

The copyright © of this thesis belongs to its rightful author and/or other copyright owner. Copies can be accessed and downloaded for non-commercial or learning purposes without any charge and permission. The thesis cannot be reproduced or quoted as a whole without the permission from its rightful owner. No alteration or changes in format is allowed without permission from its rightful owner.



**AN ENHANCED BINARY BAT AND MARKOV CLUSTERING
ALGORITHMS TO IMPROVE EVENT DETECTION FOR
HETEROGENEOUS NEWS TEXT DOCUMENTS**



WAFI ZUBAIR ABDULLAH AL-DYANI

**DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2022**



Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(*Certification of thesis / dissertation*)

Kami, yang bertandatangan, memperakukan bahawa
(*We, the undersigned, certify that*)

WAFA ZUBAIR ABDULLAH

calon untuk Ijazah **PhD**
(*candidate for the degree of*)

telah mengemukakan tesis / disertasi yang bertajuk:
(*has presented his/her thesis / dissertation of the following title*):

**"AN ENHANCED BINARY BAT AND MARCOV CLUSTERING ALGORITHMS TO IMPROVE EVENT
DETECTION FOR HETEROGENEOUS NEWS TEXT DOCUMENTS"**

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(*as it appears on the title page and front cover of the thesis / dissertation*).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : **25 April 2022.**

That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:
25 April 2022.

Pengerusi Viva:
(*Chairman for VIVA*)

Assoc. Prof. Dr. Nur Haryani Zakaria

Tandatangan
(*Signature*)

Pemeriksa Luar:
(*External Examiner*)

Prof. Ts. Dr. Mokhairi Mokhtar

Tandatangan
(*Signature*)

Pemeriksa Dalam:
(*Internal Examiner*)

Assoc. Prof. Dr. Yuhanis Yusof

Tandatangan
(*Signature*)

Nama Penyelia/Penyelia-penyelia:
(*Name of Supervisor/Supervisors*)

Dr. Farzana Kabir Ahmad

Tandatangan
(*Signature*)

Nama Penyelia/Penyelia-penyelia:
(*Name of Supervisor/Supervisors*)

Assoc. Prof. Dr. Siti Sakira Kamaruddin


Tandatangan
(*Signature*)

Tarikh:
(*Date*) **25 April 2022**

Permission to Use

I am presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:



Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Pengesanan Peristiwa (ED) bertindak untuk mengenal pasti peristiwa dari pelbagai jenis teks. Membina model ED untuk dokumen teks berita sangat membantu pembuat keputusan dalam pelbagai disiplin dalam meningkatkan strategi mereka. Walau bagaimanapun, mengenal pasti dan meringkaskan peristiwa daripada data tersebut adalah tugas yang tidak mudah kerana jumlah besar dokumen teks berita heterogen yang diterbitkan. Dokumen sedemikian mewujudkan ruang fitur berdimensi tinggi yang mempengaruhi kaedah dasar dalam model ED. Untuk menangani masalah sedemikian, penyelidikan ini memperkenalkan model ED yang dipertingkatkan yang merangkumi kaedah yang ditambahbaik untuk fasa paling penting model ED seperti Pemilihan Fitur (FS), ED dan ringkasan. Penyelidikan ini berfokuskan kepada masalah FS dengan mengesan peristiwa secara automatik melalui kaedah FS *wrapper* baharu berdasarkan Algoritma Kelawar Binari Tersuai (ABBA) dan Algoritma Pengelompokan Markov Tersuai (AMCL), yang dinamakan ABBA-AMCL. Teknik penyesuaian ini dibangunkan untuk mengatasi penumpuan pramatang dalam BBA dan kadar penumpuan cepat dalam MCL. Tambahan pula, penyelidikan ini mencadangkan empat kaedah peringkasan untuk menghasilkan ringkasan yang berinformasi. Model ED yang dipertingkat diuji pada 10 set data penanda aras dan 2 set data berita Facebook. Keberkesanan ABBA-AMCL dibandingkan dengan 8 kaedah FS berdasarkan algoritma meta-heuristik dan 6 kaedah ED berasaskan graf. Keputusan empirikal dan statistik membuktikan bahawa ABBA-AMCL mengatasi kaedah lain pada kebanyakan set data. Ciri perwakilan utama menunjukkan bahawa kaedah ABBA-AMCL berjaya mengesan peristiwa dunia sebenar daripada set data berita Facebook dengan 0.96 *Precision* dan 1 *Recall* untuk dataset 11, manakala untuk set data 12, *Precision* ialah 1 dan *Recall* ialah 0.76. Sebagai kesimpulan, ABBA-AMCL baharu yang ditunjukkan dalam penyelidikan ini telah berjaya merapatkan jurang penyelidikan dan menyelesaikan permasalahan ruang fitur berdimensi tinggi. Oleh itu, model ED yang dipertingkatkan boleh menyusun dokumen berita mengikut peristiwa yang berbeza dan dapat menyediakan informasi bermanfaat kepada pembuat dasar dalam membuat keputusan.

Kata Kunci: Pengesanan peristiwa, Pemilihan Fitur, Dokumen teks berita heterogen, Algoritma Kelawar Binari, Algoritma Pengelompokan Markov.

Abstract

Event Detection (ED) works on identifying events from various types of data. Building an ED model for news text documents greatly helps decision-makers in various disciplines in improving their strategies. However, identifying and summarizing events from such data is a non-trivial task due to the large volume of published heterogeneous news text documents. Such documents create a high-dimensional feature space that influences the overall performance of the baseline methods in ED model. To address such a problem, this research presents an enhanced ED model that includes improved methods for the crucial phases of the ED model such as Feature Selection (FS), ED, and summarization. This work focuses on the FS problem by automatically detecting events through a novel wrapper FS method based on Adapted Binary Bat Algorithm (ABBA) and Adapted Markov Clustering Algorithm (AMCL), termed ABBA-AMCL. These adaptive techniques were developed to overcome the premature convergence in BBA and fast convergence rate in MCL. Furthermore, this study proposes four summarizing methods to generate informative summaries. The enhanced ED model was tested on 10 benchmark datasets and 2 Facebook news datasets. The effectiveness of ABBA-AMCL was compared to 8 FS methods based on meta-heuristic algorithms and 6 graph-based ED methods. The empirical and statistical results proved that ABBA-AMCL surpassed other methods on most datasets. The key representative features demonstrated that ABBA-AMCL method successfully detects real-world events from Facebook news datasets with 0.96 Precision and 1 Recall for dataset 11, while for dataset 12, the Precision is 1 and Recall is 0.76. To conclude, the novel ABBA-AMCL presented in this research has successfully bridged the research gap and resolved the curse of high dimensionality feature space for heterogeneous news text documents. Hence, the enhanced ED model can organize news documents into distinct events and provide policymakers with valuable information for decision making.

Keywords: Event detection, Feature selection, Heterogeneous news text documents, Binary bat algorithm, Markov clustering algorithm.

Acknowledgement

All praise is due to Allah, who, by His grace and blessings, I have completed my thesis. I would like to express my appreciation and gratitude to everyone who has contributed to complete this thesis. I would like to thank Dr. Farzana Kabir Ahmad's for her valuable support, guidance, and feedback that helped me achieve my goal. I would like to express my thanks to my co-supervisor, Prof. Madya Dr. Siti Sakira Binti Kamaruddin, for her guidance, comments, and kindness, which helped me to improve my work.

I would like to express my heart-felt gratitude to my family, especially to the most important person in my life, my beloved mother, Zakia Mohammed Al-Dyani, who has been a mother and a father to me throughout my life and without her prayers, I might not be able to achieve what I have achieved until now. I would like to dedicate this research to my deceased father, Zubair Abdullah Al-Dyani for whom I hold all the appreciation, respect and pride. May Allah (SWT), forgive him and have mercy on him and make his abode Jannat al-Firdaws. Also, I would like to thank my beloved and supportive sister, Dr. Iman Zubair as well as my beloved brothers, sisters-in-law, my lovely nieces, and nephews for their constant source of love, concern, support, and strength through all these years.

I would like to thank all of my friends for their encouragement and support during my PhD journey. I greatly value their friendship and I deeply appreciate their belief in me. I am also very grateful to the examiners for their valuable comments during the viva and corrections period.

I would like to express my appreciation to Hadhramout University and Hadhramout Foundation for Human Development for giving me the opportunity (scholarship) to study in Malaysia.

Finally, I had a very enjoyable study at Universiti Utara Malaysia (UUM). Not only does it have a beautiful natural environment, but the university also has helpful and kind staff.

Table of Contents

PERMISSION TO USE	I
ABSTRAK	II
ABSTRACT	III
ACKNOWLEDGEMENT	IV
TABLE OF CONTENTS	V
LIST OF TABLES	XI
LIST OF FIGURES	XIII
LIST OF APPENDICES	XIV
LIST OF ABBREVIATIONS.....	XV
CHAPTER ONE INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	8
1.3 Research Questions	16
1.4 Research Objectives.....	17
1.5 Scope of Study	17
1.6 Significant of Study	19
1.7 Proposal Organization.....	20
CHAPTER TWO LITERATURE REVIEW	23
2.1 Introduction.....	23
2.2 Event Detection Definitions and Concepts	23
2.3 Event Detection Models.....	25
2.3.1 Event Detection Models for News Text Documents	30
2.3.2 Event Detection Models for Facebook News Posts.....	32
2.3.3 Variations between News Articles and Facebook News Posts.....	34
2.4 Limitations and Motivation.....	36
2.5 Feature Selection Phase	39
2.5.1 Feature Selection Methods.....	41
2.5.2 Feature Selection Methods Based on Meta-Heuristic Algorithms	47
2.5.3 Bat Algorithm	50
2.5.3.1 Binary Bat Algorithm.....	55

2.5.3.1.1	Key Advantages of Binary Bat Algorithm	55
2.5.3.1.2	Key Disadvantages of Binary Bat Algorithm.....	56
2.5.3.2	Related Work: Binary Bat Algorithm for Feature Selection Problem	58
2.5.3.3	Related Works: Limitations of Binary Bat Algorithm	61
2.5.3.4	Tuning and Controlling Techniques for Binary Bat Algorithm	65
2.6	Event Detection Phase	70
2.6.1	Event Detection Methods.....	70
2.6.1.1	Query-Based Methods.....	73
2.6.1.2	Statistical-Based Methods	73
2.6.1.3	Probabilistic\Topical Based Methods.....	74
2.6.1.4	Clustering-Based Methods	76
2.6.1.5	Graph-Based Methods.....	79
2.6.2	Markov Clustering Method.....	88
2.6.2.1	Key Advantages of Markov Clustering Method	89
2.6.2.2	Key Disadvantages of Markov Clustering Method.....	90
2.6.2.3	Parameter Setting Techniques for Markov Clustering Method.....	92
2.7	Summarization Phase.....	94
2.7.1	Summarizing Methods.....	94
2.7.2	Related Works: Summarization Methods	97
2.7.3	Limitations of Related Works: Summarization Methods	99
2.7.4	LUHN Summarization Technique	103
2.7.5	Text Rank Summarization Technique	104
2.8	Discussion	106
2.9	Chapter summary	109
CHAPTER THREE RESEARCH METHODOLOGY.....		111
3.1	Introduction.....	111
3.2	Data Collection Phase	113
3.2.1	Facebook News Posts	113
3.2.1.1	Collection of Facebook News Posts	113
3.2.1.2	Labelling Facebook News Posts.....	116

3.2.2	20Newsgroup	120
3.2.3	News Aggregator Dataset	121
3.2.4	Benchmark Datasets: News articles and Really Simple Syndication News Feeds	122
3.2.5	Dataset Preparation	125
3.3	Preprocessing Phase	127
3.3.1	Filtering Step	128
3.3.2	Remove URL, Digits, Extra White Space, and Special Characters Step	129
3.3.3	Converting to Lowercase Text Step	129
3.3.4	Tokenization Step	129
3.3.5	Remove Stop Words Step	130
3.3.6	Text Normalization Step	131
3.3.7	Document Representation Step	132
3.4	Feature Selection Phase	133
3.5	Event Detection Phase	138
3.6	Summarization Phase	140
3.7	Evaluation Phase	141
3.8	Chapter Summary	146
CHAPTER FOUR WRAPPER FEATURE SELECTION METHOD BASED ON BASIC BINARY BAT AND BASIC MARKOV CLUSTERING ALGORITHMS		147
4.1	Introduction	147
4.2	Developed Wrapper Feature Selection Method Based on Basic Binary Bat and Basic Markov Clustering Algorithms	148
4.2.1	Feature Selection Phase	148
4.2.2	Event Detection Phase	149
4.2.2.1	Graph Construction Process	152
4.2.2.2	Graph clustering: Detection of Event Clusters	152
4.2.3	Evaluation Phase	153
4.3	Parameter Settings	154
4.4	Experimental Results	156

