

Marina Yusoff · Tao Hai ·
Murizah Kassim · Azlinah Mohamed ·
Eisuke Kita (Eds.)

Communications in Computer and Information Science

1771

Soft Computing in Data Science


7th International Conference, SCDS 2023
Virtual Event, January 24–25, 2023
Proceedings

 Springer




Communications in Computer and Information Science 1771

Editorial Board Members

Joaquim Filipe , *Polytechnic Institute of Setúbal, Setúbal, Portugal*

Ashish Ghosh , *Indian Statistical Institute, Kolkata, India*

Raquel Oliveira Prates , *Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil*

Lizhu Zhou, *Tsinghua University, Beijing, China*

Rationale

The CCIS series is devoted to the publication of proceedings of computer science conferences. Its aim is to efficiently disseminate original research results in informatics in printed and electronic form. While the focus is on publication of peer-reviewed full papers presenting mature work, inclusion of reviewed short papers reporting on work in progress is welcome, too. Besides globally relevant meetings with internationally representative program committees guaranteeing a strict peer-reviewing and paper selection process, conferences run by societies or of high regional or national relevance are also considered for publication.

Topics

The topical scope of CCIS spans the entire spectrum of informatics ranging from foundational topics in the theory of computing to information and communications science and technology and a broad variety of interdisciplinary application fields.

Information for Volume Editors and Authors

Publication in CCIS is free of charge. No royalties are paid, however, we offer registered conference participants temporary free access to the online version of the conference proceedings on SpringerLink (<http://link.springer.com>) by means of an http referrer from the conference website and/or a number of complimentary printed copies, as specified in the official acceptance email of the event.

CCIS proceedings can be published in time for distribution at conferences or as postproceedings, and delivered in the form of printed books and/or electronically as USBs and/or e-content licenses for accessing proceedings at SpringerLink. Furthermore, CCIS proceedings are included in the CCIS electronic book series hosted in the SpringerLink digital library at <http://link.springer.com/bookseries/7899>. Conferences publishing in CCIS are allowed to use Online Conference Service (OCS) for managing the whole proceedings lifecycle (from submission and reviewing to preparing for publication) free of charge.

Publication process

The language of publication is exclusively English. Authors publishing in CCIS have to sign the Springer CCIS copyright transfer form, however, they are free to use their material published in CCIS for substantially changed, more elaborate subsequent publications elsewhere. For the preparation of the camera-ready papers/files, authors have to strictly adhere to the Springer CCIS Authors' Instructions and are strongly encouraged to use the CCIS LaTeX style files or templates.

Abstracting/Indexing

CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compindex, Mathematical Reviews, SCImago, Scopus. CCIS volumes are also submitted for the inclusion in ISI Proceedings.

How to start

To start the evaluation of your proposal for inclusion in the CCIS series, please send an e-mail to ccis@springer.com.

Marina Yusoff · Tao Hai · Murizah Kassim ·
Azlinah Mohamed · Eisuke Kita
Editors

Soft Computing in Data Science

7th International Conference, SCDS 2023
Virtual Event, January 24–25, 2023
Proceedings

Editors

Marina Yusoff 

Universiti Teknologi MARA
Shah Alam, Malaysia

Murizah Kassim 

Universiti Teknologi MARA
Shah Alam, Malaysia

Eisuke Kita 

ISSN 1865-0929

Communications in Computer and Information Science

ISBN 978-981-99-0404-4

<https://doi.org/10.1007/978-981-99-0405-1>

Institute of Liberal Arts and Sciences
Nagoya, Japan

Tao Hai 

Baoji University of Arts and
Sciences
Baoji, China

Azlinah Mohamed 

Universiti Teknologi MARA
Shah Alam, Malaysia

ISSN 1865-0937 (electronic)

ISBN 978-981-99-0405-1 (eBook)

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Singapore Pte Ltd. 2023

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Welcome to the proceedings of the Seventh International Conference on Soft Computing in Data Science (SCDS 2023). SCDS 2023 was a virtual conference co-hosted by China's Baoji University of Arts and Sciences. 'Science in Analytics: Harnessing Data and Simplifying Solutions' was the conference's theme. SCDS 2023 aimed to provide a knowledge-sharing platform for Big Data Analytics and Artificial Intelligence theory and applications. Big Data Analytics enables businesses to better leverage data for data-driven decisions and AI for optimal performance. The world is shifting toward automation and innovation through Artificial Intelligence. Experts in

industry are capitalizing on emerging technologies such as artificial intelligence, big data, and cloud computing.

The papers in these proceedings address issues, challenges, theory, and novel applications of big data analytics and artificial intelligence, including but not limited to Artificial Intelligence techniques and applications, Computing and Optimization, Data Mining and Image Processing, and Machine and Statistical Learning.

In the advanced society of the twenty-first century, knowledge and technology on Big Data and Artificial Intelligence must be transferred to industrial applications and used to solve real-world problems that benefit the global community. Collaborations between academia and industry in research can result in novel, innovative computing applications that enable real-time insights and solutions.

We are thrilled to have received paper submissions from a diverse group of national and international researchers this year. We received 61 paper submissions, with 31 of them being accepted. The SCDS 2023 review process was double-blind. All accepted submissions were assigned to at least two independent reviewers to ensure a rigorous, thorough, and convincing evaluation process. The review process included 15 international and 30 local reviewers. The conference proceeding volume editors and the Springer CCIS Editorial Board made the final acceptance decisions, with 24 of the 61 submissions (39.34%) published in the conference proceedings.

We would like to express our gratitude to the authors who submitted manuscripts to SCDS 2023. We appreciate the reviewers taking the time to read the papers. We gratefully acknowledge all conference committee members for their valuable time, ideas, and efforts in ensuring the success of SCDS 2023. We would also like to thank the Springer CCIS Editorial Board, organizations, and sponsors for their ongoing support. We sincerely hope SCDS 2023 has provided a forum for sharing knowledge, publishing promising research findings, and forming new research collaborations. We hope everyone

took something away from the Keynote and Parallel Sessions and enjoyed interacting with other researchers via this virtual conference.

