract— **Online Health Consultation (OHC) can improve healthcare quality, efficiency, and equity by offering medical services online and overcoming geographical barriers for those far from health institutions. However, this huge dataset of Q&As from OHCs has yet to reveal Indonesians' online health information needs for low-risk diseases. Therefore, the distribution of topics should be analyzed and identified to find trends in low-risk diseases and group the topics into the most important clusters. The proposed approach for this research is topic modeling using LDA. This study aims to identify topics derived from textual analysis of physicians' answers on low-risk diseases in OHC and investigate the trends of low-risk disease topics based on topic modeling results in order to provide personalized information service recommendations and improve OHC service quality, especially for information seekers or users of OHC services on low-risk diseases. This research contributes to identifying trends in low-risk diseases based on OHC history from Indonesia and describing topic clusters from topic modeling based on physician responses. The results show that based on topic modeling with LDA, physicians identified 33 topics and two trends in online health consultations for low-risk diseases (12 diseases) and non-diseases, such as healthy lifestyle, pregnancy, and medical procedures.**

Keywords—Topic Modeling, LDA, Low-risk disease, Online Health Consultation

I. INTRODUCTION

Understanding the causes of death and the risk factors associated with early death is critical. Noncommunicable diseases (NCDs) not only dominate mortality worldwide, but also cause the majority of deaths in high-income countries, according to ourworldindata.org, which used data from 1990 to 2019 [1], [2]. In low-middle-income countries, deaths from causes such as infectious diseases, malnutrition, and neonatal and maternal mortality are common and sometimes dominant [2], [3].

Online Health Consultation (OHC) are rapidly growing and contributing to providing OHC for the public, including the need for information on non-communicable diseases (NCDs) or other low-risk diseases. OHC can improve the quality, efficiency, and equity of health services by providing health services to anyone with medical needs online at any time and overcoming geographical barriers to provide services to people far from health institutions. Investigating public adoption of online healthcare can benefit industry and policy in the healthcare sector [4].

A collection of questions and answers regarding low-risk diseases is available on the OHC service. Low-risk diseases are medical diseases that often do not have a high risk of death or cause serious health problems. In this study, low-risk diseases are not high-risk diseases according to WHO standards in 2020 [5][6]. However, this massive data set has not produced the necessary information to understand Indonesians' online health information needs for low-risk diseases. Therefore, it is necessary to analyze the Q&A data and then identify the distribution of topics to find the trend of low-risk diseases and group the topics into the most significant clusters. The proposed approach for this study is topic modeling, namely LDA.

Topic modeling is an automated process that helps analyze large datasets without the need for manual coding or annotation, as well as an appropriate tool for performing other topic extraction based on short texts from the internet [7]. One of the well-known topic modeling algorithms is the Latent Dirichlet Allocation (LDA) model, which efficiently identifies potential topic information in a document or large corpus [7], delivers more meaningful extracted topics, and obtains good results that describe the entire corpus [8].

Several studies that use a topic modeling approach to understand the information contained in online health services have been conducted previously, including topic modeling in understanding the information needs of Chinese online health communities regarding COVID-19, which aims to discover the information needs topic distribution, trend changes, and impact factors of COVID-19 in Chinese OHCs [7].

The study explores media's role in health communication during the COVID-19 crisis in China, using topic modeling to analyze Chinese news and articles from the WiseSearch database [9]. The study examines physicians' responses to high-risk diseases in online health consultations, with a focus on whether they contain words implying preventative measures against the disease [10]. Using a dataset from the most significant healthcare websites in China, topic modeling helps understand medical terms, symptoms, and anamneses in textual data, enabling the development of recommendation systems to connect patients with the right physicians [11].

Another study uses the LDA model to analyze Chinese medicine user reviews on online medical platforms, identifying various themes and topics related to the subject matter [12] . Another study also uses a similar model to identify the key topics discussed in patient reviews. It investigates the impact of these topics on patient selection and the moderating effect of a physician's specialty [13].

Based on previous research, this study aims to highlight the essential trends in low-risk diseases in Indonesia based on OHC. As a result, the following contributed to this paper:

a. Provide an overview of trends in low-risk disease subjects from online health consultations using a topic modelling method.
b. Identify and explain the clusters formed by the topic modeling technique on a collection of text containing physicians' responses from online health consultations.

The findings of this study have great practical value for online health services to provide personalized information services and improve service quality, especially for information seekers or users who conduct OHC on low-risk diseases in Indonesia.

II. Topic Modeling Technique

This section outlines the research approach utilized in this study, which includes topic modeling procedures in online health consultation for low-risk disorders utilizing a dataset from a website that provides Online Health Consultation (OHC) services, as illustrated in Fig 1. The experimental phase of this research uses 2 python libraries named Gensim and mallet for topic modeling. Described below are each of the stages in Figure 1.