4.4.1 Evaluation Metrics	156
4.4.2 Convergence Rate	159
4.4.3 Statistical Results	163
4.5 Discussion	164
4.6 Chapter Summary	167
 CHAPTER FIVE WRAPPER FEATURE SELECTION METHOD BASED ON ADAPTIVE BINARY BAT AND BASIC MARKOV CLUSTERING ALGORITHMS..... 168	
5.1 Introduction.....	168
5.2 Developed Wrapper Feature Selection Method Based on Adaptive Binary Bat Algorithm and Basic Markov Clustering Method	169
5.2.1 Feature Selection Phase	169
5.2.1.1 Update Velocity Equation	170
5.2.1.2 Accept New Generated Solution Condition	172
5.2.1.3 Developed Adaptive Techniques for Updating A and r Equations.....	173
5.2.2 Event Detection Phase	176
5.2.3 Evaluations Phase	176
5.3 Parameter Settings	177
5.4 Experimental Results	178
5.4.1 Evaluation Metrics	178
5.4.2 Convergence Rate	181
5.4.3 Statistical Results.....	185
5.5 Discussion	186
5.6 Chapter Summary	188
 CHAPTER SIX WRAPPER FEATURE SELECTION METHOD BASED ON ADAPTIVE BINARY BAT AND ADAPTIVE MARKOV CLUSTERING ALGORITHMS..... 190	
6.1 Introduction.....	190
6.2 Developed Wrapper Feature Selection Method Based on Adaptive Binary Bat and Adaptive Markov Clustering Algorithms	191
6.2.1 Feature Selection Phase	191

6.2.2 Event Detection Phase	192
6.2.2.1 Adapting Pruning (<i>p</i>) Parameter.....	192
6.2.2.2 Adapting Inflation (<i>inf</i>) Parameter	194
6.2.3 Evaluations Phase	196
6.3 Parameter Settings	197
6.4 Experimental Results and Discussions	197
6.4.1 Evaluation Metrics	197
6.4.1.1 ABBA-AMCL vs MHAs-Based Methods	198
6.4.1.2 ABBA-AMCL vs Graph-Based ED Methods.....	201
6.4.2 Statistical Results.....	203
6.4.2.1 ABBA-AMCL vs MHAs-Based Methods	203
6.4.2.2 ABBA-AMCL vs Graph ED Methods	205
6.4.3 Visualize Event Clusters.....	207
6.5 Chapter Summary	211
CHAPTER SEVEN SUMMARIZATON AND REPRESENTATION OF	
EVENTS.....	213
7.1 Introduction.....	213
7.2 Developed Summarization Methods.....	214
7.2.1 Hybrid TextRank-LUHN Summarization Method.....	214
7.2.1.1 Summary by Text Rank Technique.....	215
7.2.1.2 Summary by LUHN Technique	216
7.2.1.3 Merging Summaries	217
7.2.2 Voting Summarization Techniques	218
7.2.2.1 Comment Voting Summarization Technique.....	218
7.2.2.2 Share Voting Summarization Technique	218
7.2.2.3 Engagement Voting Summarization Technique.....	218
7.3 Event Cluster Representation.....	219
7.4 Evaluation Metrics	219
7.5 Parameter Settings	223
7.6 Results and Discussion	224
7.6.1 Summary Evaluation Results for the First Experiment	224

7.6.2 Summary Evaluation Results for the Second Experiment.....	227
7.6.3 Summary Evaluation Results for the Third Experiment.....	229
7.6.4 Representation of Events	231
7.7 Chapter Summary	250
CHAPTER EIGHT CONCLUSIONS AND FUTURE WORK	251
8.1 Conclusions.....	251
8.2 Research Objectives and Contributions	252
8.3 Limitation of the Study	256
8.4 Recommendation for Future Work	258
REFERENCES.....	260
APPENDIX A List of Publications.....	299



UUM
Universiti Utara Malaysia

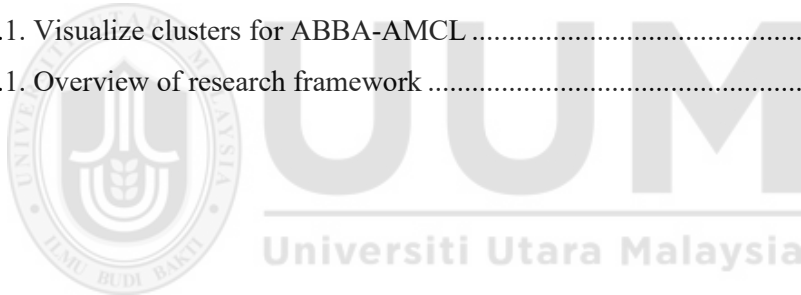
List of Tables

Table 2.1 Summary of ED Studies for Text Data	26
Table 2.2 Comparison of Official News Articles and Facebook News Posts	35
Table 2.3 Comparison of Feature Reduction Methods	42
Table 2.4 Comparison of Feature Selection Methods	46
Table 2.5 Advantages and Disadvantages of BBA	57
Table 2.6 Summary of BBA Related Works	63
Table 2.7 Limitations of Graph-Based Methods	87
Table 2.8 Limitations of Parameter Settings Techniques for MCL	93
Table 2.9 Summarization Methods used by ED Studies	97
Table 3.1 Description of Facebook News Posts Metadata	115
Table 3.2 Statistics Analysis of Facebook News Posts (January 2010 to May 2020)	115
Table 3.3 Extracted Events from Facebook News Posts (2010 to 2014)	119
Table 3.4 Extracted Events from Facebook News Posts (2015 to 2020)	120
Table 3.5 20Newsgroup Categories	121
Table 3.6 Categories of News Aggregator Dataset	121
Table 3.7 News Articles and RSS News Feeds	123
Table 3.8 Categories of News Articles and RSS News Feeds	124
Table 3.9 Characteristics of Text News Datasets	126
Table 4.1 Initial Parameters Setting for BBA, GA, BPSO, and MCL Algorithms	155
Table 4.2 Performance of FS Methods Based on F_{avg}	156
Table 4.3 Performance of FS Methods Based on P_{avg}	157
Table 4.4 Performance of FS Methods Based on R_{avg}	157
Table 4.5 Performance of FS Methods Based on SFR	157
Table 4.6 Results of Friedman Rank Test Based on F_{avg}	163
Table 4.7 Results of Wilcoxon Signed-Rank Test Based on F_{avg}	164
Table 5.1 Initial Parameters Setting for BCS, BGSA, BDFFA, and DIWBBA Algorithms ...	177
Table 5.2 Performance of FS Methods Based on F_{avg}	179
Table 5.3 Performance of FS Methods Based on P_{avg}	179
Table 5.4 Performance of FS Methods Based on R_{avg}	179
Table 5.5 Performance of FS Methods Based on SFR	180
Table 5.6 Results of Friedman Rank Test Based on F_{avg}	185
Table 5.7 Results of Wilcoxon Signed-Rank Test Based on F_{avg}	186
Table 6.1 Performance of Methods Based on F_{avg}	198

Table 6.2 Performance of Methods Based on P_{avg}	198
Table 6.3 Performance of Methods Based on R_{avg}	199
Table 6.4 Performance of Methods Based on RPD_{avg}	199
Table 6.5 Performance of Methods Based on Best F Measure.....	202
Table 6.6 Performance of Methods Based on RPD for Best F Measure.....	202
Table 6.7 Friedman Rank Test Based on F_{avg}	204
Table 6.8 Wilcoxon Signed-Rank Test Based on F_{avg}	205
Table 6.9 Friedman Rank Test Based on Best F Measure.....	206
Table 6.10 Wilcoxon Signed-Rank Test Based on Best F Measure.....	206
Table 7.1 Performance of Summarization Methods Based on FROUGE-1.....	225
Table 7.2 Performance of Summarization Methods Based on FROUGE-2.....	225
Table 7.3 Performance of Summarization Methods Based on FROUGE-3.....	225
Table 7.4 Performance of Summarization Methods Based on FROUGE-1.....	227
Table 7.5 Performance of Summarization Methods Based on FROUGE-2.....	227
Table 7.6 Performance of Summarization Methods Based on FROUGE-3.....	228
Table 7.7 Results of all Applied Methods using TR-LH with TFIDF for DS11.....	230
Table 7.8 Results of all Applied Methods using TR-LH with TFIDF for DS12.....	230
Table 7.9 Japan Tsunami Event Features.....	232
Table 7.10 Trapped of Chilean Miners Event Features.....	233
Table 7.11 Sinking of the South Korean Ferry Event Features.....	234
Table 7.12 Malaysia Airlines Flight MH370 Lost Event Features.....	235
Table 7.13 Jamal Khashoggi Murder Event Features.....	236
Table 7.14 Kenya's Capital Nairobi Attack Event Features.....	237
Table 7.15 Iran Nuclear Deal Event Features.....	238
Table 7.16 Rohingya Crisis Event Features.....	238
Table 7.17 Features and Descriptions of Events for DS6.....	240
Table 7.18 Features and Descriptions of Events for DS8.....	241
Table 7.19 Features and Descriptions of Events for DS9.....	243
Table 7.20 Features and Descriptions of Events for DS10.....	244
Table 7.21 Features and Descriptions of Events for DS11.....	246
Table 7.22 Features and Descriptions of Events for DS12.....	248

List of Figures

Figure 2.1. Main phases of ED model	39
Figure 2.2. Taxonomy of feature reduction methods.....	43
Figure 2.3. Parameter setting taxonomy according to Parpinelli et al. (2019).....	65
Figure 2.4. Classification of ED methods.....	71
Figure 2.5. Graph based methods	79
Figure 2.6. Taxonomy of summarization methods	95
Figure 3.1. Research methodology	112
Figure 3.2. Standard BBA algorithm	135
Figure 3.3. Adaptive BBA (ABBA) algorithm	137
Figure 3.4. (a) Standard MCL and (b) Adaptive MCL(AMCL).....	139
Figure 4.1. The developed wrapper BBA-MCL FS method.....	150
Figure 4.2. Convergence graph of all FS methods for DS1-DS12 datasets.....	160
Figure 5.1. Convergence graph of all FS methods for DS1-DS12 datasets.....	182
Figure 6.1. Visualize clusters for ABBA-AMCL	209
Figure 8.1. Overview of research framework	257



List of Appendices

Appendix A List of Publications	294
---------------------------------------	-----



List of Abbreviations

ED	Event Detection
SNS	Social Networks sites
NED	New Event Detection
RED	Retrospective Event Detection
FS	Feature Selection
TF	Term Frequency
TFIDF	Term Frequency Inverse Document Frequency
LDA	Latent Dirichlet Allocation
NER	Named Entity Relation
POS	Part Of Speech
MHAs	Meta-Heuristic Algorithms
BBA	Binary Bat Algorithm
BA	Bat Algorithm
r	emission rate
A	Loudness
MCL	Markov Clustering
inf	inflation
p	pruning
TDT	Topic Detection and Tracking
API	Application Programming Interface
NLP	Natural Language Processing
FE	Feature Extraction
LSI	Latent Semantic Indexing
PCA	Principal Component Analysis
CHI	Chi-square
MI	Mutual Information
DF	Document Frequency
IG	Information Gain
VSM	Vector Space Model

PSO	Particle Swarm Optimization
GA	Genetic Algorithm
GWO	Grey Wolf Optimizer
BKH	Binary Krill Herd
BCS	Binary Cuckoo Search
BBF	Binary Butterfly (BF)
BDFA	Binary Dragonfly Algorithm
BFA	Binary Firefly Algorithm
ACO	Ant Colony Optimization
ABC	Artificial Bee Colony
BWOA	Binary Whale Optimization Algorithm
BAI	Binary Ant Lion
BGSA	Binary Gravitational Search Algorithm
BFPA	Binary Flower Pollination Algorithm
SA	Simulated Annealing
HS	Harmony Search
NB	Naïve Bayes
SVM	Support Vector Machine
WBC	White Blood Cells
LR	Linear Regression
DIWBBA	Dynamic Inertia Weight BBA
CRF	Conditional Random Field
KNN	K-Nearest Neighbour
IDF	Inverse Document Frequency
DFT	Discrete Fourier Transformation
WT	Wavelet Transformation
CWT	Continues WT
AHC	Agglomerative Hierarchical Clustering
CD	Community Detection
PR	Page Rank
M	stochastic matrix

exp	expansion
TR	TextRank
LH	LUHN
CV	Comments Voting
SV	Share Voting
EV	Engagement Voting
ROUGE	Recall-Oriented Understudy for Gisty Evaluation
MMR	Maximal Marginal Relevance
BOW	Bag of Words
SFR	Selected Feature Ratio
RPD	Relative Percentage Deviation
Q	Modularity
<i>F</i>	<i>F</i> measure
<i>P</i>	<i>Precision</i>
<i>R</i>	<i>Recall</i>
Bestp	Best pruning
<i>p</i> -prob	pruning probability
Bestinf	Best <i>inf</i>
EIG	Eigenvector
GN	Girvan–Newman
LEI	Leiden
LOV	Louvain
GM	Greedy Modularity
WT	WalkTrap
LSA	Latent Semantic Analysis
LEX	LexRank
KL	KL-Sum

CHAPTER ONE

INTRODUCTION

This chapter presents the research background and the main motivation behind this study followed by an indication of the most important unresolved problems found in studies of detecting events from heterogeneous news text documents. Later, research questions and objectives were introduced along with the scope and significance of the current study.