January 2023

Marina Yusoff
Tao Hai
Murizah Kassim
Azlinah Mohamed
Eisuke Kita

Organization

Advisor

Jasni Mohamad Zain

Universiti Teknologi MARA, Malaysia

Conference Chair

Marina Yusoff

Universiti Teknologi MARA, Malaysia

Co-chair

Hai Tao

Baoji University of Arts and Sciences, China

Secretaries

Nurain Ibrahim (Head)

Universiti Teknologi MARA, Malaysia

Siti Afiqah Muhamad Jamil

Universiti Teknologi MARA, Malaysia

Registration and Presentation

Lu Ye

Baoji University of Arts and Sciences, China

Fang Xu

Baoji University of Arts and Sciences, China

Yang Xuelan

Baoji University of Arts and Sciences, China

Wan Fairos Wan Yaacob

Universiti Teknologi MARA, Malaysia

Finance Committee

Nurbaizura Borhan (Head)

Universiti Teknologi MARA, Malaysia

Nur Asyikin Abdullah

Universiti Teknologi MARA, Malaysia

Azira Mohamed Amin

Universiti Teknologi MARA, Malaysia

Organization

Technical Program Committee

Murizah Kassim (Head)

Universiti Teknologi MARA, Malaysia

Abdul Kadir Jumaat

Universiti Teknologi MARA, Malaysia

Mohd Abdul Talib Mat Yusoh

Universiti Teknologi MARA, Malaysia

Publicity & Sponsorship Committee

Memiyanty Abdul Rahim (Head) Universiti Teknologi MARA, Malaysia

Nooritawati Md Tahir Universiti Teknologi MARA, Malaysia Muhammad Imran

Mohd Tamrin Universiti Teknologi MARA, Malaysia

Siti Nurul Aslinda Mohd Universiti Teknologi MARA, Malaysia Nahrowei

Mario Köppen

Sri Hartati

Chidchanok Lursinsap

Kyutech Institute of Technology, Japan

Universitas Gadjah Mada, Indonesia

Chulalongkorn University, Thailand

Hizir Sofyan
Sofiya Ahmad Kamal

Syiah Kuala University, Indonesia Nur Yasmin
Universiti Teknologi MARA, Malaysia

Publication Committee (Program/Book of Abstracts)

Siti Afiqah Muhamad Jamil Universiti Teknologi MARA, Malaysia (Head)
Mohammed Adam Kunna Azrag Universiti Teknologi MARA, Malaysia Saiful
Farik Mat Yatim Universiti Teknologi MARA, Malaysia
Yang Xuelan Baoji University of Arts and Sciences, China
Nur Aliah Khairina Mohd Haris UiTM Technoventure Sdn. Bhd, Malaysia Putri
Azmira R Azmi UiTM Technoventure Sdn. Bhd, Malaysia

Website & Online Committee

Mohamad Asyraf Abdul Latif Universiti Teknologi MARA, Malaysia
Mohamad Ridhwan Mohamad Universiti Teknologi MARA, Malaysia
Razali

International Scientific Committee

	Organization	ix
Aiden Doherty	University of Oxford, UK	
Hamid R. Rabiee	Sharif University of Technology, Iran	
Ruhana Ku Mahamud	Universiti Utara Malaysia, Malaysia	
Nur Iriawan	Institut Teknologi Sepuluh Nopember, Indonesia	
Rajalida Lipikorn	Chulalongkorn University, Thailand	
Agus Harjoko	Universitas Gadjah Mada, Indonesia	
Min Chen	University of Oxford, UK	
Simon Fong	University of Macau, China	
Mohammed Bennamoun	University of Western Australia, Australia	
Yasue Mitsukura	Keio University, Japan	
Dhiya Al-Jumeily	Liverpool John Moores University, UK	
Dariusz Krol	Wroclaw University of Science and Technology, Poland	
Richard Weber	University of Chile, Chile	
Jose Maria Pena	Technical University of Madrid, Spain	
Yusuke Nojima	Osaka Prefecture University, Japan	
Siddhivinayak Kulkarni	University of Ballarat, Australia	

Tahir Ahmad	Universiti Teknologi Malaysia, Malaysia
Daud Mohamed	Universiti Teknologi MARA, Malaysia
Wahyu Wibowo	Institut Teknologi Sepuluh Nopember, Indonesia
Edi Winarko	Universitas Gadjah Mada, Indonesia
Retantyo Wardoyo	Universitas Gadjah Mada, Indonesia
Rayner Alfred	Universiti Malaysia Sabah, Malaysia
Faiz Ahmed	.
Mohamed Elfaki	Qatar University, Qatar
Mohamed Chaouch	Qatar University, Qatar
Abdul Haris Rangkuti, S.	BINUS University, Indonesia
Eisuke Kita	Nagoya University, Japan
Waheb Abdullah	.
Abduljabbar Shaif	Birmingham City University, UK
Achmad Choiruddin	Institut Teknologi Sepuluh Nopember, Indonesia
Jerry Dwi Purnomo	Institut Teknologi Sepuluh Nopember, Indonesia

Reviewers SCDS 2023

Abdul Kadir Jumaat	Universiti Teknologi MARA, Malaysia
Albert Guvenis	Bog̃aziçi University, Turkey
Azlan Iqbal	Universiti Tenaga Nasional & College of Information Technology, Malaysia
Dipti Prakash Theng	G. H. Rasoni College of Engineering, India
Ensar Gul Organization	Maltepe University, Turkey
Fadhlan Hafizhelmi Kamaru Zaman	Universiti Teknologi MARA, Malaysia
Faroudja Abid	Centre for Development of Advanced Technologies, CDTA, Algeria
Gandhiya Vendhan S.	Bharathiar University, India
Hamidah Jantan	Universiti Teknologi MARA, Malaysia
Hasan A. Alsarraf	Northern Technical University, Iraq
Hasniaty A.	Universitas Hasanuddin, Indonesia
Hoo Yann Seong	Universiti Pertahanan Nasional Malaysia, Malaysia
Jasni Mohamad Zain	Universiti Teknologi MARA, Malaysia
Khairul Anwar Sedek	Universiti Teknologi MARA, Malaysia
Marina Yusoff	Universiti Teknologi MARA, Malaysia
Marshima Mohd Rosli	Universiti Teknologi MARA, Malaysia
Mazlina Ibrahim	Universiti Pertahanan Nasional Malaysia, Malaysia
Mohd Abdul Talib Mat Yusoh	Universiti Teknologi MARA, Malaysia

Muhammad Firdaus Mustapha
Murali Krishna Enduri
Murizah Kassim
Noor Azilah Muda
Nor Aziyatul Izni Mohd Rosli
Norhaslinda Kamaruddin
Noriko Etani
Nur Intan Raihana Ruhaiyem
Peraphon Sophatsathit
Rahmad Hidayat

Richard C. Millham
Rizauddin Saian
Rizwan Aslam Butt

Rodrigo Campos Bortoleeto
Shafaf Ibrahim
Shamimi A. Halim
Siow Hoo Leong
Siti Afiqah Muhamad Jamil
Smita Rath
Sofianita Mutalib
Suraya Masrom
Syerina Azlin Md Nasir
Wan Fairos Wan Yaacob