A. Dataset preparation

This study uses a dataset containing a text corpus of answers from general Physicians for low-risk diseases in Indonesia for seven years, from December 4, 2014, to March 4, 2021. The dataset source from four sites providing free services for online health consultations in Indonesia: alodokter.com, sehatq.com, dokter.id, klikdokter.com. Low-risk diseases are diseases other than the list of high-risk diseases issued by WHO in 2020 (WHO, 2020). Fig 2. shows the distribution of the dataset by year.

The text dataset of physicians' responses on low-risk diseases consists of several topics: female reproductive system, autoimmune hematology, nervous disorders, dyspepsia, digestive disorders, dermatology, metabolism, eye diseases, psychological diseases, lung diseases, gastroenteritis, and many more. The total dataset of this study contained 18,307 cases. Fig. 3 illustrates an example dataset: the text of general physicians' responses to questions on low-risk diseases found on the free Online Health Consultation (OHC) service.

B. Data Preprocessing

In the data preprocessing phase, which is the process of cleaning the dataset with the following steps [14], [15]; The

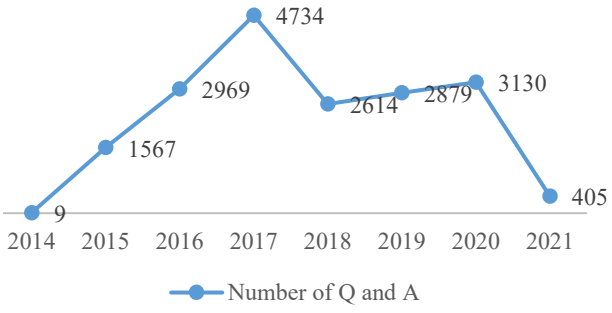

Fig 2. The distribution of the dataset by year contains a collection of texts for general physicians' responses (Q and A) sourced from free Online Health Consultation services (OHCs) in Indonesia

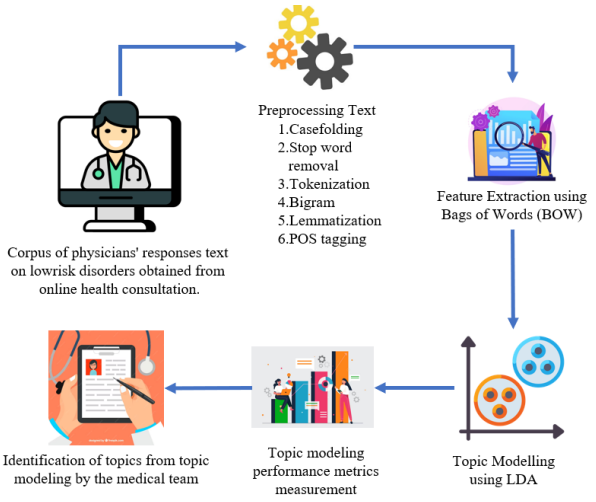

Fig 1. A Proposed Research Framework for Topic Modeling with LDA on Low-Risk Diseases in Online Health Consultation

initial step is to convert all letters in the corpus of Physicians' response text to lowercase. The next step is the word filtering procedure known as the stop word removal procedure. This procedure eliminates words that are unimportant or irrelevant to the research.

The subsequent step is tokenization, which involves splitting text into tokens, such as phrases, paragraphs, and documents. The following stage is bigram [16], which divides a sentence into two words. The next stage is lemmatization, which is the process of standardizing words to determine a word's fundamental form. The next stage is part-of-speech (POS) tagging, which automatically categorizes words formed during the lemmatization stage into word-class labels, i.e., adjectives and nouns.

C. Feature Extraction

Following the data preparation, the Bags of Words (BOW) approach is used to extract features. BOW is a technique for training machine learning models that employs the frequency of occurrence of each word in a document as a feature [17]. This study uses BOW representation because based on research, BOW models are reported to achieve 90% accuracy or higher for various TC tasks [17]. The BOW model creates 1,792,321 tokens in this investigation.

D. Topic Model using LDA

LDA is one of the most often-used for topic modeling approaches [18], [19]. LDA is a Probability model for generating a text data set (corpus). In this study, LDA detects