1.1 Background

Event Detection (ED) is the process of automatically recognizing events from multiple sources of data, such as text, video, photos, and audio data (Goswami & Kumar, 2016). The majority of ED experts are interested in textual data because 80% of the data generated on the web is in the form of digital text data, which reports on real-world events (Q. Chen et al., 2017; Goswami & Kumar, 2016). Different platforms produce and circulate such data, including various news media, forums, weblogs, emails, and Social Networks Sites (SNS) like Facebook and Twitter (Goswami & Kumar, 2016). As a result, many ED scholars have developed numerous ED models, which are typically categorized into either New Event Detection (NED) models or Retrospective Event Detection (RED) models (Panagiotou et al., 2016).

Unlike the NED model, the RED model is applied to the entire corpus rather than a specified time window (Wei et al., 2018). Despite the fact that RED has been extensively studied for a long time, it is still an active and fascinating research topic

REFERENCES

- Abasi, A. K., Khader, A. T., Al-Betar, M. A., Naim, S., Makhadmeh, S. N., & Alyasseri, Z. A. A. (2020). Link-based multi-verse optimizer for text documents clustering. *Applied Soft Computing*, 87, 106002. <https://doi.org/10.1016/j.asoc.2019.106002>
- Abasi, A. K., Khader, A. T., Al-Betar, M. A., Naim, S., Makhadmeh, S. N., & Alyasseri, Z. A. A. (2021). An improved text feature selection for clustering using binary grey wolf optimizer. *In Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019*, 503–516. https://doi.org/10.1007/978-981-15-5281-6_34
- Abdul-Mageed, M. M. (2008). Online news sites and journalism 2.0: Reader comments on Al Jazeera Arabic. *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 6(2), 59–76. <https://doi.org/10.31269/triplec.v6i2.78>
- Abhik, D., & Toshniwal, D. (2013). Sub-event detection during natural hazards using features of social media data. *In Proceedings of the 22nd International Conference on World Wide Web*, 783–788. <https://doi.org/10.1145/2487788.2488046>
- Abualigah, L., Gandomi, A. H., Elaziz, M. A., Hamad, H. Al, Omari, M., Alshinwan, M., & Khasawneh, A. M. (2021). Advances in meta-heuristic optimization algorithms in big data text clustering. *Electronics*, 10(2), 101.
- Abualigah, L., Gandomi, A. H., Elaziz, M. A., Hussien, A. G., Khasawneh, A. M., Alshinwan, M., & Houssein, E. H. (2020). Nature-inspired optimization algorithms for text document clustering—a comprehensive analysis. *Algorithms*, 13(12), 345.
- Abualigah, L. M., & Khader, A. T. (2017). Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. *The Journal of Supercomputing*, 73(11), 4773–4795. <https://doi.org/10.1109/csit.2016.7549453>
- Abualigah, L. M., Khader, A. T., & Al-Betar, M. A. (2016). Unsupervised feature selection technique based on genetic algorithm for improving the text clustering. *2016 7th International Conf. on Computer Science and Information Technology (CSIT)*, 1–6. <https://doi.org/10.1109/csit.2016.7549453>
- Abualigah, L. M., Khader, A. T., Al-Betar, M. A., & Awadallah, M. A. (2016). A krill herd algorithm for efficient text documents clustering. *In 2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 67–72. <https://doi.org/10.1109/iscaie.2016.7575039>
- Abualigah, L. M., Khader, A. T., Al-Betar, M. A., & Hanandeh, E. S. (2016). Unsupervised text feature selection technique based on particle swarm optimization algorithm for improving the text clustering. *1st EAI International Conf. on Computer Science and Engineering*, 169. <https://doi.org/10.4108/eai.27->

- Abualigah, L. M., Khader, A. T., & Hanandeh, E. S. (2018). A new feature selection method to improve the document clustering using particle swarm optimization algorithm. *Journal of Computational Science*, 25(March), 456–466. <https://doi.org/https://doi.org/10.1016/j.jocs.2017.07.018>
- Abulaish, M., Sharma, S., & Fazil, M. (2019). A multi-attributed graph-based approach for text data modeling and event detection in Twitter. *2019 11th International Conf. on Communication Systems & Networks (COMSNETS)*, 703–708. <https://doi.org/10.1109/COMSNETS.2019.8711451>
- Afrabandpey, H., Ghaffari, M., Mirzaei, A., & Safayani, M. (2014). A novel bat algorithm based on chaos for optimization tasks. *Intelligent Systems (ICIS), 2014 Iranian Conference On*, 1–6. <https://doi.org/10.1109/IranianCIS.2014.6802527>
- Afriyani, R., Bustamam, A., & Sarwinda, D. (2021). Analyzing protein-protein interactions of coronavirus using markov clustering with cuckoo search and ant lion optimization. *Journal of Physics: Conference Series*, 1722(1), 12009.
- Agarwal, S., & Ranjan, P. (2016). Dimensionality reduction methods classical and recent trends: A survey. *IJCTA*, 9(10), 4801–4808.
- Aggarwal, C. C., & Subbian, K. (2012). Event detection in social streams. *In Proceedings of the 2012 SIAM International Conference on Data Mining*, 624–635. <https://doi.org/10.1137/1.9781611972825.54>
- Ahmed, F., & Abulaish, M. (2012). An MCL-based approach for spam profile detection in online social networks. *2012 IEEE 11th International Conf. on Trust, Security and Privacy in Computing and Communications*, 602–608. <https://doi.org/10.1109/TrustCom.2012.83>
- Ahn, B. G., Van Durme, B., & Callison-Burch, C. (2011). WikiTopics: What is popular on Wikipedia and why. *In Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages*, 33–40.
- Akachar, E., Ouhbi, B., & Frikh, B. (2018). Community detection in social networks using structural and content information. *In Proceedings of the 20th International Conference on Information Integration and Web-Based Applications & Services*, 282–288. <https://doi.org/10.1145/3282373.3282399>
- Akhtar, N., Beg, M. M. S., & Javed, H. (2019). Textrank enhanced topic model for query focussed text summarization. *In 2019 Twelfth International Conference on Contemporary Computing (IC3)*, 1–6. <https://doi.org/10.1109/IC3.2019.8844939>
- Akila, S., & Christe, S. A. (2022). A wrapper based binary bat algorithm with greedy crossover for attribute selection. *Expert Systems with Applications*, 187, 115828.
- Akinyelu, A. A., & Adewumi, A. O. (2018). On the performance of cuckoo search and

- bat algorithms based instance selection techniques for SVM speed optimization with application to e-fraud detection. *KSII Transactions on Internet and Information Systems (TIIS)*, 12(3), 1348–1375. <https://doi.org/10.3837/tiis.2018.03.021>
- Al-fath, A. M. U., & Sa, S. (2016). Implementation of MCL Algorithm in clustering digital news with graph representation. In *2016 4th International Conference on Information and Communication Technology (ICoICT)*, 1–6. <https://doi.org/10.1109/ICoICT.2016.7571917>
- Al-Rawi, A. (2017). News values on social media: News organizations' Facebook use. *Journalism*, 18(7), 871–889. <https://doi.org/10.1177/1464884916636142>
- Al-Taani, A. T., & Al-Omour, M. M. (2014). An extractive graph-based Arabic text summarization approach. *The International Arab Conference on Information Technology*, 158–163.
- Alam, M. W. U. (2019). *Improved binary bat algorithm for feature selection*. Åbo Akademi University.
- Alami, N., El Mallahi, M., Amakdouf, H., & Qjidaa, H. (2021). Hybrid method for text summarization based on statistical and semantic treatment. *Multimedia Tools and Applications*, 80(13), 19567–19600. <https://doi.org/10.1007/s11042-021-10613-9>
- Alashri, S., Kandala, S. S., Bajaj, V., Ravi, R., Smith, K. L., & Desouza, K. C. (2016). An analysis of sentiments on Facebook during the 2016 US presidential election. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 795–802. <https://doi.org/10.1109/ASONAM.2016.7752329>
- Aleti, A., & Moser, I. (2013). Studying feedback mechanisms for adaptive parameter control in evolutionary algorithms. *2013 IEEE Congress on Evolutionary Computation*, 3117–3124. <https://doi.org/10.1109/CEC.2013.6557950>
- Ali, Z. H., & Malallah, A. P. D. S. (2019). Multilingual text summarization based on LDA and modified PageRank. *Iraqi Journal of Information Technology*, 5, 9(3), 2018. <https://doi.org/10.34279/0923-009-003-013>
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *International Journal of Advanced Computer Science and Applications (Ijacs)*, 8(10), 397.
- Alomari, O. A., Khader, A. T., Al-Betar, M. A., & Abualigah, L. M. (2017). Gene selection for cancer classification by combining minimum redundancy maximum relevancy and bat-inspired algorithm. *International Journal of Data Mining and Bioinformatics*, 19(1), 32–51. <https://doi.org/10.1504/ijdmb.2017.10009480>

- Alsaedi, N., Burnap, P., & Rana, O. (2016). Automatic summarization of real world events using Twitter. *In Proceedings of the International AAAI Conference on Web and Social Media*, 10(1).
- Alsaedi, N., Burnap, P., & Rana, O. F. (2014). A combined classification-clustering framework for identifying disruptive events. *ASE SocialCom Conference, Stanford University, CA., USA*, 1–10.
- Altuncu, M. T., Mayer, E., Yaliraki, S. N., & Barahona, M. (2019). From free text to clusters of content in health records: An unsupervised graph partitioning approach. *Applied Network Science*, 4(1), 2. <https://doi.org/https://doi.org/10.1007/s41109-018-0109-9>
- Altuncu, M. T., Yaliraki, S. N., & Barahona, M. (2018). Content-driven, unsupervised clustering of news articles through multiscale graph partitioning. *ArXiv, abs/1808.0*, 1–8. <https://arxiv.org/abs/1808.01175v1>
- Alzaqebah, M., Abdullah, S., & Jawarneh, S. (2016). Modified artificial bee colony for the vehicle routing problems with time windows. *SpringerPlus*, 5(1), 1298. <https://doi.org/10.1186/s40064-016-2940-8>
- Aramaki, E., Maskawa, S., & Morita, M. (2011). Twitter catches the flu: Detecting influenza epidemics using Twitter. *In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1568–1576.
- Arora, S., & Anand, P. (2019). Binary butterfly optimization approaches for feature selection. *Expert Systems with Applications*, 116, 147–160. <https://doi.org/10.1016/j.eswa.2018.08.051>
- Atefeh, F., & Khreich, W. (2015). A survey of techniques for event detection in Twitter. *Computational Intelligence*, 31(1), 133–164. <https://doi.org/10.1111/coin.12017>
- Atefi, K., Hashim, H., & Khodadadi, T. (2020). A hybrid anomaly classification with deep learning (DL) and binary algorithms (BA) as optimizer in the intrusion detection system (IDS). *2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, 29–34.
- Atia, S., & Shaalan, K. (2015). Increasing the accuracy of opinion mining in Arabic. *In 2015 First International Conference on Arabic Computational Linguistics (ACLing)*, 106–113. <https://doi.org/10.1109/ACLing.2015.22>
- Azad, A., Pavlopoulos, G. A., Ouzounis, C. A., Kyrpides, N. C., & Buluç, A. (2018). HipMCL: A high-performance parallel implementation of the Markov clustering algorithm for large-scale networks. *Nucleic Acids Research*, 46(6), e33–e33. <https://doi.org/10.1093/nar/gkx1313>
- Azam, N., Abulaish, M., & Haldar, N. A.-H. (2015). Twitter data mining for events classification and analysis. *2015 Second International Conf. on Soft Computing and Machine Intelligence (ISCM)*, 79–83. <https://doi.org/10.1109/ISCM.2015>

- Bacan, H., Pandzic, I. S., & Gulija, D. (2005). Automated news item categorization. *In Proceedings of the 19th Annual Conference of The Japanese Society for Artificial Intelligence*, 251–256.
- Balcerzak, B., Jaworski, W., & Wierzbicki, A. (2014). Application of TextRank algorithm for credibility assessment. *In 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 1, 451–454. <https://doi.org/10.1109/WI-IAT.2014.70>
- Baldwin, T., Cook, P., Han, B., Harwood, A., Karunasekera, S., & Moshtaghi, M. (2012). A support platform for event detection using social intelligence. *In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 69–72.
- Bangyal, W. H., Ahmad, J., Rauf, H. T., & Pervaiz, S. (2018a). An improved bat algorithm based on novel initialization technique for global optimization problem. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 9(7), 158–166.
- Bangyal, W. H., Ahmad, J., Rauf, H. T., & Pervaiz, S. (2018b). An overview of mutation strategies in bat algorithm. *International Journal of Advanced Computer Science and Applications*, 9(8), 523–534. <https://doi.org/https://doi.org/10.14569/ijacsa.2018.090866>
- Barbosa, C. E. M., & Vasconcelos, G. C. (2018). Eight bio-inspired algorithms evaluated for solving optimization problems. *International Conf. on Artificial Intelligence and Soft Computing*, 290–301. https://doi.org/https://doi.org/10.1007/978-3-319-91253-0_28
- Basheer, S., Anbarasi, M., Sakshi, D. G., & Kumar, V. V. (2020). Efficient text summarization method for blind people using text mining techniques. *International Journal of Speech Technology*, 23(4), 713–725. <https://doi.org/10.1007/s10772-020-09712-z>
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *In Proceedings of the International AAAI Conference on Web and Social Media*, 3(1), 361–362.
- Battiti, R. (1994). Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4), 537–550. <https://doi.org/10.1109/72.298224>
- Baziar, A., Kavooosi-Fard, A., & Zare, J. (2013). A novel self adaptive modification approach based on bat algorithm for optimal management of renewable MG. *Journal of Intelligent Learning Systems and Applications*, 5(01), 11. <https://doi.org/10.4236/jilsa.2013.51002>