Weng Siew Lam

Yusnani Mohd Yusoff
Yuzi Mahmud
Zainura Idrus

Organized by

Universiti Teknologi MARA, Malaysia
SRM University AP, India
Universiti Teknologi MARA, Malaysia
Universiti Teknikal Malaysia Melaka, Malaysia
UCSI University, Malaysia
Universiti Teknologi MARA, Malaysia
Kyoto University, Japan
Universiti Sains Malaysia, Malaysia
Chulalongkorn University, Thailand
Politeknik Negeri Lhokseumawe & Universitas
Gadjah Mada, Indonesia
Durban University of Technology, South Africa
Universiti Teknologi MARA, Malaysia
NED University of Engineering and
Technology, Pakistan
Instituto Federal de São Paulo, Brazil
Universiti Teknologi MARA, Malaysia
Universiti Teknologi MARA, Malaysia
Universiti Teknologi MARA, Malaysia
Universiti Teknologi MARA, Malaysia
Siksha 'O' Anusanshan SOA University, India
Universiti Teknologi MARA, Malaysia
Universiti Teknologi MARA, Malaysia
Universiti Teknologi MARA, Malaysia
Universiti Teknologi MARA, Malaysia

Organization

xi



Industry and International Collaboration



**Universitas
Gadjah Mada**



Contents

Artificial Intelligence Techniques and Applications

Explainability for Clustering Models	3
<i>Mahima Arora and Ankush Chopra</i>	

Fault Diagnosis Methods of Deep Convolutional Dynamic Adversarial Networks	18
<i>... Tao Hai and Fuhao Zhang</i>	

Carbon-Energy Composite Flow for Transferred Multi-searcher Q-Learning Algorithm with Reactive Power Optimization	32
<i>Jincheng Zhou and Hongyu Xue</i>	

Multi-source Heterogeneous Data Fusion Algorithm Based on Federated Learning	46
<i>... Jincheng Zhou and Yang Lei</i>	

Dynamic Micro-cluster-Based Streaming Data Clustering Method for Anomaly Detection	61
<i>... Xiaolan Wang, Md Manjur Ahmed, Mohd Nizam Husen, Hai Tao, and Qian Zhao</i>	

Federated Ensemble Algorithm Based on Deep Neural Network	76
<i>Dan Wang and Ting Wang</i>	

Performance Comparison of Feature Selection Methods for Prediction in Medical Data	92
<i>... Nur Hidayah Mohd Khalid, Amelia Ritahani Ismail, Normaziah Abdul Aziz, and Amir Aatieff Amir Hussin</i>	

An Improved Mask R-CNN Algorithm for High Object Detection Speed and Accuracy	107
<i>... Qingchuan Liu, Muhammad Azmi Ayub, Fazlina Ahmat Ruslan, Mohd Nor Azmi Ab Patar, and Shuzlina Abdul-Rahman</i>	

Computing and Optimization

Federated Learning with Class Balanced Loss Optimized by Implicit	
---	--

Stochastic Gradient Descent.	121
<i>Jincheng Zhou and Maoxing Zheng</i>	
xiv	Contents
Electricity Energy Monitoring System for Home Appliances Using Raspberry Pi and Node-Red	136
..	
<i>Nur Shazwany Zamani, Muhammad Shafie Mohd Ramli, Khairul Huda Yusof, and Norazliani Md. Sapari</i>	
Short-Time Fourier Transform with Optimum Window Type and Length: An Application for Sag, Swell and Transient.	147
<i>Muhammad Sufyan Safwan Mohamad Basir, Nur-Adibah Raihan Affendy, Mohamad Azizan Mohamad Said, Rizalafande Che Ismail, and Khairul Huda Yusof</i>	
Data Analytics and Technologies	
Cox Point Process with Ridge Regularization: A Better Approach for Statistical Modeling of Earthquake Occurrences	163
<i>Alissa Chintyana, Achmad Choiruddin, and Sutikno</i>	
Discovering Popular Topics of Sarawak Gazette (SaGa) from Twitter Using Deep Learning	178
. . . <i>Nur Ain Binti Nor Azizan, Suhaila Binti Sae, and Muhammad Abdullah Bin Yusof</i>	
Comparison Analysis of LSTM and CNN Variants with Embedding Word Methods for Sentiment Analysis on Food Consumption Behavior	193
<i>Nurul Izleen Ramzi, Marina Yusoff, and Norzaidah Md Noh</i>	
Data Mining and Image Processing	
Towards Robust Underwater Image Enhancement	211
<i>Jahroo Nabila Marvi and Laksmi Rahadianti</i>	
A Comparative Study of Machine Learning Classification Models on Customer Behavior Data	222
. . . <i>Nur Ida Aniza Rusli, Farizuwana Akma Zulkifle, and Intan Syaherra Ramli</i>	
Performance Evaluation of Deep Learning Algorithms for Young	

and Mature Oil Palm Tree Detection232
Soh Hong Say, Nur Intan Raihana Ruhaiyem, and Yusri Yusup

Object Detection Based Automated Optical Inspection of Printed Circuit
Board Assembly Using Deep Learning246
Ong Yee Chiun and Nur Intan Raihana Ruhaiyem

Contents xv

Mathematical and Statistical Learning

The Characterization of Rainfall Data Set Using Persistence Diagram and
Its Relation to Extreme Events: Case Study of Three Locations
in Kemaman, Terengganu 261
..
Z. A. Hasan and R. U. Gobithaasan

Clustering Stock Prices of Industrial and Consumer Sector Companies in
Indonesia Using Fuzzy C-Means and Fuzzy C-Medoids Involving ACF
and PACF 275
.. . . . *Muhammad Adlansyah Muda, Dedy Dwi Prastyo, and Muhammad Sjahid
Akbar*

Zero-Inflated Time Series Model for Covid-19 Deaths in Kelantan Malaysia . . . 291
*Muhammad Hazim Ismail, Hasan Basri Roslee,
Wan Fairos Wan Yaacob, and Nik Nur Fatin Fatihah Sapri*

Author Index 303
.. .

Artificial Intelligence Techniques and Applications



Discovering Popular Topics of Sarawak Gazette (SaGa) from Twitter Using Deep Learning

Nur Ain Binti Nor Azizan, Suhaila Binti Saeed^(✉),
and Muhammad Abdullah Bin Yusof

Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
{68283,21020459}@siswa.unimas.my, ssuhaila@unimas.my

Abstract. The emergence of social media as an information-sharing platform is progressively increasing. With the progress of artificial intelligence, it is now feasible to analyze historical document from social media. This study aims to understand more about how people use their social media to share the content of the Sarawak Gazette (SaGa), one of the valuable historical documents of Sarawak. In the study, a short text of Tweet corpus relating to SaGa was built (according to some keyword search criteria). The Tweet corpus will then be analyzed to extract the topic based on a topic modeling, specifically, Latent Dirichlet Allocation (LDA). Then, the topics will be further classified with Convolutional Neural Network (CNN) classifier.

Keywords: Topic modeling · Twitter analysis · Sarawak Gazette

1 Introduction

The introduction of new media such as Twitter, TikTok, Youtube, and Instagram impacted how society started sharing information. The Twitter platform, for instance, is widely used by people as it focuses on communication and virtual community-based network [5].