Keluhan kulit yang merah dan kering pada anak dapat disebabkan oleh dermatitis atopik. Dermatitis atopik adalah peradangan pada kulit yang terjadi pada anak-anak yang menyebabkan munculnya ruam merah yang biasanya terasa gatal. Penyebab terjadinya kondisi ini belum dapat dipastikan tetapi beberapa pemicunya adalah perubahan cuaca. konsumsi makanan/ minuman tertentu, kontak dengan bahan tertentu. Selain karena dermatitis atopik, penyebab lainnya adalah dermatitis kontak (contohnya pada ruam popok). Saran saya, silakan segera bawa ke dokter anak terdekat agar dokter dapatmelakukan pemeriksaan pada kulit anak sehingga dokter dapat memastikan sebabnya sebelum diberikan penanganan yang sesuai. Untuk saat ini, silakan gunakan sabun yang berlabel hypoallergenic*dan hindari penggunaan produk baru dulu untuk anak. Silakan juga menggunakan petroleum jelly untuk mengatasi kulit kering ini. Kenakan pakaian yang berbahan lembut untuk anak, ganti popok anak lebih sering, dan hindari menggunakan tisu basah untuk membersihkan area kelamin dan bokong anak. Semoga membantu*

(Complaints of red and dry skin in children can be caused by atopic dermatitis. Atopic dermatitis is an inflammation of the skin that occurs in children that causes a red rash that usually feels itchy. The cause of this condition is yet to be ascertained, but some triggers are weather changes. Consumption of certain foods/drinks, contact with certain materials. Apart from atopic dermatitis, other causes are contact dermatitis (e.g., diaper rash). My advice is to immediately take it to the nearest pediatrician so that the doctor can examine the child's skin and determine the cause before being given the appropriate treatment. For now, please use soap labeled hypoallergenic and avoid using new products for children. Please also use petroleum jelly to treat this dry skin. Wear clothes made of soft materials for children, change the child's diaper more often, and avoid using wet wipes to clean the genital area and buttocks of the child. I hope this helps)

Fig 3. An Example of A General physicians' Answer Text Dataset on The Topic of Skin Diseases Sourced from free Online Health Consultation services (OHCs) Alodokter.Com.

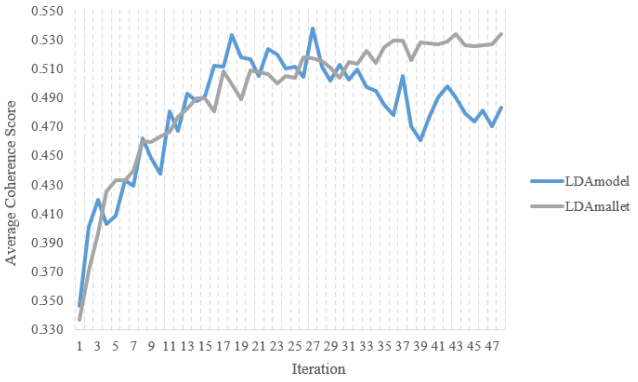

Fig 4. Comparing the average coherence scores of two LDA models, LDAmodel and LDAmallet, on topic modeling of online health consultation text history related to low-risk diseases by analyzing each iteration of both models.

topics in a collection of Physicians' response texts (corpus). The corpus is an observation variable, and the subject of each word in the corpus is a latent variable. The LDA model estimates the latent variable. In this study, we conducted experiments using two LDA models to compare them and find the best LDA model for topic modelling. The two model different libraries are named LDAmodel, which uses the Gensim library, and a model named LDAmallet, which uses the Mallet library.

The LDA approach is combined with the collapsed Gibbs sampling process using the LDAmallet model [5]. The collapsing Gibbs sampling algorithm can group words that have similarities on each topic that is formed so that they can be interpreted into sentences that are more meaningful and informative when compared to calculating the frequency of word occurrences.

In this study, the Gibbs sampling algorithm uses the mallet library version mallet-2.08. The approach for measuring the performance of the topic modeling model uses a coherence score to evaluate the quality of the generated topics [20], [21]. Procedure 1 shows the technical steps at the experimental stage using python for topic modeling using both models: LDAmodel and LDAmallet.

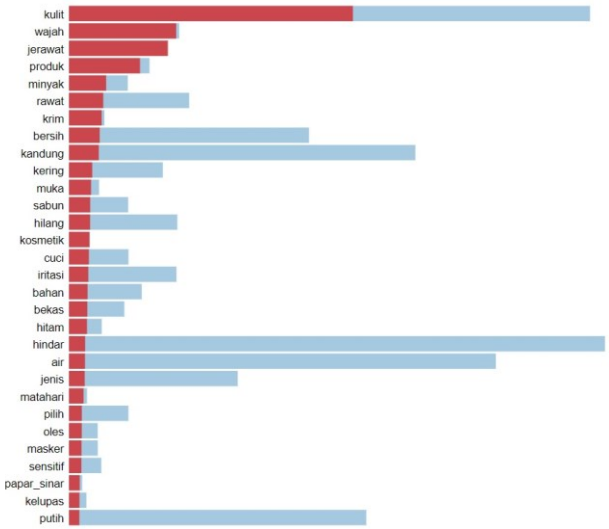

Fig 5. This group of phrases in topic four is recognized as skin illnesses as the outcome of topic modeling using the LDAmallet model based on the text history of physicians' answers.