- Becker, H., Chen, F., Iyer, D., Naaman, M., & Gravano, L. (2011). Automatic identification and presentation of Twitter content for planned events. *In Proceedings of the International AAAI Conference on Web and Social Media*, 5(1).
- Becker, H., Iyer, D., Naaman, M., & Gravano, L. (2012). Identifying content for planned events across social media sites. *In Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, 533–542. <https://doi.org/10.1145/2124295.2124360>
- Becker, H., Naaman, M., & Gravano, L. (2011a). Selecting quality Twitter content for events. *In Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 442–445.
- Becker, H., Naaman, M., & Gravano, L. (2011b). Beyond trending topics: Real-world event identification on Twitter. *In Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 1–17. <https://doi.org/10.1.1.221.2822>
- Beigh, T. M., Upadhyaya, S., & Gopal, G. (2016). Event identification in social news streams using keyword analysis. *International Research Journal of Engineering and Technology (IRJET)*, 3(5), 1781–1786. <https://doi.org/https://doi.org/10.1109/iciss.2010.5654957>
- Benson, E., Haghighi, A., & Barzilay, R. (2011). Event discovery in social media feeds. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1, 389–398.
- Bhandari, H., Shimbo, M., Ito, T., & Matsumoto, Y. (2008). Generic text summarization using probabilistic latent semantic indexing. *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*, 133–140.
- Bharti, K. K., & kumar Singh, P. (2014). A survey on filter techniques for feature selection in text mining. *Proc. of the Second International Conf. on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012*, 1545–1559. https://doi.org/10.1007/978-81-322-1602-5_154
- Bharti, K. K., & Singh, P. K. (2014). A three-stage unsupervised dimension reduction method for text clustering. *Journal of Computational Science*, 5(2), 156–169. <https://doi.org/10.1016/j.jocs.2013.11.007>
- Bharti, K. K., & Singh, P. K. (2015). Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. *Expert Systems with Applications*, 42(6), 3105–3114. <https://doi.org/10.1016/j.eswa.2014.11.038>
- Bharti, K. K., & Singh, P. K. (2016a). Chaotic gradient artificial bee colony for text clustering. *Soft Computing*, 20(3), 1113–1126. <https://doi.org/10.1007/s00500-014-1571-7>
- Bharti, K. K., & Singh, P. K. (2016b). Opposition chaotic fitness mutation based

- adaptive inertia weight BPSO for feature selection in text clustering. *Applied Soft Computing*, 43, 20–34. <https://doi.org/10.1016/j.asoc.2016.01.019>
- Biswas, B. (2014). Comparison of algorithms for social networks using ontology. *International Journal of Computer Applications*, 85(13), 31–34. <https://doi.org/10.5120/14903-3396>
- Blanco, R., & Lioma, C. (2012). Graph-based term weighting for information retrieval. *Information Retrieval*, 15(1), 54–92. <https://doi.org/10.1007/s10791-011-9172-x>
- Blanco, R., & Lioma, C. (2007). Random walk term weighting for information retrieval. *In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 829–830. <https://doi.org/10.1145/1277741.1277930>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. <https://doi.org/10.1145/2107736.2107741>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 1–12. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bokhari, M. U., & Adhami, M. K. (2015). Event evolution modeling for efficient news search. *International Journal of Computer Applications*, 117(14), 23–29. <https://doi.org/10.5120/20623-3347>
- Boukhari, N., Debbat, F., Monmarché, N., & Slimane, M. (2018). A study on self-adaptation in the evolutionary strategy algorithm. *IFIP International Conf. on Computational Intelligence and Its Applications*, 150–160. https://doi.org/10.1007/978-3-319-89743-1_14
- Brezocnik, L., Fister, I., & Podgorelec, V. (2018). Swarm intelligence algorithms for feature selection: A review. *Applied Sciences*, 8(9), 1521. <https://doi.org/10.3390/app8091521>
- Burney, A., Sami, B., Mahmood, N., Abbas, Z., & Rizwan, K. (2012). Urdu text summarizer using sentence weight algorithm for word processors. *International Journal of Computer Applications*, 46(19), 38–43. <https://doi.org/10.1.1.735.9870>
- Bustamam, A., Mujtahidah, I., & Lestari, D. (2018). Applications of fruit fly optimization algorithm for analyzing protein-protein interaction through Markov clustering on HIV virus. *AIP Conference Proceedings*, 2023(1), 20231.
- Bustamam, A., Nurazmi, V. Y., & Lestari, D. (2018). Applications of cuckoo search optimization algorithm for analyzing protein-protein interaction through Markov clustering on HIV. *In Proceedings of the 3rd International Symposium on Current*

Progress in Mathematics and Sciences 2017 (ISCPMS2017), 2023(1), 020232.
<https://doi.org/10.1063/1.5064229>

- Bustamam, A., Siswantining, T., Febriyani, N. L., Novitasari, I. D., & Cahyaningrum, R. D. (2017). Protein sequences clustering of herpes virus by using Tribe Markov clustering (Tribe-MCL). *AIP Conference Proceedings*, 1862(1), 30150.
- Bustamam, A., Wisnubroto, M. S., & Lestari, D. (2018). Analysis of protein-protein interaction network using Markov clustering with pigeon-inspired optimization algorithm in HIV (human immunodeficiency virus). *Proc. of the 3rd International Symposium on Current Progress in Mathematics and Sciences 2017 (ISCPMS2017)*, 2023(1), 20229. <https://doi.org/https://doi.org/10.1063/1.5064226>
- Cai, X., Gao, X., & Xue, Y. (2016). Improved bat algorithm with optimal forage strategy and random disturbance strategy. *International Journal of Bio-Inspired Computation*, 8(4), 205–214. <https://doi.org/10.1504/IJBIC.2016.078666>
- Cai, X., Wang, L., Kang, Q., & Wu, Q. (2014). Bat algorithm with Gaussian walk. *International Journal of Bio-Inspired Computation*, 6(3), 166–174. <https://doi.org/10.1504/IJBIC.2014.062637>
- Cataldi, M., Di Caro, L., & Schifanella, C. (2010). Emerging topic detection on twitter based on temporal and social terms evaluation. In *MDMKDD'10 Proceedings of the Tenth International Workshop on Multimedia Data*, 1–4. <https://doi.org/10.1145/1814245.1814249>
- Cawley, G. C., Talbot, N. L., & Girolami, M. (2007). Sparse multinomial logistic regression via bayesian L1 regularisation. In *Advances in neural information processing systems*.
- Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1), 10–17. <https://ojs.aaai.org/index.php/ICWSM/article/view/14033>
- Chakri, A., Khelif, R., Benouaret, M., & Yang, X.-S. (2017). New directional bat algorithm for continuous optimization problems. *Expert Systems with Applications*, 69, 159–175. <https://doi.org/10.1016/j.eswa.2016.10.050>
- Chang, Y.-L., & Chien, J.-T. (2009). Latent Dirichlet learning for document summarization. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1689–1692. <https://doi.org/10.1109/ICASSP.2009.4959927>
- Chatra, K., Kuppili, V., Edla, D. R., & Verma, A. K. (2019). Cancer data classification using binary bat optimization and extreme learning machine with a novel fitness function. *Medical & Biological Engineering & Computing*, 57(12), 2673–2682.
- Chechelnytsky, D. (2018). *Deep neural models to represent news events*. University

of Stavanger, Norway.

- Chen, H., Hou, Q., Han, L., Hu, Z., Ye, Z., Zeng, J., & Yuan, J. (2019). Distributed text feature selection based on bat algorithm optimization. *2019 10th IEEE International Conf. on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 1*, 75–80. <https://doi.org/10.1109/IDAACS.2019.8924308>
- Chen, H. P., Hsu, K. W., & Chiu, S. I. (2016). Event detection in an ego network on Facebook. *In Pacific Asia Conference on Information Systems, PACIS 2016 - Proceedings*, 172.
- Chen, Q., Guo, X., & Bai, H. (2017). Semantic-based topic detection using Markov decision processes. *Neurocomputing*, 242(June), 40–50. https://doi.org/https://doi.org/10.1007/978-3-642-12538-6_6
- Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., & Leskovec, J. (2014). Can cascades be predicted? *In Proceedings of the 23rd International Conference on World Wide Web*, 925–936. <https://doi.org/10.1145/2566486.2567997>
- Cheng, S., Liu, B., Shi, Y., Jin, Y., & Li, B. (2016). Evolutionary computation and big data: key challenges and future directions. *International Conference on Data Mining and Big Data*, 3–14.
- Cheong, C., & Cheong, F. (2011). Social media data mining: A social network analysis of Tweets during the 2010-2011 Australian floods. *PACIS 2011 Proceedings*, 46. <https://aisel.aisnet.org/pacis2011/46>
- Cheruku, R., Edla, D. R., Kuppili, V., & Dharavath, R. (2018). RST-BatMiner: A fuzzy rule miner integrating rough set feature selection and bat optimization for detection of diabetes disease. *Applied Soft Computing*, 67, 764–780. <https://doi.org/10.1016/J.ASOC.2017.06.032>
- Chowdhury, S. R., Sarkar, K., & Dam, S. (2017). An approach to generic Bengali text summarization using latent semantic analysis. *2017 International Conference on Information Technology (ICIT)*, 11–16. <https://doi.org/10.1109/ICIT.2017.12>
- Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), 66111-1-66111–66116. <https://doi.org/10.1103/PhysRevE.70.066111>
- Cordeiro, M. (2012). Twitter event detection: Combining wavelet analysis and topic inference summarization. *In Doctoral Symposium on Informatics Engineering, 1*, 11–16.
- Cracs, C. S., & Porto, P. (2018). A three-step data-mining analysis of top-ranked higher education institutions' communication on Facebook. *In Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 923–929. <https://doi.org/10.1145/3284179.3284342>

- CrowdTangle. (2015). *The Most Influential News Pages on Facebook in 2015*. Trending Top Most. <https://blog.crowdtangle.com/the-biggest-news-pages-on-facebook-in-2015-c429f9307a8f>
- Culotta, A. (2010). Towards detecting influenza epidemics by analyzing Twitter messages. *In Proceedings of the First Workshop on Social Media Analytics*, 115–122. <https://doi.org/10.1145/1964858.1964874>
- Cvijikj, I. P., & Michahelles, F. (2011). Monitoring trends on Facebook. *In Proceedings - IEEE 9th International Conference on Dependable, Autonomic and Secure Computing, DASC 2011*, 895–902. <https://doi.org/10.1109/DASC.2011.150>
- Dai, X., He, Y., & Sun, Y. (2010). A two-layer text clustering approach for retrospective news event detection. *In 2010 International Conference on Artificial Intelligence and Computational Intelligence, 1*, 364–368. <https://doi.org/10.1109/AICI.2010.83>
- Dai, X., & Sun, Y. (2010). Event identification within news topics. *2010 International Conf. on Intelligent Computing and Integrated Systems*, 498–502. <https://doi.org/10.1109/ICISS.2010.5654957>
- de Lacerda, M. G. P. de. (2021). *Out-of-the-box parameter control for evolutionary and swarm-based algorithms with distributed reinforcement learning*.
- de Lacerda, M. G. P., de Araujo Pessoa, L. F., de Lima Neto, F. B., Ludermir, T. B., & Kuchen, H. (2021). A systematic literature review on general parameter control for evolutionary and swarm-based algorithms. *Swarm and Evolutionary Computation*, 60, 100777.
- Deng, J., Qiao, F., Li, H., Zhang, X., & Wang, H. (2015). An overview of event extraction from Twitter. *In 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 251–256. <https://doi.org/10.1109/CyberC.2015.24>
- Dewan, P., & Kumaraguru, P. (2014). It doesn't break just on Twitter. Characterizing Facebook content during real world events. *ArXiv E-Prints*, 1405–4820.
- Dewan, P., & Kumaraguru, P. (2015). Towards automatic real time identification of malicious posts on Facebook. *In 2015 13th Annual Conference on Privacy, Security and Trust (PST)*, 85–92. <https://doi.org/10.1109/PST.2015.7232958>
- Dhal, K. G., & Das, S. (2018). A dynamically adapted and weighted Bat algorithm in image enhancement domain. *Evolving Systems*, 10(2), 1–19. <https://doi.org/https://doi.org/10.1007/s12530-018-9216-1>
- Dhar, A., Dash, N. S., & Roy, K. (2018). Efficient feature selection based on modified Cuckoo search optimization problem for classifying Web text documents. *In International Conference on Recent Trends in Image Processing and Pattern Recognition*, 640–651. https://doi.org/10.1007/978-981-13-9187-3_57