Some people shared historical context in their Tweets and comments with different aims, such as educating and reminiscing memories [10]. With so many user-generated posts, social media offers an extensive environment to learn about past and ongoing events significant to the flow of history. As there is seamless sharing, it is now possible to analyze historical content from social media with the advancement of artificial intelligence.

There are many considerable amounts of research that have been done to extract history-oriented content data from social media. [10] used machine learning to conduct an extensive comprehensive analysis of history-oriented content in microblogs to examine the peculiarities of tweets about history. Be that as it may, there is a low amount of studies done on Malaysian historical documents. This study will be using Sarawak Gazette as a medium for historical documents to understand historical document-oriented content in social media.

The study aims to build a short-text corpus of SaGa from Twitter and discover the pattern of popular topics from the corpus. It gathers information to preserve and immerse one's memories in society and participation in the historical subject in their social media; in this case, SaGa. As a result, researchers can examine how people refer to the past, why such connections emerge, and what functions they provide.

2 Related Works

2.1 Data Acquisition

In recent years, several types of research have been conducted that focus on discovering a certain topic from the Twitter corpus. However, from the collections of similar articles, only a few of them attempted to come across the historical topics [3, 10] or past events topics [9]; [13].

Several of the notable publications assessed use history-related and topic-modeling Twitter analysis, in which Tweets are examined using keywords or hashtags that have the historical context or a certain topic as features. A few relevant studies also jumped straight to the process and did not define the data acquisition procedure.

Most of the reviews of the literature focus on how academics used Twitter API as a data-collecting tool. The study from [3] analyzed Twitter profiles for the important historical events Tweet: – the 2nd Spanish Republic, The Spanish Civil War, and the Franco Regime. The research utilized Twitter API. Twitter provides a free Application Programming Interface (API), which is very useful to retrieve and analyze Twitter data programmatically and develop for the Twitter conversation. In addition to that, a technological aid, the tool t-hoarder kit_2, was employed. It is open-source software that allows developers to build and analyze Twitter data, making it easier to utilize in the analysis and data visualization.

In the study, [10] conduct an extensive comprehensive analysis of history-oriented content on Twitter to examine the peculiarities of Tweets about history. The research also utilized Twitter API. To collect the data from Tweets that imply history-oriented content and collective memories of previous events/entities, they apply hashtags-based crawling and a bootstrapping technique.

The first step involves defining seed hashtags selected by experts such as #history and hashtags that refer to past events such as #historical events. All of the selected hashtags were selected by historian choices and then queried to collect Tweets. All the tags linked to the seed hashtags were then collected and reviewed manually to ensure that the hashtags were indeed related to history. Finally, 147 collected hashtags were attained along with 2370252 Tweets.

2.2 Feature Extraction

Term Frequency times Inverse Document Frequency (TF-IDF) is a standard measure of how dense the frequency of a specific term is in several documents. It is commonly used to re-weight feature count in order to reduce the effect of more frequent but less valuable words while increasing the influence of rarer but more significant terms [15]. [7] use TF IDF to obtain the weight for terms in the Tweets document for the data structure.

The bag-of-words model was used in NLP processing and information retrieval in the past. The grammar and sentence structure will be ignored, and a phrase can be perceived or represented as bags of words.

Text-term Matrix will then be used to look for instances of specific words in the collection of words to comprehend the document based on its most important terms. Tokenization in [2] uses BoW for feature extraction in NMF model. [1] also utilized BoW in preprocessing by constructing the model on W (tokenized terms of each document) and V (a corpus of terms with frequency).

2.3 Different Approaches in Topic Modeling

Rule-based is usually used for text classification. [11] describe ruled-based to any categorization technique that uses IF-THEN features for class prediction. For example, [14] utilized rule-based classification on Tweets to identify the topic in Tweets regarding HPV vaccination.

One of the popular machine learning topic models used to extract topics from a text corpus is Latent Dirichlet Allocation (LDA). According to [8], The 'latent' term refers to something hidden or not fully formed. It creates probabilities of topics and words categorized into documents.

A study by [9] in 2016 applied the LDA model. The LDA setting was experimented on to create a more useful and cohesive topic and transpired. It is also worth noting that all retrieved topics were linked to the event described in the data.

According to the subjective analysis of the findings, LDA can derive extensive insights from the data. It retrieved all of the important event elements, such as the participants, geographic location, and how the event. For example, in the Gikomba dataset, topic 0: 'bomb' and 'blast' give initial insights into the bombing.

2.4 Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN)

[12] present a supervised NTM (Topic Attention Model) with a Recurrent Neural Network (RNN). They demonstrate a unique approach to using document-specific (long text) topic modeling with a global topic vector obtained from NTM. In addition, they created a backpropagation inference that optimized combined models. TAM has improved learning tasks, such as regression and classification, and produced less ambiguity for document modeling.

RNN effectively retains the local patterns of a text string- both semantic and syntactic, but they struggle with long-range dependencies. [4] designed an RNN-based language model:- TopicRNN that uses latent topics to capture the semantic information of a text document.

There is two evaluation of performance for the methodology: word prediction and sentiment analysis. Penn TreeBank (PTB), a traditional benchmark for evaluating a new language model, was used for word prediction, while IMDB 100k was used for sentiment analysis. They also utilized RNN (TopicRNN), LSTM (TopicLSTM), and GRU(TopicGRU).

In the same study by [4], TopicRNN was trained with various network sizes. [1] utilized data from the USA and offer a five layers adaptive approach to follow the

trending topics of health on Twitter. The supervised approach is used as Convolutional Neural Network (CNN) to categorize and label the Tweets. Because the approach also used the word2vec model, it collects the correlation of words depending on the context, which improves the run time, prediction, and accuracy of Tweet tracking.

Tweepy, a python tool to collect Tweets, was used in the first layer. The cleaning and preparation procedures are detailed in the second layer, which changed Tweets into vectors. Word2vec was used in the third layer that turned the vector into a matrix used to initialize a neural network to predict the tagged Tweet. CNN classifier was then used in the fourth layer to label undetected Tweets from Word2Vec. Unlabeled data was then sent into the LDA model at the fifth layer, where new topics might be produced. The CNN model will be retrained for both the topics and Word2Vec.

2.5 Evaluation

There must be assessments of results for the study to be clear and logical. These tests were carried out to see how well the algorithms used in the research performed. The topic coherence analysis and the precision, recall, and F1 measurements evaluated the outcomes. It was also utilized to analyze the models to achieve accurate performance.