Procedure 1: Topic Models for Explore Clinical Answer Text

```
Inputs: Corpus = collection of physicians' clinical answers
Outputs: Modeling of Clinical Answer Pattern Topics
Begin
clean_data:
        Apply preprocessing (Corpus)
Function make_bigrams(texts):
   return [bigram_mod[doc] for doc in texts]
apply nlp_id.postag:
        PosTag()
clean_data = lemmatization (PosTag())
apply gensim.corpora import Dictonary
import genism.corpora as corpora:
        id2word=corpora.Dictionary (clean_data)
//LDA Method
Apply LDA = genism.model.ldamulticore.LdaModel:
        corpus=corpus,
        num_topics=n,

Apply coherence_model_LDA ()
        CoherenceModel (model=LDA, texts=clean_Data,
        dictionary (id2word), coherence='c_v')
//LDA Mallet
apply LDAmallet =gensim.models. wrappers.LdaMallet:
        mallet_path,
        corpus=corpus,
        num_topics=n,

Apply coherence_model_LDAMallet ()
        CoherenceModel (model=LDAMallet,
        texts=clean_Data, dictionary (id2word),
        coherence='c_v')
End Procedure
```

In this study, the approach for measuring the performance of the topic modeling model uses a coherence score to evaluate the quality of the generated topics [20], [21].

E. Topic Identification with Medical Expert

In this study, the topic identification process still requires human assistance. At this stage, the output of the topic modeling method with the best performance will be distributed to medical experts. This stage is to identify in

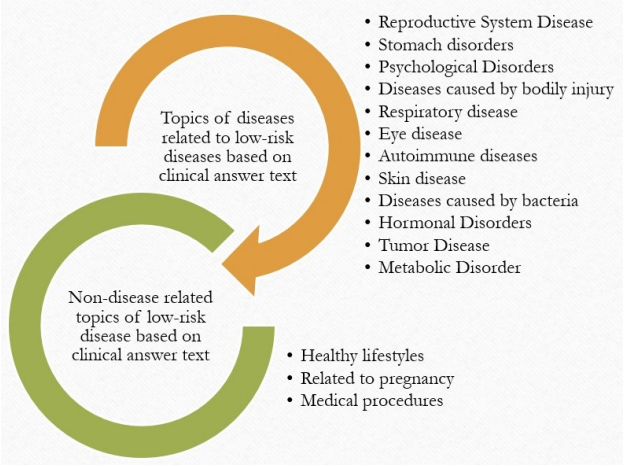

Fig 6. Overview of low-risk disease trends based on Online Health Consultation (OHC) history using Topic Modeling with the LDA method (LDAmallet model) identified by the medical team resulted in 2 large groups: disease topics and non-disease topics.

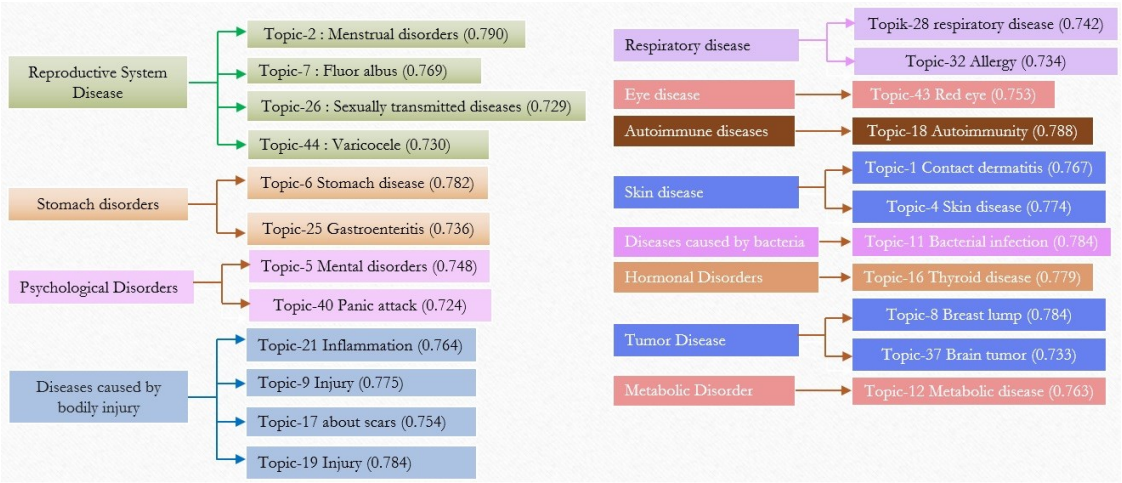

Fig 7. All Topics in the Disease Topic Groups Divided Into Multiple Disease Group Categories Based On Medical Team Identification

depth whether the collection of words on a topic contains a pattern of clinical answers that lead to a group of diseases or a particular type of disease.