- Dhiman, A., & Toshniwal, D. (2020). An approximate model for event detection from Twitter data. *IEEE Access*, 8, 122168–122184. <https://doi.org/10.1109/ACCESS.2020.3007004>
- Diao, Q., Jiang, J., Zhu, F., & Lim, E.-P. (2012). Finding bursty topics from microblogs. *In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers, 1*, 536–544.
- Divyashesh, V., & Pandey, A. (2019). High-dimensional data classification using PSO and bat algorithm. *In Computational Intelligence: Theories, Applications and Future Directions, 1*, 41–51. <https://doi.org/10.1007/s10618-015-0421-2>
- Dong, X., Mavroeidis, D., Calabrese, F., & Frossard, P. (2015). Multiscale event detection in social media. *Data Mining and Knowledge Discovery*, 29(5), 1374–1405. <https://doi.org/10.1007/s10618-015-0421-2>
- Doreswamy, H., & Salma, U. M. (2016). A binary bat inspired algorithm for the classification of breast cancer. *International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI)*, 5(2), 1–21. <https://doi.org/10.5121/ijscai.2016.5301>
- Durán, C., Muscoloni, A., & Cannistraci, C. V. (2021). Geometrical inspired pre-weighting enhances Markov clustering community detection in complex networks. *Applied Network Science*, 6(1), 1–16.
- Dutta, S., Chandra, V., Mehra, K., Ghatak, S., Das, A. K., & Ghosh, S. (2019). Summarizing microblogs during emergency events: A comparison of extractive summarization algorithms. *Emerging Technologies in Data Mining and Information Security*, 813, 859–872. https://doi.org/10.1007/978-981-13-1498-8_76
- Edouard, A. (2018). *Event detection and analysis on short text messages*. Université Côte d'Azur.
- Edouard, A., Cabrio, E., Tonelli, S., & Le Thanh, N. (2017). Graph-based event extraction from Twitter. *Proc. of the International Conf. Recent Advances in Natural Language Processing, RANLP 2017*, 222–230. https://doi.org/10.26615/978-954-452-049-6_031
- Eiben, Á. E., Hinterding, R., & Michalewicz, Z. (1999). Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 3(2), 124–141. <https://doi.org/10.1109/4235.771166>
- El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2020). EdgeSumm: Graph-based framework for automatic text summarization. *Information Processing & Management*, 57(6), 102264. <https://doi.org/10.1016/j.ipm.2020.102264>
- Elbarougy, R., Behery, G., & El Khatib, A. (2020). Extractive Arabic text

- summarization using modified PageRank algorithm. *Egyptian Informatics Journal*, 21(2), 73–81. <https://doi.org/10.1016/j.eij.2019.11.001>
- Elena. (2018). Top 10 Most Watched News Channels In The World 2018 | Trendrr. In *trendrr*. <https://www.trendrr.net/5197/ten-best-most-watched-news-channels-in-the-world-famous-biggest/>
- Emary, E., Yamany, W., & Hassanien, A. E. (2014). New approach for feature selection based on rough set and bat algorithm. *2014 9th International Conf. on Computer Engineering & Systems (ICCES)*, 346–353. <https://doi.org/10.1109/icces.2014.7030984>
- Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary ant lion approaches for feature selection. *Neurocomputing*, 213, 54–65. <https://doi.org/10.1016/j.neucom.2016.03.101>
- Enache, A.-C., & Sgarciu, V. (2015). An improved bat algorithm driven by support vector machines for intrusion detection. *Computational Intelligence in Security for Information Systems Conference*, 41–51.
- Enache, A. C., & Science, C. (2015). Intelligent feature selection method rooted in binary bat algorithm for intrusion detection. *2015 IEEE 10th Jubilee International Symposium on Applied Computational Intelligence and Informatics*, 517–521. <https://doi.org/10.1109/saci.2015.7208259>
- Enache, A. C., & Sgarciu, V. (2015). A feature selection approach implemented with the binary bat algorithm applied for intrusion detection. *2015 38th International Conference on Telecommunications and Signal Processing (TSP)*, 11–15.
- Enache, A. C., & Sgarciu, V. (2014a). Anomaly intrusions detection based on support vector machines with bat algorithm. *2014 18th International Conf. on System Theory, Control and Computing (ICSTCC)*, 856–861. <https://doi.org/10.1109/icstcc.2014.6982526>
- Enache, A. C., & Sgarciu, V. (2014b). Enhanced intrusion detection system based on bat algorithm-support vector machine. *Proc. of the 11th International Conf. on Security and Cryptography*, 184–189. <https://doi.org/10.5220/0005015501840189>
- Enache, A. C., Sgarciu, V., & Togan, M. (2017). Comparative Study on Feature Selection Methods Rooted in Swarm Intelligence for Intrusion Detection. *Proc. - 2017 21st International Conf. on Control Systems and Computer, CSCS 2017*, 239–244. <https://doi.org/10.1109/CSCS.2017.40>
- Erkan, G., & Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22, 457–479. <https://doi.org/10.1613/jair.1523>
- Fang, Y., Zhang, H., Ye, Y., & Li, X. (2014). Detecting hot topics from Twitter: A multiview approach. *Journal of Information Science*, 40(5), 578–593.

<https://doi.org/10.1177%2F0165551514541614>

- Fister, I., Fong, S., & Brest, J. (2014). A novel hybrid self-adaptive bat algorithm. *The Scientific World Journal*, 2014, 709–738. <https://doi.org/10.1155/2014/709738>
- Fister, I., Yang, X. S., Fong, S., & Zhuang, Y. (2014). Bat algorithm: Recent advances. *2014 IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI)*, 163–167. <https://doi.org/10.1109/cinti.2014.7028669>
- Florence, R., Nogueira, B., & Marcacini, R. (2017). Constrained hierarchical clustering for news events. *In Proceedings of the 21st International Database Engineering & Applications Symposium*, 49–56. <https://doi.org/10.1145/3105831.3105859>
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5), 75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>
- Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659, 1–44. <https://doi.org/10.1016/j.physrep.2016.09.002>
- Fung, G. P. C., Yu, J. X., Yu, P. S., & Lu, H. (2005). Parameter free bursty events detection in text streams. *In Proceedings of the 31st International Conference on Very Large Data Bases*, 181–192.
- GabAllah, N. A., & Rafea, A. (2019). Unsupervised topic extraction from Twitter: A feature-pivot approach. *In WEBIST 2019-Proceedings of the 15th International Conference on Web Information Systems and Technologies*, 185–192. <https://doi.org/10.5220/0007959001850192>
- Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: A survey. *Artificial Intelligence Review*, 47(1), 1–66. <https://doi.org/10.1007/s10462-016-9475-9>
- Gandomi, A. H., & Yang, X. S. (2014). Chaotic bat algorithm. *Journal of Computational Science*, 5(2), 224–232. <https://doi.org/10.1016/j.jocs.2013.10.002>
- García, S., Molina, D., Lozano, M., & Herrera, F. (2009). A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. *Journal of Heuristics*, 15(6), 617–644. <https://doi.org/10.1007/s10732-008-9080-4>
- Garg, D. (2012). Comparative analysis of dynamic graph techniques and data structure. *International Journal of Computer Applications*, 45(5), 41–46.
- Garg, M., & Kumar, M. (2016). Review on event detection techniques in social multimedia. *Online Information Review*, 40(3), 347–361. <https://doi.org/10.1108/OIR-08-2015-0281>
- Gashi, R., & Ahmeti, H. G. (2021). Impact of social media on the development of new products, marketing and customer relationship management in Kosovo. *Emerging*

Science Journal, 5(2), 125–138. <https://doi.org/10.28991/esj-2021-01263>

- Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821–7826. <https://doi.org/10.1073/pnas.122653799>
- Gong, Y., & Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 19–25. <https://doi.org/10.1145/383952.383955>
- Goswami, A., & Kumar, A. (2016). A survey of event detection techniques in online social networks. *Social Network Analysis and Mining*, 6(1), 107. <https://doi.org/10.1007/s13278-016-0414-1>
- Gunawan, D., Harahap, S. H., & Rahmat, R. F. (2019). Multi-document summarization by using TextRank and maximal marginal relevance for text in bahasa Indonesia. *In 2019 International Conference on ICT for Smart Society (ICISS)*, 7, 1–5. <https://doi.org/10.1109/ICISS48059.2019.8969785>
- Gupta, D., Arora, J., Agrawal, U., Khanna, A., & de Albuquerque, V. H. C. (2019). Optimized binary bat algorithm for classification of white blood cells. *Measurement*, 143, 180–190. <https://doi.org/10.1016/j.measurement.2019.01.002>
- Gustavsson, P., & Jönsson, A. (2010). Text summarization using random indexing and pagerank. *In Proceedings of the Third Swedish Language Technology Conference (Sltc-2010), Linköping, Sweden*.
- Gwadera, R., & Crestani, F. (2009). Mining and ranking streams of news stories using cross-stream sequential patterns. *In Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 1709–1712. <https://doi.org/10.1145/1645953.1646210>
- Haghighi, A., & Vanderwende, L. (2009). Exploring content models for multi-document summarization. *In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 362–370.
- Harish, B. S. (2010). Representation and classification of text documents : A brief review. *IJCA, Special Issue on RTIPPR (2)*, 110–119. <https://doi.org/10.1.1.206.3120>
- Harish, B. S., & Revanasiddappa, M. B. (2017). A comprehensive survey on various feature selection methods to categorize text documents. *International Journal of Computer Applications*, 164(8), 1–7. <https://doi.org/10.5120/ijca2017913711>

- Hasan, M., Orgun, M. A., & Schwitter, R. (2018). A survey on real-time event detection from the Twitter data stream. *Journal of Information Science*, 44(4), 443–463. <https://doi.org/10.1177/0165551517698564>
- Hassanian-esfahani, R., & Kargar, M. (2016). A survey on web news retrieval and mining. *2016 Second International Conf. on Web Research (ICWR)*, 90–101. <https://doi.org/10.1109/ICWR.2016.7498452>
- Hille, S., & Bakker, P. (2013). I like news. Searching for the “Holy Grail” of social media: The use of Facebook by Dutch news media and their audiences. *European Journal of Communication*, 28(6), 663–680. <https://doi.org/10.1177/0267323113497435>
- Hogenboom, F., Frasinca, F., Kaymak, U., & Jong, F. De. (2011). An overview of event extraction from text. In *Proceedings of Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2011), Workshop in Conjunction with the 10th International Semantic Web Conference 2011 (ISWC 2011)*, 48–57. <https://doi.org/10.1.1.369.7040>
- Hogenboom, F., Frasinca, F., Kaymak, U., Jong, F. De, & Caron, E. (2016). A survey of event extraction methods from text for decision support systems. *Decision Support Systems*, 85, 12–22. <https://doi.org/10.1016/j.dss.2016.02.006>
- Holcomb, J., Gottfried, J., Mitchell, A., & Schillinger, J. (2013). News use across social media platforms. *Pew Research Journalism Project*.
- Hong, S.-S., Lee, W., & Han, M.-M. (2015). The feature selection method based on genetic algorithm for efficient of text clustering and text classification. *International Journal of Advances in Soft Computing & Its Applications*, 7(1), 22–40. <https://doi.org/https://scholarworks.bwise.kr/gachon/handle/2020.sw.gachon/11004>
- Hospedales, T., Gong, S., & Xiang, T. (2009). A markov clustering topic model for mining behaviour in video. *2009 IEEE 12th International Conf. on Computer Vision*, 1165–1172. <https://doi.org/10.1109/ICCV.2009.5459342>
- Hossny, A. H., Mitchell, L., Lothian, N., & Osborne, G. (2020). Feature selection methods for event detection in Twitter: A text mining approach. *Social Network Analysis and Mining*, 10(1), 1–15. <https://doi.org/10.1007/s13278-020-00658-3>
- Hsu, H.-H., & Hsieh, C.-W. (2010). Feature Selection via Correlation Coefficient Clustering. *Journal of Software*, 5(12), 1371–1377. <https://doi.org/10.4304/jsw.5.12.1371-1377>
- Hu, L., Zhang, B., Hou, L., & Li, J. (2017). Adaptive online event detection in news streams. *Knowledge-Based Systems*, 138, 105–112. <https://doi.org/10.1016/j.knsys.2017.09.039>
- Huang, D., Hu, S., Cai, Y., & Min, H. (2014). Discovering event evolution graphs based