Train-test split is splitting the data into two parts, the first part is to sit or train the data using the algorithm, and the other part is to test and assess the accuracy of the data. Typical topic modeling such as LDA, LSA, and NMF did not rely on training data. But deep learning topic modeling approaches usually use train-test split for assessment reasons. For example, [4] also utilized a train-test split. The methodology, the TopicRNN, was initially tested on a word prediction task utilizing the Penn Treebank (PTB) component of the database. The Wall Street Journal is a newspaper published in the United States. The normal split is applied, with Sections. 0–20 (930K tokens) being used. Sections 21–22 (74K tokens) are for training, Sections. 23–24 (82K tokens) are for validation, and Sections. 23–24 (82K tokens) are for testing.

Calculating accuracy, precision, recall, and f-score are also vital. Precision was calculated to quantify the prediction of the class that belongs to the class, and recall quantifies the number of the class predictions made from all the datasets, fl-score was measured to provide a single score that stabilizes both the concerns precision and recall in one number. This type of evaluation will be useful if the algorithm used is the supervised algorithm.

In Nugruho et al. 2016, after the topic extraction and topic coherence evaluation, they evaluate the Tweet-topic accuracy by comparing the result of a manually labeled training dataset and measuring the precision (P) and recall (R) using F-score. The findings depicted that LDA outperforms NMF in terms of topic coherence.

[1] uses the confusion metrics of accuracy (1), recall (2), precision (3), and F1-score (4) to evaluate the results. In order to compare and contrast the suggested system, to predict and tag new Tweets, SVM and CNN algorithms were used. Unfortunately, both approaches have limitations due to the imbalanced datasets (Fig. 1).

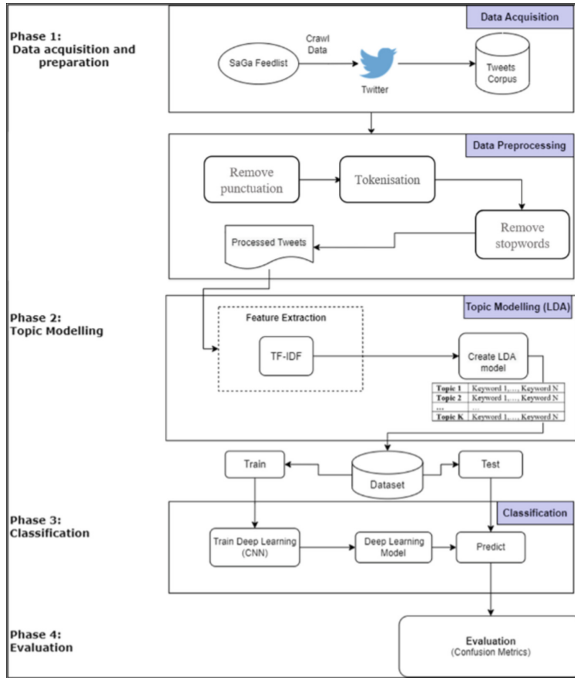


Fig. 1. Pipeline overview.

3 Methodology

This study will analyze the popular topics of the Sarawak Gazette from Twitter. It consists of 4 phases. Phase 1: Data acquisition and preparation, Phase 2: Topic Modelling, Phase 3: Classification, and finally, Phase 4: Evaluation.

The first phase is data acquisition and preparation. This phase is partitioned into two processes:- Data acquisition and data preprocessing. The study utilized a web crawler to extract the Tweets related to SaGa to create a short text corpus of Tweets, and then the acquired Tweets will be cleaned and normalized in the data preprocessing process. The Tweet acquired are dated from January 1, 2011, to December 31, 2021, and written in English.

Phase 2 is the deployment of the topic modeling to extort the relevant topic. Phase 2 is sectioned into two parts. The first part is feature extraction and the second part is, topic modeling with Latent Dirichlet Allocation (LDA). The study deployed TF-IDF as the feature extraction technique which merely entails running through the provided corpus once and computing document frequencies for all of its attributes. The LDA model was built with Gensim and the results were visualized by using pyLDAvis.

Phase 3 is the classification, where the Convolutional Neural Network (CNN) classifier will be used to predict the recurring topics from the Tweets. Finally, the fourth phase is where the confusion metric (accuracy, precision, recall, and f1-score) will be calculated and used as an evaluation method.

3.1 Data Acquisition and Data Pre-processing

An online free web crawler was utilized to extract the Tweets related to SaGa. Phantombuster is a cloud-based data extraction platform that scrapes important data from many social media websites to assist organizations in automating sales and advertising activities.

The Tweets will be collected by an online web crawler, Phantombuster, by feeding them with a list of relevant keywords from SaGa. The Tweets will be presented in a CSV file to ease the tasks after the tweet extraction.

The Tweets acquired are dated from January 1, 2011, to December 31, 2021. (10 years). Most of the Tweets obtained are in English because the study is gathering all Tweets from worldwide. This data collection activity generates unstructured Tweets that must be preprocessed. The first step is to crawl the selected keywords through Phantombuster. Tweets containing the queried keywords were fed into the crawler. [2] and [6] also utilized search using keywords in data collection method.

The study implemented keyword frequency analysis. The keywords are chosen from the e-Sarawak gazette websites, the Kuching in Pictures 1841–1946 by Ho Ah Chong and Sarawak Historical Event 1946–1960 by the same author. The three sources of keywords were named Document A, Document B, and Document C.

The study creates two types of search term sets: hashtags and strings. The search terms, which serve as the crawler's feed list, are provided in English and some Malay words. Table 1 displays the SaGa Feed list of keywords to crawl through Phantombuster. The hashtags were chosen as they depicted the most straightforward keywords; for example, #sarawakhistory is related closely to the historical context of Saga. In keyword frequency analysis, the keywords with a minor frequency were removed and replaced with new keywords with a better frequency. 51 keywords were finalized (Table 1).

Table 1. Keywords and frequencies

Keywords	Frequency in document A	Frequency in document B	Frequency in document C	Total
#sarawakhistory	0	0	0	0
#sejarahsarawak	0	0	0	0
Sarawak history	0	0	0	0
Sarawak gazette	9	12	3	24
Abang Haji Mustapha	0	2	2	4
Anthony Abell Sarawak	0	1	4	5
Anthony Brooke	2	2	2	6
anti cession Sarawak	0	4	13	17
Battle of North Borneo	0	0	0	0
Batu Lintang Camp	0	3	12	15

(continued)

Table 1. (continued)