III. Result and Discussion

This section discusses the results of the topic modeling research trials, including the results of the performance measurement process on two Latent Dirichlet Allocation (LDA) topic modeling methods named LDAmodel and LDAmallet. To compare their performance, both models were tested and reviewed. Then, using the best model's output, we conduct a complete analysis with the assistance of medical specialists.

A. Comparision Performance of Topic Models

Fig 4 depicts the outcomes of topic modeling studies using two LDA models: the LDA model and the LDAmallet model. At this step, fifty iterations of the topic modeling procedure were performed on both models. Fig 4 shows that after 50 iterations, the LDA model outperforms the LDAmallet model. At the 28th iteration, the LDA model receives a score of 0.538, while the LDAmallet model receives a score of 0.534, which is lower than the LDA model. However, the LDAmallet model has improved with each iteration, but the LDA model has dropped in performance after the 28th iteration. As a result, by adding iterations to both models, the LDAmallet model's average coherence score may be much greater than the LDA models'.

B. Topic Modeling Visualization

A comprehensive evaluation of the effectiveness of the LDAmodel and LDAmallet models is conducted in this paper. This study, however, only demonstrates the outcomes of topic modeling with the LDAmallet, which is more effective than the LDAmodel. Using the results of the LDAmallet model, the medical team attempts to determine the disease associated with each topic. Fig. 5 is an illustration of topic modeling outcomes utilizing the LDAmallet method. The results of the topic modeling use Indonesian terminology because the dataset used in this study is a corpus of general physician responses gathered from the website of free Online Health Consultation services (OHC) in Indonesia.

The detailed description in Fig. 5 represents the words in Topic 4, which medical experts then identified as a collection of terms related to skin diseases. These words include "*kulit*" ("skin"), "*wajah*" ("face"), *"jerawat"* ("acne"), "*produk*" ("product"), "*minyak*" ("oil"), *rawat* ("treat"), "*krim*" ("cream"), "*bersih*" ("clean"), "*kandung*" ("contain"), "*kering*" ("*dry*"), "*muka*" ("face"), "*sabun*" ("soap"), "*hilang*" ("lost"), "*kosmetik*" *("*cosmetic"), "*cuci*" ("wash"), *"iritasi"* ("irritation"), "*bahan*" ("ingredient"), "*bekas*" ("imprint"), "*hitam*" ("black"), "*hindar*" ("avoid"), "*air*" ("water"), "*jenis*" ("type"), *"matahari"* ("sun"), "*pilih*" ("choose"), "*oles*" ("rub"), "*masker*" ("mask"), "*sensitive*" ("sensitive"), "*papar_matahari*" ("sun_exposure"), "*kelupas*" *(*"exfoliate"), "*putih*" *("*white").

C. Low-Risk Disease Trends Based on Online Health Consultation (OHC)

Experiments with the LDAmallet model revealed 33 topics, with each topic having the highest coherence value in each iteration compared to other topics. The 33 subject models are then distributed to medical professionals, who divide them into two groups: disease topics and non-disease topics. The LDAmallet clustering results are shown in Fig 6.

Based on Fig 6. In the first group, namely the disease topic group, there are several sub-topics such as Reproductive System Diseases, Stomach Disorders, Psychological Disorders, Diseases Due to Body Wounds, Respiratory Diseases, Eye Diseases, Autoimmune Diseases, Skin Diseases, Bacterial Diseases, Hormone Disorders, Tumor Diseases, Metabolic Disorders as in the picture. While in the second group, namely the non-disease group, there is a sub-group of a healthy lifestyle related to pregnancy and medical procedures.

D. Topic Identification

The medical professionals then defined and classified low-risk Disease Topics within each disease subject category based on the findings of topic modeling applied to collecting physicians' response texts in OHC. According to the study results of the medical team, 33 topics were recognized as including a collection of phrases. Table 1 provides a detailed explanation of the medical team's identification results. Then, at the subsequent stage, the medical team reconfirmed that the 33 topics may be categorized into two broad categories, namely the disease group and the non-disease group depicted in Fig 7.

TABLE I. DISCOVERY TOPICS GROUPS ON LOW-RISK DISEASES USING THE LDAMALLET METHOD BASED ON THE TEXT OF THE ANSWERS OF GENERAL PHYSICIANS IN THE ONLINE HEALTH CONSULTATION.