- on news articles relationships. In *2014 IEEE 11th International Conference on E-Business Engineering*, 246–251. <https://doi.org/10.1109/ICEBE.2014.49>
- Huang, P. C., Wu, J. S., & Lee, C. N. (2015). Negative emotion event detection for Chinese posts on Facebook. In *2015 International Conference on Cloud Computing and Big Data (CCBD)*, 329–335. <https://doi.org/10.1109/CCBD.2015.32>
- Huang, W., Li, Z., Zhang, L., & Li, Y. (2016). Review of intelligent microblog short text processing. *Web Intelligence*, 14(3), 211–228. <https://doi.org/10.3233/WEB-160340>
- Huang, Xiaohui, Zhang, X., Ye, Y., Deng, S., & Li, X. (2013). A topic detection approach through hierarchical clustering on concept graph. *Applied Mathematics & Information Sciences*, 7(6), 2285. <https://doi.org/10.12785/amis/070619>
- Huang, Xingwang, Zeng, X., & Han, R. (2017). Dynamic inertia weight binary bat algorithm with neighborhood search. *Computational Intelligence and Neuroscience*, 2017, 1–15. <https://doi.org/10.1155/2017/3235720>
- Hussien, A. G., Hassanien, A. E., Houssein, E. H., Bhattacharyya, S., & Amin, M. (2019). S-shaped binary whale optimization algorithm for feature selection. In *Recent trends in signal and image processing* (pp. 79–87). Springer. https://doi.org/10.1007/978-981-10-8863-6_9
- Ibrahim, A., Elghazaly, T., & Gheith, M. (2013). A novel Arabic text summarization model based on rhetorical structure theory and vector space model. *International Journal of Computational Linguistics and Natural Language Processing*, 2(8), 480–485.
- Inouye, D., & Kalita, J. K. (2011). Comparing twitter summarization algorithms for multiple post summaries. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 298–306. <https://doi.org/10.1109/PASSAT/SocialCom.2011.31>
- Ishikawa, S., Arakawa, Y., Tagashira, S., & Fukuda, A. (2012). Hot topic detection in local areas using Twitter and Wikipedia. In *2012 International Conference on Architecture of Computing Systems, ARCS 2012*, 1–5.
- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS One*, 9(6), e98679. <https://doi.org/10.1371/journal.pone.0098679>
- Jain, D., Borah, M. D., & Biswas, A. (2020). Fine-tuning textrank for legal document summarization: A bayesian optimization based approach. In *Forum for Information Retrieval Evaluation*, 41–48. <https://doi.org/10.1145/3441501.3441502>

- Janani, G., Rajamohana, S. P., Anitha, M., & Umamaheswari, K. (2017). Feature Selection Using Hybrid Approach for Opinion Spam Detection. *The International Journal of Engineering and Science (IJES)*, 6(9), 61–72.
- Jayabarathi, T., Raghunathan, T., & Gandomi, A. H. (2018). The bat algorithm, variants and some practical engineering applications: A review. In *Nature-Inspired Algorithms and Applied Optimization* (pp. 313–330). Springer. https://doi.org/10.1007/978-3-319-67669-2_14
- Jensi, R., & Jiji, D. G. W. (2014). A survey on optimization approaches to text document clustering. *International Journal on Computational Sciences & Applications (IJCSA)*, 3(6), 1–14. <https://doi.org/10.5121/ijcsa.2013.3604>
- Jordehi, R. (2015). Chaotic bat swarm optimisation (CBSO). *Applied Soft Computing*, 26, 523–530. <https://doi.org/10.1016/j.asoc.2014.10.010>
- Joshi, A., Fidalgo, E., Alegre, E., & Al Nabki, M. W. (2018). Extractive text summarization in dark web: A preliminary study. *International Conference of Applications of Intelligent Systems*.
- Jovic, a., Brkic, K., & Bogunovic, N. (2015). A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1200–1205. <https://doi.org/10.1109/mipro.2015.7160458>
- Ju, A., Jeong, S. H., & Chyi, H. I. (2014). Will social media save newspapers? Examining the effectiveness of Facebook and Twitter as news platforms. *Journalism Practice*, 8(1), 1–17. <https://doi.org/10.1080/17512786.2013.794022>
- Jung, J., Miyake, M., & Akama, H. (2005). Recurrent Markov Cluster (RMCL) algorithm for the refinement of the semantic network. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, 1428–1431.
- Kaleel, S. B., AlMeshary, M., & Abhari, A. (2013). Event detection and trending in multiple social networking sites. In *Proceedings of the 16th Communications & Networking Symposium*, 1–5.
- Kannan, S., & Gurusamy, V. (2014). Preprocessing techniques for text mining. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.
- Kashef, S., & Nezamabadi-pour, H. (2014). An advanced ACO algorithm for feature subset selection. *Neurocomputing*, 147, 271–279. <https://doi.org/10.1016/j.neucom.2014.06.067>
- Kashef, S., & Nezamabadi-pour, H. (2013). A new feature selection algorithm based on binary ant colony optimization. In *The 5th Conference on Information and Knowledge Technology*, 50–54. <https://doi.org/10.1109/IKT.2013.6620037>

- Katragadda, S., Benton, R., & Raghavan, V. (2017). Framework for real-time event detection using multiple social media sources. *In Proceedings of the 50th Hawaii International Conference on System Sciences*, 1–10.
- Khan, B. S., & Niazi, M. A. (2017). Network community detection : A review and visual survey. *ArXiv E-Prints*, arXiv:1708.00977.
- Khatdeo, S., Shrawane, S., Kumbhare, P., & Nimbarte, P. M. S. (2017). Detection and visualization of events from online news. *International Journal on Recent and Innovation Trends in Computing and Communication*, 1(21), 241–244.
- Kumar, A., Sharma, A., Sharma, S., & Kashyap, S. (2017). Performance analysis of keyword extraction algorithms assessing extractive text summarization. *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, 408–414. <https://doi.org/10.1109/COMPTELIX.2017.8004004>
- Kumar, M., & Rehan, P. (2020). Graph node rank based important keyword detection from Twitter. *Applied Computing and Informatics*.
- Kumar, V. (2014). Feature Selection: A literature Review. *The Smart Computing Review*, 4(3), 211–229. <https://doi.org/10.6029/smarter.2014.03.007>
- Kumar, Y. J., Goh, O. S., Ghani, M. K. A., Salim, N., & Albaham, A. T. (2014). Voting models for summary extraction from text documents. *In 2014 International Conference on IT Convergence and Security (ICITCS)*, 1–4. <https://doi.org/10.1109/ICITCS.2014.7021826>
- Kushwaha, N., & Pant, M. (2018). Link based BPSO for feature selection in big data text clustering. *Future Generation Computer Systems*, 82, 190–199. <https://doi.org/10.1016/j.future.2017.12.005>
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *In Proceedings of the 19th International Conference on World Wide Web*, 591–600. <https://doi.org/10.1145/1772690.1772751>
- Kwan, E., Hsu, P.-L., Liang, J.-H., & Chen, Y.-S. (2013). Event identification for social streams using keyword-based evolving graph sequences. *In 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, 450–457. <https://doi.org/10.1145/2492517.2492528>
- Lam, S. L. Y., & Lee, D. L. (1999). Feature reduction for neural network based text categorization. *In Proc. of Sixth International Conference on Database Systems for Advanced Applications*, 195–202. <https://doi.org/10.1109/DASFAA.1999.765752>
- Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E*, 80(5), 1–12. <https://doi.org/10.1103/PhysRevE.80.056117>

- Lang, K. (1995). Newsweeder: Learning to filter netnews. In *In Machine Learning Proceedings 1995* (pp. 331–339). Elsevier.
- Lavanya, S., Kavipriya, R., Yang, Y., Carbonell, J. Q., Brown, R. D., Archibald, B., & Liu, X. (2014). A survey on event detection in news streams. *International Journal of Computer Science Trends and Technology (IJCST)*, 2(5), 33–35.
- Leban, G., Fortuna, B., Brank, J., & Grobelnik, M. (2014). Event registry – Learning about world events from news. *Proc. of the 23rd International Conf. on World Wide Web*, 107–110. <https://doi.org/10.1145/2567948.2577024>
- Leban, G., Fortuna, B., & Grobelnik, M. (2016). Using news articles for real-time cross-lingual event detection and filtering. *In NewsIR@ ECIR*, 33–38.
- Lee, R., & Sumiya, K. (2010). Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. *In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, 1–10. <https://doi.org/10.1145/1867699.1867701>
- Lei, K., & Pu, C. (2014). Complex optimization problems using highly efficient particle swarm optimizer. *Telkomnika*, 12(4), 1023–1030. <https://doi.org/10.12928/TELKOMNIKA.v12i4.535>
- Lei, X., Wang, F., Wu, F.-X., Zhang, A., & Pedrycz, W. (2016). Protein complex identification through Markov clustering with firefly algorithm on dynamic protein–protein interaction networks. *Information Sciences*, 329, 303–316.
- Li, A., Jiang, T., Wang, Q., & Yu, H. (2016). The mixture of TextRank and LexRank techniques of single document automatic summarization research in Tibetan. *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 1, 514–519. <https://doi.org/10.1109/IHMSC.2016.278>
- Li, C., Sun, A., & Datta, A. (2012). Twevent: Segment-based event detection from tweets. *In Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 155–164. <https://doi.org/10.1145/2396761.2396785>
- Li, G., & Le, C. (2019). Hybrid Binary Bat Algorithm with Cross-Entropy Method for Feature Selection. *2019 4th International Conference on Control and Robotics Engineering (ICCRE)*, 165–169.
- Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C.-C. (2012). Tedas: A twitter-based event detection and analysis system. *In 2012 IEEE 28th International Conference on Data Engineering*, 1273–1276. <https://doi.org/10.1109/ICDE.2012.125>
- Li, Y., Luo, C., & Chung, S. M. (2008). Text clustering with feature selection by using statistical data. *IEEE Transactions on Knowledge and Data Engineering*, 20(5), 641–652. <https://doi.org/10.1109/TKDE.2007.190740>

- Liang, X., Qu, Y., & Ma, G. (2012). Research on extension LexRank in summarization for opinionated texts. *In 2012 13th International Conference on Parallel and Distributed Computing, Applications and Technologies*, 517–522. <https://doi.org/10.1109/PDCAT.2012.117>
- Lin, C. Y. (2004). Rouge: A package for automatic evaluation of summaries. *In Proceedings of the ACL-02 Workshop on Automatic Summarization*, 4, 74–81.
- Liu, J., Abbass, H. A., & Tan, K. C. (2019). Evolutionary community detection algorithms. *In Evolutionary Computation and Complex Networks* (pp. 77–115). Springer, Cham. https://doi.org/10.1007/978-3-319-60000-0_5
- Liu, L., Kang, J., Yu, J., & Wang, Z. (2005). A comparative study on unsupervised feature selection methods for text clustering. *In 2005 International Conference on Natural Language Processing and Knowledge Engineering*, 597–601. <https://doi.org/10.1109/NLPKE.2005.1598807>
- Liu, N., Tang, X.-J., Lu, Y., Li, M.-X., Wang, H.-W., & Xiao, P. (2014). Topic-sensitive multi-document summarization algorithm. *In 2014 Sixth International Symposium on Parallel Architectures, Algorithms and Programming*, 69–74. <https://doi.org/10.1109/PAAP.2014.22>
- Liu, X., & Qi, D. (2016). A self-adaptive bat algorithm for camera calibration. *In 2nd International Conference on Advances in Mechanical Engineering and Industrial Informatics (AMEII 2016)*.
- Liu, Y., Tang, F., & Zeng, Z. (2015). Feature selection based on dependency margin. *IEEE Transactions on Cybernetics*, 45(6), 1209–1221. <https://doi.org/10.1109/TCYB.2014.2347372>
- Liu, Z., Lin, G., Yang, S., Feng, J., Lin, W., & Goh, W. L. (2018). Learning Markov clustering networks for scene text detection. *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 6936–6944. <https://doi.org/10.1109/CVPR.2018.00725>
- Lloret, E., Plaza, L., & Aker, A. (2018). The challenging task of summary evaluation: An overview. *Language Resources and Evaluation*, 52(1), 101–148. <https://doi.org/10.1007/s10579-017-9399-2>
- Long, R., Wang, H., Chen, Y., Jin, O., & Yu, Y. (2011). Effective event detection, tracking and summarization on microblog data. *In Proceedings of the 12th International Conference on Web-Age Information Management*, 652–663. https://doi.org/10.1007/978-3-642-23535-1_55
- Lu, Z., Yu, W., Zhang, R., Li, J., & Wei, H. (2015). Discovering event evolution chain in Microblog. *In 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*, 635–640. <https://doi.org/10.1109/HPCC-CSS->

- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), 159–165. <https://doi.org/10.1147/rd.22.0159>
- Lyu, S., Li, Z., Huang, Y., Wang, J., & Hu, J. (2019). Improved self-adaptive bat algorithm with step-control and mutation mechanisms. *Journal of Computational Science*, 30, 65–78. <https://doi.org/10.1016/j.jocs.2018.11.002>
- Ma, X.-X., & Wang, J.-S. (2018). Optimized parameter settings of binary bat algorithm for solving function optimization problems. *Journal of Electrical and Computer Engineering*, 2018.
- Mafarja, M. M., Eleyan, D., Jaber, I., Hammouri, A., & Mirjalili, S. (2017). Binary dragonfly algorithm for feature selection. *2017 International Conf. on New Trends in Computing Sciences (ICTCS)*, 12–17. <https://doi.org/10.1109/ICTCS.2017.43>
- Mafarja, M. M., & Mirjalili, S. (2017). Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, 260, 302–312. <https://doi.org/10.1016/j.neucom.2017.04.053>
- Maheswari, S., & Arunesh, K. (2020). Unsupervised Binary BAT algorithm based Network Intrusion Detection System using enhanced multiple classifiers. *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, 885–889.
- Mallick, C., Das, A. K., Dutta, M., Das, A. K., & Sarkar, A. (2019). Graph-based text summarization using modified TextRank. In *Soft computing in data analytics* (pp. 137–146). Springer. https://doi.org/10.1007/978-981-13-0514-6_14
- Manaskasemsak, B., Chinthanet, B., & Rungsawang, A. (2016). Graph clustering-based emerging event detection from twitter data stream. *Proc. of the Fifth International Conf. on Network, Communication and Computing*, 37–41. <https://doi.org/10.1145/3033288.3033312>
- Mathioudakis, M., & Koudas, N. (2010). Twittermonitor: Trend detection over the twitter stream. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, 1155–1158. <https://doi.org/10.1145/1807167.1807306>
- Meena, M. J., Chandran, K. R., Karthik, A., & Samuel, A. V. (2012). An enhanced ACO algorithm to select features for text categorization and its parallelization. *Expert Systems with Applications*, 39(5), 5861–5871. <https://doi.org/10.1016/j.eswa.2011.11.081>
- Meena, Y. K., & Gopalani, D. (2016). Improvement in quality of extractive text summaries using modified reciprocal ranking. In *Proceedings of International Conference on ICT for Sustainable Development*, 307–315. https://doi.org/10.1007/978-981-10-0135-2_30