Keywords	Frequency in document A	Frequency in document B	Frequency in document C	Total
British Borneo	3	0	2	5
Brooke's government	3	1	0	4
Brooketon colliery	6	0	0	6
Bujang Suntong	0	0	13	13
Charles Arden Clarke	0	1	7	8
Charles brooke	11	31	0	42
Charles Vyner Brooke	2	19	1	22
council negri Sarawak	18	43	33	94
Crown colony of Sarawak	0	0	6	6
Diocese of Labuan and Sarawak	0	0	0	0
Duncan Stewart sarawak	0	0	7	7
Dyaks	6	1	0	7
Fort Margherita	1	6	1	8
Ibans Sarawak	2	4	17	23
Japanese occupation Sarawak	0	8	5	13
Kuching Market Price List	126	0	0	126
liberation of kuching	0	0	0	0
Malay National Union	0	1	4	5
Malaya-Borneo	0	0	0	0
Parti Komunis Malaya	0	0	0	0
pergerakan pemuda melayu	0	0	3	3
Perjanjian Malaysia 1963	1	0	0	1
Perjanjian Persekutuan	0	0	0	0
pieces from the brooke past	1	0	0	1
Rajah of Sarawak	49	357	73	479
Ranee of Sarawak	4	28	2	34

(continued)

Table 1. (continued)

Keywords	Frequency in document A	Frequency in document B	Frequency in document C	Total
rosli dhoby	0	0	41	41
sarawak headhunter	0	0	0	0
Sarawak library	39	0	0	39
Sarawak Museum	14	10	0	24
Sarawak north borneo	8	9	1	18
Sarawak Ranger	1	11	4	16
Sarawak regatta	11	0	2	13
Sarawak Supreme Council	4	15	18	37
Sarawak Volunteer Force	4	5	1	10
Sumpitan	1	0	0	1
suruhanjaya cobbold	0	0	0	0
the cobbold comission	0	0	0	0
Tun Sir Henry H.S. Lee	0	0	0	0
white rajah	2	14	1	17
Prince Philip	2	15	0	17
upper Rejang	4	1	0	5
Cession Sarawak	38	125	101	264
Malaya Patriotic Fund	63	3	0	66
Rajah muda Sarawak	1	20	1	22
Charles Brooke Memorial	11	2	0	13
Sarawak turf club	63	3	0	66
Brooke Rule	0	11	5	16
Betram Brooke	0	12	1	13
Sarawak Union	7	0	1	8
Astana	7	51	22	80
Kayan	2	11	8	21
Rajah muda Sarawak	1	20	2	23

For the Phantombuster's setup, the first step is connecting to Twitter. Then the SaGa feedlist saved in the spreadsheet in google drive was used to scrape the Tweets. Phantombuster is non-case sensitive; thus, the inconsistency of each capital or lower letter would not affect the crawling process. The behavior is the limit of Tweets that the study aimed

to scrape. Meanwhile, the setting was set so that the process was done automatically and repeated eight times per day.

The Data Acquisition process has yielded 11435 Tweets saved in a CSV file named SaGaTweets. Taking the precedent research of [1], the study randomly checked the acquired Tweets and removed the unrelated and insignificant Tweets to get more meaningful data. Data preprocessing involves three processes which are removing punctuation, tokenizing, and removing stopwords.

3.2 Topic Modeling with Latent Dirichlet Allocation

The study utilized the feature extraction technique Term Frequency-Inverse Document Frequency (TF-IDF). In TF-IDF, feature extraction merely entails running through the provided corpus once and computing document frequencies for all of its attributes.

After TF-IDF, Latent Dirichlet Allocation (LDA) topic modeling was applied. Models.TfidfModel function was used to create a model with six topics, one of which is a collection of terms, each of which provides a particular amount of weight to the topic. In LDA, the study initiated six topics, each with ten generated terms.

After developing the LDA model, the study needs to analyze the generated topics and related keywords. There is no better tool than the interactive chart in the pyLDAvis package, designed to function effectively with Google Colaboratory. The output of the LDA model is displayed in pyLDAvis.

To discover the ideal number of topics, the study creates numerous LDA models with varying numbers of topics (k) and selects the one with the highest coherence value. Choosing a 'k' that represents the end of a fast increase in topic coherence typically yields meaningful and interpretable topics. The optimal number of topics with the highest coherence value is six. Six topics were chosen considering that the dataset the study used is small. This can be regarded as the optimal number of topics to summarise the corpus. [1, 7, 9] also find the optimal topic number with topic coherence score.

3.3 Classification

One utilization of topic modeling is determining the topic of a given text. To do so, the study looks for that document's subject number with the largest topic percentage contribution. To identify the main topic in each Tweet, the study defined a new function and compared the dominant topic and its percentage contribution.

Before the classification procedure began, the dataset was divided into 80% for training and 20% for testing. The study has an 80:20 data splitting ratio on 4301 data, the train size would be 3440, and the test size would be 861. The first step in CNN to search for specific characteristics or patterns in the Tweets is to employ a pre-defined word embedding from the library. There are several open-source embeddings available; the study has utilized Word2Vec.

Then using the Keras model, the study inserted the input layers linked to convolutional layers, which conduct several jobs such as padding, striding, kernel operation, etc. This layer is regarded as a building block of convolutional neural networks. The filter/kernel is then slid over these embeddings to discover convolutions, which are then

dimensionally lowered using the Max Pooling layer to reduce complexity and processing. Finally, there are entirely linked layers and an activation function on the outputs that will provide values for each class. The dense was set to 2 as the study has two topics.

3.4 Evaluation

The evaluation for the study utilized the sparse categorical cross-entropy (scc) function to generate a category index of the most probable matched category. There are three convolutional layers with kernel sizes 32, 64, and 128. The model is trained for 30 epochs or passes over 300 batch records of the training data. The study also employed the efficient Adam-Optimizer implementation to calculate the weights of stochastic gradient descent and measured accuracy and loss during training.

Then, the classification report is started with a confusion matrix, thus exploiting the function from `sklearn.metrics` and importing `confusion_matrix` and `classification_report`. The Confusion Matrix has two parameters: test and projected data values. The results display the precision, recall, f1-score, and support for all four topics.

4 Results

The results of the study can be divided into three analyses. The first one is the data analysis which involves the Tweets acquired during Phase 1. The study collected 11435 Tweets using 51 keywords and chose 4301 total Tweets for the study to be saved as the raw data. After going through a preprocessing phase, the Tweets are safely conserved in a CSV file and saved as a short text corpus.