Reproductive System Disease	**Topic-7 (Fluor albus)**: white, color, clean, vagina, infection, normal, liquid, wear, bacteria, organ, guard, avoid, smell, itching, dressing, female, fungus, pants, mucus, intimate, yellow, area, soap, shy, tight, dressing, change, water, brown, area
	Topic-2 (Menstrual disorders): menstruation, cycle, hormone, uterus, birth_control, hormonal, female, contraception, disrupt, inject, bladder, regulate, normal, blood, birth_control_pills, device, slow, side_effect, pill, change, endometriosis, reproduction, nature, pelvis, wall, usg, stop, menopause, iud, ovarian_cyst
	Topic-26 (Sexually transmitted diseases): intercourse, sexual, vagina, behavior, contagious, penis, intimate, tide, genitals, sex, liquid, cervix, gonorrhea, cancer, tool, woman, condom, syphilis, organ, genital_warts, lubricated, replace, safe, masturbation, swipe, excitatory, hope, blood, uterus, hpv
	Topic-44 (Varicocele): behavior, appropriate, therapy, assist, suggest, consultation, hope, advice, avoid, need, specialist, medicine, direct, testicles, routine, condition, evaluation, minimum, basic, primary, tips, varicocele, scrotum, penis, ER, maximum, pack, management, indication
Stomach disorders	**Topic-6 (Stomach disease)**: meal, stomach, acid, avoid, pain, spicy, consumption, drink, ulcer, fat, alcohol, coffee, gastric_ulcer, portion, caffeine, soda, gastritis, lying, cigarettes, esophagus, digest, level, ulcer, oil, adjust, irritation, nausea, endoscopy, medicine, dyspepsia
	Topic-25 (Gastroenteritis): eat, drink, vomit, nausea, diarrhea, consumption, digest, consume, appetite, liquid, dehydration, easy, toxic, hours, hope, try, advice, headache, avoid, rest, lighthearted, attention, down, gastroenteritis, portion, type, house, frequency, intake, process
	eat, drink, vomit, nausea, diarrhea, consume, digest, consume, appetite, liquid, dehydration, easy, poison, hour, may, try, advice, headache, avoid, rest, light, mindful, down, gastroenteritis, portion, type, home, frequency, intake, olah
Psychological Disorders	**Topic-5 (Mental disorders)**: disturb, sleep, person, anxious, thought, difficult, depression, friend, try, nature, across, work, soul, fear, behavior, change, activity, mental, life, easy, psychological, talk, psychiatrist, awake, clock, positive, night, personal, property, family
	Topic-40 (Panic attack): heart, chest, disturb, pain, check, sweat, lung, tightness, muscle, weight, left, pounding, anxious, physical, breath, attack, breath, symptom, press, activity, fail, record, cold, hope, ekg, rest, arm, X-ray, panic_attack, spread
Diseases caused by bodily injury	**Topic-21 (Inflammation)**: pain, subside, inflammation, warm, compress, neck, muscle, medication, region, paracetamol, area, comfortable, intense, rest, water, consume, over-the-counter, effect, mild, paracetamol, cold, sweet, endure, paracetamol_ibuprofen, consume, massage, sore, contraction, heavy, painkiller
	Topic-9 (Injury): appear, thanks, help, direct, disturb, street, hand, advice, complain, enter, danger, condition, step, result, spark, hope, activity, side, diligent, specialist, natural, nerve, refer, hook, injury, effort, sign, activity
	Topic-17 (about scars): wound, surgery, heal, act, trace, attitude, clean, process, infection, care, hand, close, burn, post, surgery, net, procedure, open, sewing, pus, liquid, complication, quick, guard, shape, broad, keloid, dirty, risk, post
	Topic-19 (Injury): bone, muscle, injury, position, motion, nerve, behavior, fracture, spine, activity, weight, body, sit, waist, joints, trauma, stretch, clasps, train, area, wrong, exercise, strong, posture, nets, thigh, severe, shoulder, bump, stiff
Respiratory disease	**Topik-28 (Respiratory Disease)**: cough, throat, breath, airway, breath, congested, inflammation, infection, lung, asthma, cigarette, water, cold, phlegm, smoke, drink, home, tb, symptom, flu, tonsils, swallow, warm, tuberculosis, check, sick, air, pneumonia, difficult, virus
	Topic-32 (Allergy): allergy, nose, cold, reaction, avoid, exposure, trigger, across, medicine, water, plug, air, cavity, sinusitis, inflammation, rhinitis, symptom, substance, trigger, dust, sinus, appear, polyp, hot, hand, complain, cold, body, clean, nosebleed
Eye disease	**Topic-43 (Red eye):** eye, vision, red, petal, disturb, light, ball, glasses, hand, drop, cornea, dry, sharp, specialist, water, lens, tired, condition, contact_lens, hope, object, layer, sellable, minus, distance, protected, irritation, membrane, bright, due
Autoimmune diseases	**Topic-18 (Autoimmunity)**: sick, symptom, appearance, suffer, appear, natural, attack, experience, sign, disappear, history, mild, chronic, major, diagnosis, wrong, severe, difficult, arise, relapse, hello, autoimmune, typical, bad, syndrome, nature, term, familiar, cover, gather
Skin disease	**Topic-1 (Contact dermatitis)**: skin, itching, infection, wear, red, rash, fungus, clean, allergy, scratch, bath, irritation, area, material, freckle, appear, avoid, dry, water, soap, keep, contact_dermatitis, insect_bite, contact, area, sweat, wash, replace, spot, direct
	Topic-4 (Skin disease): skin, facial, acne, product, oil, care, cream, clean, bladder, dry, face, soap, lose, cosmetics, wash, irritation, material, used, black, dodge, water, type, sun, select, smear, mask, sensitive, exposure, peel, white
Diseases caused by bacteria	**Topic-11 (Bacterial infection)**: medicine, consume, side-effects, antibiotic, dose, prescription, class, suitable, recommend, anti, consume, type, stop, over-the-counter, look, watch, bacteria, indication, buy, tablet, hang, effect, work, cure, combination, term, prevent, fill, control, corticosteroid
Hormonal Disorders	**Topic-16 (Thyroid disease)**: weight, body, down, body, thyroid, level, ideal, exercise, hormone, normal, activity, diet, obesity, gland, nutrition, regulate, calories, eat, drastic, content, factor, nutrition, balance, influence, intake, grease, fat, metabolism, fast, in
Tumor Disease	**Topic-8 (Breast lump)**: lump, breast, cancer, large, cyst, tumor, gland, clear, lymph_nodes, swelling, web, malignant, neck, abscess, measure, shape, change, nipple, biopsy, surgery, examine, fill, liquid, palpate, lipoma, ulcer, benign_tumor, pus, infection, growing
	Topic-37 (Brain tumor): annoy, nerves, brain, headache, body, vertigo, seizure, condition, change, conscious, draw, symptom, lost, natural, press, low, motion, sensation, rotate, ant, medicine, position, tumor, system, down, check, sudden, faint, function, hopefully
Metabolic Disorder	**Topic-12 (Metabolic disease)**: blood, press, reed, levels, body, function, anemia, cells, down, red, levels, diabetes, hypertension, normal, kidney, low, liver, sugar, damaged, flow, sugar_level, frozen, group, stop, fail, rupture, control, platelets, suffering, complications