- Mehmood, A., Palli, A. S., & Khan, M. N. A. (2014). A study of sentiment and trend analysis techniques for social media content. *International Journal of Modern Education and Computer Science*, 6(12), 47. <https://doi.org/10.5815/ijmecs.2014.12.07>
- Mehrotra, R., Sanner, S., Buntine, W., & Xie, L. (2013). Improving LDA topic models for microblogs via tweet pooling and automatic labeling. *In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 889–892. <https://doi.org/10.1145/2484028.2484166>
- Mei, Q., Ling, X., Wondra, M., Su, H., & Zhai, C. (2007). Topic sentiment mixture: Modeling facets and opinions in weblogs. *In Proceedings of the 16th International Conference on World Wide Web*, 171–180. <https://doi.org/10.1145/1242572.1242596>
- Mele, I., Bahrainian, S. A., & Crestani, F. (2019). Event mining and timeliness analysis from heterogeneous news streams. *Information Processing & Management*, 56(3), 969–993. <https://doi.org/10.1016/j.ipm.2019.02.003>
- Mele, I., & Crestani, F. (2019). A multi-source collection of event-labeled news documents. *In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, 205–208. <https://doi.org/10.1145/3341981.3344253>
- Mele, I., & Crestani, F. (2017). Event detection for heterogeneous news streams. *In International Conference on Applications of Natural Language to Information Systems*, 110–123. https://doi.org/10.1007/978-3-319-59569-6_11
- Melvin, S., Yu, W., Ju, P., Young, S., & Wang, W. (2017). Event detection and summarization using phrase network. *In Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 89–101. https://doi.org/10.1007/978-3-319-71273-4_8
- Meng, J., Lin, H., & Yu, Y. (2011). A two-stage feature selection method for text categorization. *Computers & Mathematics with Applications*, 62(7), 2793–2800. <https://doi.org/10.1016/j.camwa.2011.07.045>
- Merchant, K., & Pande, Y. (2018). Nlp based latent semantic analysis for legal text summarization. *In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1803–1807. <https://doi.org/10.1109/ICACCI.2018.8554831>
- Mhamdi, C., Al-Emran, M., & Salloum, S. A. (2018). Text mining and analytics: A case study from news channels posts on Facebook. *In Intelligent Natural Language Processing: Trends and Applications* (pp. 399–415). Springer. https://doi.org/10.1007/978-3-319-67056-0_19
- Mihalcea, R. (2005). Language independent extractive summarization. *In Proceedings of the 43th Annual Meeting of the Association for Computational Linguistics (ACL*

2005), 5, 49–52. <https://doi.org/10.3115/1225753.1225766>

- Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into text. *In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 404–411.
- Mirjalili, S., Mirjalili, S. M., & Yang, X. S. (2014). Binary bat algorithm. *Neural Computing and Applications*, 25(3–4), 663–681. <https://doi.org/10.1007/s00521-013-1525-5>
- Mohamad, A., Syed Mustapha, S., & Razali, M. (2010). Automatic event detection on Reuters news. *In Knowledge Management International Conference 2010 (KMICe2010)*, 25–27. <https://repo.uum.edu.my/id/eprint/11252>
- Mohamed, R., Technology, I., Tun, U., Onn, H., Pahat, B., & Wahid, N. (2016). The effectiveness of bat algorithm for data handling in various applications. *In 2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), November*, 151–156. <https://doi.org/10.1109/ICCSCE.2016.7893562>
- Mohammed, A. J., Yusof, Y., & Husni, H. (2016). Integrated bisect k-means and firefly algorithm for hierarchical text clustering. *Journal of Engineering and Applied Sciences*, 11(3), 522–527. <https://repo.uum.edu.my/id/eprint/20644>
- Montané, F. A. T., & Galvão, R. D. (2002). Vehicle routing problems with simultaneous pick-up and delivery service. *Opsearch*, 39(1), 19–33. <https://doi.org/10.1007/BF03398667>
- Moratanch, N., & Chitrakala, S. (2016). A survey on abstractive text summarization. *In 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, 1–7. <https://doi.org/10.1109/ICCPCT.2016.7530193>
- Moutidis, I., & Williams, H. T. P. (2020). Complex networks for event detection in heterogeneous high volume news streams. *ArXiv E-Prints*, arXiv:2005.13751. https://doi.org/10.1007/978-3-030-36687-2_55
- Moutidis, I., & Williams, H. T. P. (2019). Utilizing complex networks for event detection in heterogeneous high-volume news streams. *International Conf. on Complex Networks and Their Applications*, 659–672. https://doi.org/10.1007/978-3-030-36687-2_55
- Nakamura, R. Y. M., Pereira, L. A. M., Costa, K. A., Rodrigues, D., Papa, J. P., & Yang, X.-S. (2012). BBA: A binary bat algorithm for feature selection. *2012 25th SIBGRAPI Conf. on Graphics, Patterns and Images*, 291–297. <https://doi.org/10.1109/sibgrapi.2012.47>
- Nakamura, R. Y. M., Pereira, L. A. M., Rodrigues, D., Costa, K. A. P., Papa, J. P., & Yang, X.-S. (2013). Binary bat algorithm for feature selection. *In Swarm Intelligence and Bio-Inspired Computation* (pp. 225–237). Elsevier.

<https://doi.org/10.1016/B978-0-12-405163-8.00009-0>

- Nanba, H., Saito, R., Ishino, A., & Takezawa, T. (2013). Automatic extraction of event information from newspaper articles and web pages. *International Conf. on Asian Digital Libraries*, 171–175. https://doi.org/10.1007/978-3-319-03599-4_21
- Newman, D., Baldwin, T., Cavedon, L., Huang, E., Karimi, S., Martinez, D., Scholer, F., & Zobel, J. (2010). Visualizing search results and document collections using topic maps. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(2–3), 169–175. <https://doi.org/10.1016/j.websem.2010.03.005>
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6), 66133. <https://doi.org/10.1103/physreve.69.066133>
- Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3). <https://doi.org/10.1103/PhysRevE.74.036104>
- Nguyen, B. H., Xue, B., & Andreae, P. (2017). A novel binary particle swarm optimization algorithm and its applications on knapsack and feature selection problems. In *Intelligent and Evolutionary Systems* (pp. 319–332). Springer.
- Nguyen, B. H., Xue, B., & Zhang, M. (2020). A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation*, 54, 100663.
- Nguyen, S., Ngo, B., Vo, C., & Cao, T. (2019). Hot topic detection on twitter data streams with incremental clustering using named entities and central centroids. In *2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*, 1–6. <https://doi.org/10.1109/RIVF.2019.8713730>
- Nurwidiantoro, A., & Winarko, E. (2013). Event detection in social media: A survey. In *International Conference on ICT for Smart Society*, 1–5. <https://doi.org/10.1109/ICTSS.2013.6588106>
- Ochieng, P. J., Kusuma, W. A., & Haryanto, T. (2017). Detection of protein complex from protein-protein interaction network using Markov clustering. *Journal of Physics: Conference Series*, 835(1), 12001.
- Osborne, M., & Dredze, M. (2014). Facebook, Twitter and Google Plus for breaking news: Is there a winner? In *Proceedings of the International AAI Conference on Web and Social Media*, 8(1), 611–614.
- Osborne, M., Petrovic, S., McCreadie, R., Macdonald, C., & Ounis, I. (2012). Bieber no more: First story detection using Twitter and Wikipedia. In *Sigir 2012 Workshop on Time-Aware Information Access*, 16–76.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation*

ranking: Bringing order to the web. Stanford InfoLab.

- Panagiotou, N., Katakis, I., & Gunopulos, D. (2016). Detecting events in online social networks: Definitions, trends and challenges. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9580, pp. 42–84). Springer, Cham. https://doi.org/10.1007/978-3-319-41706-6_2
- Parikh, R. (2013). ET : Events from Tweets. In *Proceedings of the 22nd International Conference on World Wide Web*, 613–620. <https://doi.org/10.1145/2487788.2488006>
- Parpinelli, R. S., Plichoski, G. F., Da Silva, R. S., & Narloch, P. H. (2019). A review of techniques for online control of parameters in swarm intelligence and evolutionary computation algorithms. *International Journal of Bio-Inspired Computation*, 13(1), 1–20. <https://doi.org/10.1504/ijbic.2019.097731>
- Passaro, L. C., Bondielli, A., & Lenci, A. (2016). FB-NEWS15: A topic-annotated Facebook corpus for emotion detection and sentiment analysis. In *Third Italian Conference on Computational Linguistics (CLiC-It 2016)*, 228–232.
- Perez, J., Valdez, F., & Castillo, O. (2015). Modification of the bat algorithm using fuzzy logic for dynamical parameter adaptation. In *2015 IEEE Congress on Evolutionary Computation (CEC)*, 464–471. <https://doi.org/10.1109/CEC.2015.7256926>
- Perez, J., Valdez, F., Castillo, O., & Roeva, O. (2016). Bat algorithm with parameter adaptation using interval type-2 fuzzy logic for benchmark mathematical functions. In *2016 IEEE 8th International Conference on Intelligent Systems (IS)*, 120–127. <https://doi.org/10.1109/IS.2016.7737409>
- Petrović, S., Osborne, M., & Lavrenko, V. (2010). Streaming first story detection with application to twitter. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 181–189.
- Phung, V., & De Vine, L. (2015). A study on the use of word embeddings and pagerank for vietnamese text summarization. In *Proceedings of the 20th Australasian Document Computing Symposium*, 1–8. <https://doi.org/10.1145/2838931.2838935>
- Phuvipadawat, S., & Murata, T. (2010). Breaking news detection and tracking in Twitter. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 3, 120–123. <https://doi.org/10.1109/WI-IAT.2010.205>
- Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. In *International Symposium on Computer and Information Sciences*, 284–293. https://doi.org/10.1007/11569596_31

- Popescu, A.-M., & Pennacchiotti, M. (2010). Detecting controversial events from Twitter. *In Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 1873–1876. <https://doi.org/10.1145/1871437.1871751>
- Popescu, A.-M., Pennacchiotti, M., & Paranjpe, D. (2011). Extracting events and event descriptions from Twitter. *In Proceedings of the 20th International Conference Companion on World Wide Web*, 105–106. <https://doi.org/10.1145/1963192.1963246>
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137. <https://doi.org/10.1108/eb046814>
- Prasad, R., Bisandu, D., & Liman, M. (2018). Clustering news articles using efficient similarity measure and N-grams. *International Journal of Knowledge Engineering and Data Mining*, 5(1), 1. <https://doi.org/10.1504/ijkedm.2018.10016103>
- Purushothaman, R., Rajagopalan, S. P., & Dhandapani, G. (2020). Hybridizing Gray Wolf Optimization (GWO) with Grasshopper Optimization Algorithm (GOA) for text feature selection and clustering. *Applied Soft Computing*, 96, 106651.
- Qasim, O. S., & Algamal, Z. Y. (2020). Feature selection using different transfer functions for binary bat algorithm. *Int J Math Eng Manag Sci*, 5(4), 697–706.
- R. Motooka, T. Yumoto, M. Nii, Y. T., & Sumiya, K. (2011). A similar event search system using hashtag of Twitter. *The Database Society of Japan*, 3, 1–5.
- Rad, F., Nadri, A. A., & Parvin, H. (2016). Optimal feature selection for data classification and clustering: Techniques and guidelines. *International Journal of Information, Security and Systems Management*, 5(2), 583–590.
- Rafea, A., & Mostafa, N. A. (2013). Topic extraction in social media. *In 2013 International Conference on Collaboration Technologies and Systems (CTS)*, 94–98. <https://doi.org/10.1109/CTS.2013.6567212>
- Rahutomo, F., Kitasuka, T., & Aritsugi, M. (2012). Semantic cosine similarity. *The 7th International Student Conference on Advanced Science and Technology ICAST*, 4(1), 1.
- Rajalaxmi, R. R., & Ramesh, A. (2014). Binary bat approach for effective spam classification in online social networks. *Australian Journal of Basic & Applied Sciences*, 8(18), 383–388.
- Rajani, N. F. N., McArdle, K., & Baldrige, J. (2014). Extracting topics based on authors, recipients and content in microblogs. *In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1171–1174. <https://doi.org/10.1145/2600428.2609537>
- Ramadan, Q. H., & Mohd, M. (2011). A review of retrospective news event detection.