Table 2. The topics and the keywords

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
rosli	british	record	charles	regatta	rosli
dhobi	makam	retrieve	white	library	dhoby
sarawak	masjid	kayan	kuching	fort	sarawak
duncan	pahlawan	sarawak	rajah	margherita	melayu
sibu	borneo	kilau	brooke	state	pergerakan
stewart	kayan	rentap	sarawak	colony	union
rukun	sarawak	trial	british	drama	malaya
gabenor	surat	cinta	rajahs	sarawak	borneo
gantung	sabah	kenyah	vyner	kuching	pemuda

The second analysis is the topic modeling results. After running the LDA model in implementation, the study verified that the LDA findings could determine the event and, in some cases, important figures and give some important information about the event

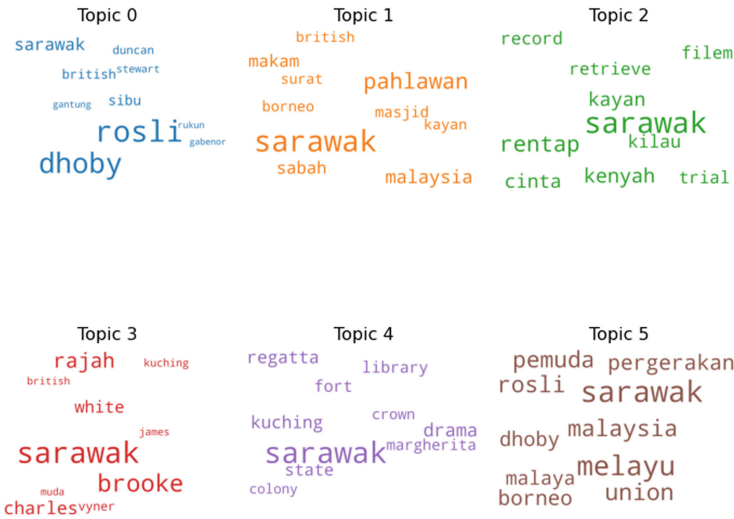


Fig. 2. The wordcloud visualization of the Topic and its keywords.

through manual examination of the obtained topics. The topics and the keywords were listed in Table 2 (Fig. 2).

The third analysis is the result analysis. The study used Convolutional Neural Network (CNN) as a model for classification. Four metrics of the confusion matrix are used to assess classification results: accuracy, recall, precision, and F1-score. The study found that the accuracy of the model is considerably low at 68% because of the inconsistencies that happened when two languages were mixed.

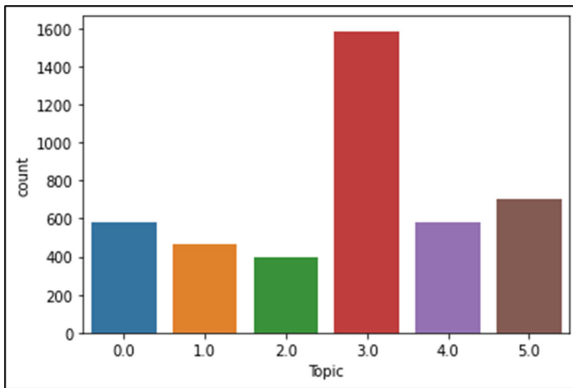


Fig. 3. Topics and frequencies.

Using the LDA model, the study also automates the model to label the dataset with its dominant topic Fig. 3. Displays the count plot for the distribution of the topic. Tweets labeled with Topic 3 have the highest distributed topic compared to other topics.

The model for classification used is Convolutional Neural Network (CNN). Four metrics are used to assess classification results: accuracy, recall, precision, and F1-score. Precision reflects the classifier's preciseness and refers to the number of linked predicted topics. Recall indicates how many accurate records are predicted, and F-scores define the harmonic average of precision and recall (Table 3).

Table 3. Four metrics to assess classification results

Topic	Precision	Recall	F1-score
0	0.71	0.53	0.61
1	0.38	0.41	0.40
2	0.51	0.39	0.44
3	0.78	0.90	0.84
4	0.64	0.49	0.55
5	0.67	0.74	0.70
Accuracy			0.68
Macro average	0.61	0.58	0.59
Weighted average	0.67	0.68	0.67

The dataset is imbalanced since 330 of the 861 instances are from Topic 3. (38% of the dataset). As a result, the predictor generally always classifies every given sample as belonging to Topic 3, achieving high precision and recall and F1-score for Topic 3 and lower scores for other topics.

5 Discussion

Overall, the study aims to understand more about how people use their social media to share the content of the Sarawak Gazette (SaGa), one of the valuable historical documents of Sarawak. By the completion of the study, three proposed objectives had been met, and the contributions of the study had also been addressed.

A list of tweets was crawled, varying from 2011 to 2021. Using 51 specific keywords, the study yielded considerably large Tweets amount from a variety of Twitter users worldwide. After going through a preprocessing phase, the Tweets are safely conserved in a CSV file and saved as a short text corpus. This corpus may be useful for future references. These Tweets offer public thoughts on Sarawak Gazette and assist us to discover how people perceive certain topics. After a manual observation, most of the Tweets related to SaGa were in picture format. Instead of writing in text, people chose to snip a part of the gazette and did not describe it more. Most Tweets also use a direct link to external websites on SaGa.

Latent Dirichlet Allocation (LDA) was used as a topic modeling approach. After running the model, the study specified six topics, each with ten generated terms. The

study can infer that Topic 0 relates to the event of Duncan Stewart's assassination as the keywords contain terms such as 'duncan', 'stewart', 'sibu', 'rosli', and 'dhobi'. Topic 1 and Topic 2 is quite unclear, but some keywords suggest that they might be on funeral and important historical figures. Topic 3 can be associated with Brooke's rule in Sarawak as it has terms such as 'sarawak', 'brooke', and 'rajah'. Additionally, the study deduced that Topic 4 mostly related to places in Sarawak.

The study demonstrated that the LDA findings could determine the event, and in some instances, significant figures mentioned in SaGa, and provide some relevant information about the event by manually analyzing the retrieved topics. Furthermore, Convolutional Neural Network (CNN) was used as the classifier to evaluate the model.

The classification method was further evaluated with a confusion matrix. The confusion matrix is utilized to calculate the accuracy of the deep learning algorithm in categorizing the data into its corresponding topics. The result shows that the model has a low accuracy result due to the mix of English and Malay languages.

One of the limitations is inadequate language resources. Because the study involves a mix of English and Malay Tweets, which include both official and informal word usage, it is difficult to find relevant work and appropriate dictionaries to utilize in the study. For example, adequate Malay stopword libraries and Malay embedding. Another limitation involves the insufficiency of SaGa resources online. The text version of SaGa cannot be found online, and the official websites only show the scanned version, with a few missing issue numbers.

6 Conclusion

This study has used Sarawak Gazette as a medium for historical documents to understand more about historical document-oriented content on Twitter. On the final note, three objectives were met by the end of the study. The first one is building a short-text corpus of Sarawak Gazette from Twitter. Next, predicting the topic used in Twitter related to Sarawak Gazette using deep learning goals has also been met. Third, the study evaluated the predicted topic's accuracy using standard measure metrics. Although all of the study's recommended methodologies produced considerable results, the experiment has its limitations, including the restricted capabilities of the model as there were inadequate Malay language library resources and SaGa resources.

The objective of enhanced text conceptualization can be reached in the future by improving automation. Topic modeling can increase the likelihood of uncovering latent topical patterns across the collection. In the future, the study may also use another topic modeling, such as Latent Semantic Analysis and Non-Negative Matrix Factorisation (NNMF), and compare the results to LDAs'. This would assist in enhancing the quality of topic modeling since they allow us to assess alternative models from both an overall and subject-specific perspective. Furthermore, instead of only using topic coherence to find the optimal topic number, it is feasible for the study to utilize NMI and topic coherence combined as done by [7]. Another thing that can be considered is the new search method using image searching, as most of the Tweets related to SaGa were in picture format.