E. Discussion

Based on the results of experiments using LDA models sourced from 2 different libraries, it was found that the LDA model was superior to LDAmallet in the initial iterations until the 28th iteration. However, after the 28th iteration, the performance of the LDAmallet model became superior, so it can be concluded that LDAmallet has better advantages for large and complex datasets, such as the dataset we used in this study. Nevertheless, based on this study, we found that the advantages of LDAmallet can be seen if the iteration process is done more or more than 25 times to find the best cluster from a large dataset.

Based on our findings, topic modeling can be a helpful method for organizing and analyzing vast volumes of text data in online health consultations using LDAmallet, which is expected to help practical applications such as analyzing online health consultation trends and the relationship between symptoms and diseases.

IV. CONCLUSION

This study investigates physicians' response texts about low-risk diseases in Online Health Consultation (OHC) with

a topic modeling approach to identify topics and trends to improve personalized information services for Indonesian users who conduct health consultations for low-risk diseases. This study employs the Latent Dirichlet Allocation (LDA) model with two alternative implementations during the testing phase: LDAmodel and LDAmallet.

The experimental results show that the LDA model using the LDAmallet method performs slightly better than other models named LDAmodel. Based on the topic modeling results, Physician[i]s identified two topic trends in OHC for low-risk diseases (12 diseases) and non-diseases, such as healthy lifestyle, pregnancy, and medical procedures. In future research, we would like to continue our experimental process in the form of ontology visualization.

ACKNOWLEDGMENT

This work was partially supported by funds from the University Center of Excellence on Artificial Intelligence for Healthcare and Society (UCE AIHeS) at Institut Teknologi Sepuluh Nopember, Indonesia, and funded by the Ministry of Research and Technology/National Research and Innovation Agency of the Republic of Indonesia. The author would also like to thank Rizaldy Ramadhan for assisting in identifying research results on topic modeling.

REFERENCES

[1] D. Dicker et al., "Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: A systematic analysis for the Global Burden of Disease Study 2017," Lancet, vol. 392, no. 10159, pp. 1684–1735, 2018, doi: 10.1016/S0140-6736(18)31891-9.

[2] H. Ritchie and M. Roser, "Causes of death," OurWorldInData.org, 2018. https://ourworldindata.org/causes-of-death (accessed May 01, 2022).

[3] World Health Organization, "Noncommunicable diseases," World Health Organisation, 2021. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (accessed Jun. 27, 2022).