Acknowledgements. The authors would like to express their heartiest thanks to Universiti Malaysia Sarawak (UNIMAS) and the Faculty of Computer Science and Information Technology (FCSIT), UNIMAS for the research opportunity and financial assistance.

The authors also extend heartfelt thanks to the family of Nur Ain. The authors were extremely grateful to the parents, Mr. Nor Azizan Bin Abdul Rashid, and Mrs. Che Wan Binti Che Man, for their love, prayers, caring, sacrifices, and understanding during the preparation of the study. Last but not least, special thanks to the author's friends for their continuous motivation and companionship throughout this research.

References

1. Asghari, M., Sierra-Sosa, D., Elmaghraby, A.: Trends on health in social media: analysis using twitter topic modeling. In: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 558–563. Louisville, KY, USA (2018). <https://doi.org/10.1109/ISSPIT.2018.8642679>
2. Casalino, G., Castiello, C., Del Buono, N., Mencar, C.: Intelligent twitter data analysis based on nonnegative matrix factorizations. In: Gervasi, O., et al. (eds.) ICCSA 2017. LNCS, vol. 10404, pp. 188–202. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62392-4_14
3. Congosto, M.: Digital sources: a case study of the analysis of the Recovery of Historical Memory in Spain on the social network Twitter. *Cult. Hist. Dig. J.* **7**, 015 (2019). <https://doi.org/10.3989/chdj.2018.015>
4. Dieng, A.B., Wang, C., Gao, J., Paisley, J.: TopicRNN: A recurrent neural network with long-range semantic dependency. In: Proceedings of ICLR (2017)
5. Komorowski, M., Huu, T., Deligiannis, N.: Twitter data analysis for studying communities of practice in the media industry. *Telematics Inform.* **35**, 195–212 (2018). <https://doi.org/10.1016/j.tele.2017.11.001>
6. Rasidi, N.M., Bakar, S.A., Razak, F.A.: Tweets clustering using latent semantic analysis. In: AIP (American Institute of Physics) Conference Proceedings, vol. 1830, no. 1, p. 020060. AIP Publishing LLC (2017)
7. Qomariyah, S., Iriawan, N., Fithriasari, K.: Topic modeling twitter data using latent dirichlet allocation and latent semantic analysis. In: AIP Conference Proceedings, vol. 2194, no. 1, p. 020093. AIP Publishing LLC (2019)
8. Seth, N.: Topic Modeling and Latent Dirichlet Allocation (LDA) using Gensim. *Analytics Vidhya* (2021)
9. Sokolova, M., et al.: Topic Modelling and Event Identification from Twitter Textual Data (2016)
10. Sumikawa, Y., Jatowt, A.: Analyzing history-related posts in twitter. *Int. J. Digit. Libr.* **22**(1), 105–134 (2020). <https://doi.org/10.1007/s00799-020-00296-2>
11. Tung, A.K.H.: Rule-based classification. In: Liu, L., Tamer Özsu, M. (eds.) *Encyclopedia of Database Systems*, pp. 2459–2462. Springer US, Boston, MA (2009). https://doi.org/10.1007/978-0-387-39940-9_559
12. Wang, X., Yang, Y.: Neural topic model with attention for supervised learning. In: *International Conference on Artificial Intelligence and Statistics*, pp. 1147–1156. PMLR (2020)
13. Yu, M., Huang, Q., Qin, H., Scheele, C., Yang, C.: Deep learning for real-time social media text classification for situation awareness – using Hurricanes Sandy, Harvey, and Irma as case studies. *Int. J. Dig. Earth* **12**, 1–18 (2019). <https://doi.org/10.1080/17538947.2019.1574316>

14. Zhang, H., et al.: Mining Twitter to assess the determinants of health behavior toward human papillomavirus vaccination in the United States. *J. Am. Med. Inform. Assoc.* **27**(2), 225–235 (2020)
15. Zhao, G., Liu, Y., Zhang, W., Wang, Y.: TFIDF based Feature Words Extraction and Topic Modeling for Short Text, pp. 188–191 (2018). <https://doi.org/10.1145/3180374.3181354>

Author Index

A

Ab Patar, Mohd Nor Azmi 107
Abdul-Rahman, Shuzlina 107
Affendy, Nur-Adibah Raihan 147
Ahmat Ruslan, Fazlina 107
Ahmed, Md Manjur 61
Akbar, Muhammad Sjahid 275
Arora, Mahima 3
Ayub, Muhammad Azmi 107
Aziz, Normaziah Abdul 92

B

Basir, Muhammad Sufyan Safwan Mohamad 147

C

Chintyana, Alissa 163
Chiun, Ong Yee 246
Choiruddin, Achmad 163
Chopra, Ankush 3

G

Gobithaasan, R. U. 261

H

Hai, Tao 18
Hasan, Z. A. 261
Husen, Mohd Nizam 61
Hussin, Amir Aatieff Amir 92

I

Ismail, Amelia Ritahani 92
Ismail, Muhammad Hazim 291
Ismail, Rizalafande Che 147

K

Khalid, Nur Hidayah Mohd 92

L

Lei, Yang 46
Liu, Qingchuan 107

M

Marvi, Jahroo Nabila 211
Muda, Muhammad Adlansyah 275

N

Noh, Norzaidah Md 193
Nor Azizan, Nur Ain Binti 178

P

Prastyo, Dedy Dwi 275

R

Rahadiani, Laksmita 211
Ramli, Intan Syaherra 222
Ramli, Muhammad Shafie Mohd 136
Ramzi, Nurul Izleen 193
Roslee, Hasan Basri 291
Ruhaiyem, Nur Intan Raihana 232, 246
Rusli, Nur Ida Aniza 222

S

Saeed, Suhaila Binti 178
Said, Mohamad Azizan Mohamad 147
Sapari, Norazliani Md. 136
Sapri, Nik Nur Fatin Fatimah 291
Say, Soh Hong 232
Sutikno 163

T

Tao, Hai 61

W

Wan Yaacob, Wan Fairos 291
Wang, Dan 76

Wang, Ting 76
Wang, Xiaolan 61

X

Xue, Hongyu 32

Y

Yusof, Khairul Huda 136, 147
Yusof, Muhammad Abdullah Bin 178

Yusoff, Marina 193
Yusup, Yusri 232

Z

Zamani, Nur Shazwany 136
Zhang, Fuhao 18
Zhao, Qian 61
Zheng, Maoxing 121
Zhou, Jincheng 32, 46, 121
Zulkifle, Farizuwana Akma 222