[4] Z. Gong, Z. Han, X. Li, C. Yu, and J. D. Reinhardt, "Factors Influencing the Adoption of Online Health Consultation Services: The Role of Subjective Norm, Trust, Perceived Benefit, and Offline Habit," Front. Public Heal., vol. 7, pp. 1–9, 2019, doi: 10.3389/fpubh.2019.00286.

[5] World Health Organization, "The top 10 causes of death - Factsheet," WHO reports, 2020. https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[6] GLOBOCAN, "The Global Cancer Observatory - All cancers," International Agent for Research on Cancer - WHO, 2020. https://gco.iarc.fr/today/data/factsheets/populations/360-indonesia-fact-sheets.pdf (accessed Apr. 21, 2020).

[7] J. Wang, L. Wang, J. Xu, and Y. Peng, "Information Needs Mining of COVID-19 in Chinese Online Health Communities," Big Data Res., vol. 24, p. 100193, 2021, doi: 10.1016/j.bdr.2021.100193.

[8] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis," Front. Artif. Intell., vol. 3, pp. 1–14, 2020, doi: 10.3389/frai.2020.00042.

[9] Q. Liu et al., "Health communication through news media during the early stage of the covid-19 outbreak in China: Digital topic modeling approach," J. Med. Internet Res., vol. 22, no. 4, pp. 1–12, 2020, doi: 10.2196/19118.

[10] S. Juanita, D. Purwitasari, I. K. E. Purnama, and M. H. Purnomo, "Identifying Precautionary Measures for High-Risk Disease from Doctor's Answer Text Using LDA," in 2021 International Electronics Symposium (IES), 2021, pp. 41–46.

[11] Q. Meng and H. Xiong, "A Doctor Recommendation Based on Graph Computing and LDA Topic Model," Int. J. Comput. Intell. Syst., vol. 14, no. 1, pp. 808–817, 2021, doi: 10.2991/ijcis.d.210205.002.

[12] G. Yang, "A Text Study of Online Medical Reviews of Traditional Chinese Medicine Based on LDA Model," in 2nd International Conference on Artificial Intelligence and Communication Technology (AICT 2023), 2023, vol. 56, pp. 261–270.

[13] M. Zhang, Y. Sun, X. Zhao, L. Wang, and J. Xiong, "The Impact of Narrative Reviews on Patient E-doctor Choice in Online Health Communities," Inquiry, vol. 60, pp. 1–9, 2023, doi: 10.1177/00469580231183695.

[14] A. P. Widyassari et al., "Review of automatic text summarization techniques & methods," J. King Saud Univ. - Comput. Inf. Sci., 2020, doi: 10.1016/j.jksuci.2020.05.006.

[15] Z. Rahimi and M. M. Homayounpour, "The impact of preprocessing on word embedding quality : a comparative study," Lang. Resour. Eval., vol. 57, pp. 257–291, 2022, doi: 10.1007/s10579-022-09620-5.

[16] N. Nasser, L. Karim, A. El Ouadrhiri, A. Ali, and N. Khan, "n-Gram based language processing using Twitter dataset to identify COVID-19 patients," Sustain. Cities Soc., vol. 72, no. May, p. 103048, 2021, doi: 10.1016/j.scs.2021.103048.

[17] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," PLoS One, vol. 15, no. 5, pp. 1–22, 2020, doi: 10.1371/journal.pone.0232525.

[18] H. Jelodar et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," Multimed. Tools Appl., vol. 78, no. 11, pp. 15169–15211, 2019, doi: 10.1007/s11042-018-6894-4.

[19] M. Bhattacharya, C. Jurkovitz, and H. Shatkay, "Co-occurrence of medical conditions: Exposing patterns through probabilistic topic modeling of snomed codes," J. Biomed. Inform., vol. 82, no. January, pp. 31–40, 2018, doi: 10.1016/j.jbi.2018.04.008.

[20] M. Rudiger, D. Antons, A. M. Joshi, and T.-O. Salge, "Topic modeling revisited : New evidence on algorithm performance and quality metrics," PLoS One, vol. 17, no. 4, pp. 1–25, 2022, doi: 10.1371/journal.pone.0266325.

[21] R. M. Omizo, "A Comparison of Topic Modeling Approaches Using Networked Discussion Forum Posts From the City-data.com Corpus," J. Open Humanit. Data, vol. 10, no. 16, pp. 1–12, 2024, doi: 10.5334/johd.182.

25th ISITIA 2024

CERTIFICATE

It is hereby certified that

Safitri Juanita

has contributed as

PRESENTER

of the paper with the title :

Topic Modeling for Online Health Consultation on Low-Risk Diseases

at the 25th International Seminar on Intelligent Technology and Its Applications (ISITIA) 2024
on Hybrid Conference, 10-12 July 2024, in Lombok, Indonesia.

International Seminar on Intelligent Technology and Its Applications
ISITIA

Dr. Puji Handayani

General Chair