2011 *International Conf. on Semantic Technology and Information Retrieval*, 209–214. <https://doi.org/10.1109/stair.2011.5995790>

Ramadhan, M. R., Endah, S. N., & Mantau, A. B. J. (2020). Implementation of TextRank algorithm in product review summarization. *In 2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, 1–5. <https://doi.org/10.1109/ICICoS51170.2020.9299005>

Ramasamy, R., & Rani, S. S. (2018). Modified binary bat algorithm for feature selection in unsupervised learning. *The International Arab Journal of Information Technology*, 15(6), 1060–1067.

Ramli, M. R., Abas, Z. A., Desa, M. I., Abidin, Z. Z., & Alazzam, M. B. (2019). Enhanced convergence of bat algorithm based on dimensional and inertia weight factor. *Journal of King Saud University - Computer and Information Sciences*, 31(4), 452–458. <https://doi.org/10.1016/j.jksuci.2018.03.010>

Rani, S. S., & Ramasamy, R. (2015). Unsupervised feature selection using binary bat algorithm. *2015 2nd International Conf. on Electronics and Communication Systems (ICECS)*, 451–456. <https://doi.org/10.1109/ECS.2015.7124945>

Rani, U., & Bidhan, K. (2021). Comparative assessment of extractive summarization: textrank TF-IDF and LDA. *Journal of Scientific Research*, 65(1), 304–311. <https://doi.org/10.37398/JSR.2021.650140>

Rashedi, E., & Nezamabadi-pour, H. (2014). Feature subset selection using improved binary gravitational search algorithm. *Journal of Intelligent & Fuzzy Systems*, 26(3), 1211–1221. <https://doi.org/10.3233/IFS-130807>

Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2010). BGSA: Binary gravitational search algorithm. *Natural Computing*, 9(3), 727–745. <https://doi.org/10.1007/s11047-009-9175-3>

Rasouli, E., Zarifzadeh, S., & Rafsanjani, A. J. (2020). WebKey: A graph-based method for event detection in web news. *Journal of Intelligent Information Systems*, 54(3), 585–604. <https://doi.org/10.1007/s10844-019-00576-7>

Rattan, S., Kaur, S., Kansal, N., & Kaur, J. (2017). An optimized lung cancer classification system for computed tomography images. *In 2017 Fourth International Conference on Image Information Processing (ICIIP)*, 1–6. <https://doi.org/10.1109/ICIIP.2017.8313676>

Rauf, H. T., Lali, M. I. U., Babar, M. H., & Ali, A. S. (2019). Bat Algorithm with Different Initialization Approaches for Numerical Optimization. *2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 1–6.

Rieder, B. (2013). Studying Facebook via data extraction: The Netvizz application. *In Proceedings of the 5th Annual ACM Web Science Conference*, 346–355.

<https://doi.org/10.1145/2464464.2464475>

- Ritter, A., Clark, S., & Etzioni, O. (2011). Named entity recognition in tweets: An experimental study. *In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1524–1534. <https://doi.org/10.1.1.228.8539>
- Ritter, A., Etzioni, O., & Clark, S. (2012). Open domain event extraction from Twitter. *In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1104–1112. <https://doi.org/10.1145/2339530.2339704>
- Rizk-Allah, R. M., & Hassanien, A. E. (2018). New binary bat algorithm for solving 0–1 knapsack problem. *Complex & Intelligent Systems*, 4(1), 31–53. <https://doi.org/10.1007/s40747-017-0050-z>
- Rizki, A., Bustamam, A., & Sarwinda, D. (2021). Applications of cuckoo search and ant lion optimization for analyzing protein-protein interaction through regularized Markov clustering on coronavirus. *Journal of Physics: Conference Series*, 1722(1), 12008.
- Rodrigues, D., Pereira, L. A. M., Almeida, T. N. S., Papa, J. P., Souza, A. N., Ramos, C. C. O., & Yang, X.-S. (2013). BCS: A binary cuckoo search algorithm for feature selection. *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, 465–468. <https://doi.org/10.1109/ISCAS.2013.6571881>
- Rodrigues, D., Pereira, L. A. M., Nakamura, R. Y. M., Costa, K. A. P., Yang, X.-S., Souza, A. N., & Papa, J. P. (2014). A wrapper approach for feature selection based on bat algorithm and optimum-path forest. *Expert Systems with Applications*, 41(5), 2250–2258. <https://doi.org/10.1016/j.eswa.2013.09.023>
- Rodrigues, D., Pereira, L. A. M., Papa, J. P., & Weber, S. A. T. (2014). A binary krill herd approach for feature selection. *In 2014 22nd International Conference on Pattern Recognition*, 1407–1412. <https://doi.org/10.1109/ICPR.2014.251>
- Rodrigues, D., Yang, X. S., De Souza, A. N., & Papa, J. P. (2015). Binary flower pollination algorithm and its application to feature selection. *In Recent advances in swarm intelligence and evolutionary computation* (pp. 85–100). Springer. https://doi.org/10.1007/978-3-319-13826-8_5
- Rogers, R. (2009). *The end of the virtual: Digital methods*. Amsterdam University Press.
- Rosa, K. Dela, Shah, R., Lin, B., Gershman, A., & Frederking, R. (2011). Topical clustering of tweets. *In Proceedings of the ACM SIGIR: SWSM*, 63. <https://doi.org/10.1.1.207.4287>
- Rossetti, G., Milli, L., & Cazabet, R. (2019). CDLIB: A python library to extract, compare and evaluate communities from complex networks. *Applied Network Science*, 4(1), 1–26. <https://doi.org/10.1007/s41109-019-0165-9>

- Rozlini, M., Munirah, M. Y., Nawi, N. M., & Wahid, N. (2015). Bat with dempster theory for feature selection: A framework. *Journal of Scientific Research and Development(JSRAD)*, 2(14), 158–165.
- Sabba, S., & Chikhi, S. (2014). A discrete binary version of bat algorithm for multidimensional knapsack problem. *International Journal of Bio-Inspired Computation*, 6(2), 140–152. <https://doi.org/10.1504/IJBIC.2014.060598>
- Sahoo, D., Bhoi, A., & Balabantaray, R. C. (2018). Hybrid approach to abstractive summarization. *Procedia Computer Science*, 132, 1228–1237. <https://doi.org/10.1016/j.procs.2018.05.038>
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, 851–860. <https://doi.org/10.1145/1772690.1772777>
- Saleem, N., Zafar, K., & Sabzwari, A. F. (2019). Enhanced feature subset selection using Niche based bat algorithm. *Computation*, 7(3), 49.
- Salloum, S. A., Al-Emran, M., Abdallah, S., & Shaalan, K. (2017). Analyzing the Arab gulf newspapers using text mining techniques. *International Conf. on Advanced Intelligent Systems and Informatics*, 396–405. https://doi.org/10.1007/978-3-319-64861-3_37
- Salloum, S. A., Al-Emran, M., & Shaalan, K. (2017a). Mining social media text: Extracting knowledge from Facebook. *International Journal of Computing and Digital Systems*, 6(2), 73–81. <https://doi.org/10.12785/ijcds/060203>
- Salloum, S. A., Al-Emran, M., & Shaalan, K. (2017b). Mining text in news channels: A case study from Facebook. *International Journal of Information Technology and Language Studies*, 1(1), 1–9.
- Salloum, S. A., Mhamdi, C., Al-Emran, M., & Shaalan, K. (2017). Analysis and classification of Arabic newspapers' Facebook pages using text mining techniques. *International Journal of Information Technology*, 1(2), 8–17.
- Sankaranarayanan, J., Samet, H., Teitler, B. E., Lieberman, M. D., & Sperling, J. (2009). Twitterstand: News in tweets. In *Proceedings of the 17th Acm Sigspatial International Conference on Advances in Geographic Information Systems*, 42–51. <https://doi.org/10.1145/1653771.1653781>
- Saranyamol, C. S., & Sindhu, L. (2014). A survey on automatic text summarization. *International Journal of Computer Science and Information Technologies*, 5(6), 7889–7893. <https://doi.org/10.22044/JADM.2018.6139.1726>
- Saritha, S. K. (2019). Community Detection Methods in Social Network Analysis. In *Emerging Technologies in Data Mining and Information Security* (pp. 849–858). Springer.

- Satuluri, V., Parthasarathy, S., & Ucar, D. (2010). Markov clustering of protein interaction networks with improved balance and scalability. *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology*, 247–256.
- Sayyadi, H., Hurst, M., & Maykov, A. (2009). Event detection and tracking in social streams. *In Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 3(1).
- Sayyadi, H., & Raschid, L. (2013). A graph analytical approach for topic detection. *ACM Transactions on Internet Technology (TOIT)*, 13(2), 4. <https://doi.org/10.1145/2542214.2542215>
- Seetharaman, A., & Sundersingh, A. C. (2022). Gene selection and classification using correlation feature selection based binary bat algorithm with greedy crossover. *Concurrency and Computation: Practice and Experience*, 34(5), e6718.
- Sehgal, S., Kumar, B., Rampal, L., & Chaliya, A. (2018). A modification to graph based approach for extraction based automatic text summarization. *In In Progress in advanced computing and intelligent engineering* (pp. 373–378). Springer. https://doi.org/10.1007/978-981-10-6875-1_36
- Setiawan, E. I., Susanto, C. P., Santoso, J., Sumpeno, S., & Purnomo, M. H. (2016). Preliminary study of spam profile detection for social media using Markov clustering : Case study on Javanese people. *2016 International Computer Science and Engineering Conf. (ICSEC)*, 1–4. https://doi.org/10.1007/978-3-319-60000-0_5
- Shah, A., Ramasubbu, D., Mathew, D., & Gadoya, M. C. (2020). Hindi history note generation with unsupervised extractive summarization. *In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop*, 44–49.
- Shah, C., & Jivani, A. G. (2017). Multi-document summarization: Study on existing techniques. *In International Conference on Advances in Computing, Communication and Informatics*.
- Sharma, J., & Annappa, B. (2016). Community detection using meta-heuristic approach: Bat algorithm variants. *2016 Ninth International Conf. on Contemporary Computing (IC3)*, 1–7. <https://doi.org/10.1109/IC3.2016.7880209>
- Sheng, J., Liu, C., Chen, L., Wang, B., & Zhang, J. (2020). Research on community detection in complex networks based on internode attraction. *Entropy*, 22(12), 1383. <https://doi.org/10.3390/e22121383>
- Shukla, S., & Naganna, S. (2014). A review on K-means data clustering approach. *International Journal of Information & Computation Technology*, 4(17), 1847–1860.

- Singh, K. N., Devi, H. M., & Mahanta, A. K. (2017). Document representation techniques and their effect on the document clustering and classification: A review. *International Journal of Advanced Research in Computer Science*, 8(5). <https://doi.org/10.26483/IJARCS.V8I5.3822>
- Song, S., Wang, Z., Xu, S., Ni, S., & Xiao, J. (2019). A novel text classification approach based on Word2vec and TextRank keyword extraction. *In 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC)*, 536–543. <https://doi.org/10.1109/DSC.2019.00087>
- Song, W., & Park, S. C. (2009). Genetic algorithm for text clustering based on latent semantic indexing. *Computers & Mathematics with Applications*, 57(11), 1901–1907. <https://doi.org/10.1016/j.camwa.2008.10.010>
- Sörensen, K. (2015). Metaheuristics—The metaphor exposed. *International Transactions in Operational Research*, 22(1), 3–18. <https://doi.org/10.1111/itor.12001>
- Sormanen, N., Rohila, J., Lauk, E., Uskali, T., Jouhki, J., & Penttinen, M. (2016). Chances and challenges of computational data gathering and analysis: The case of issue-attention cycles on Facebook. *Digital Journalism*, 4(1), 55–74. <https://doi.org/10.1080/21670811.2015.1096614>
- Sowmiya, J. S., & Chandrakala, S. (2014). Joint sentiment/topic extraction from text. *In 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, 611–615. <https://doi.org/10.1109/ICACCCT.2014.7019160>
- Srividhya, S., & Mallika, R. (2018). Multimodal feature selection using invasive weed optimization and improved bat for high dimensional imbalanced datasets. *International Journal of Applied Engineering Research*, 13(2), 960–966. <https://doi.org/10.1016/j.neucom.2015.06.068>
- Steinberger, J., & Jezek, K. (2004). Using latent semantic analysis in text summarization and summary evaluation. *In Proc. ISIM*, 4, 93–100.
- Subasic, I., & Berendt, B. (2011). Peddling or creating? Investigating the role of twitter in news reporting. *In European Conference on Information Retrieval*, 207–213. https://doi.org/10.1007/978-3-642-20161-5_21
- Sucharitha, Y., Vijayalata, Y., & Prasad, V. K. (2019). Analysis of early detection of emerging patterns from social media networks: A data mining techniques perspective. *In In Soft Computing and Signal Processing* (pp. 15–25). Springer. https://doi.org/10.1007/978-981-13-3600-3_2
- Sy, E., Jacobs, S. A., Dagnino, A., & Ding, Y. (2016). Graph-based clustering for detecting frequent patterns in event log data. *In 2016 IEEE International Conference on Automation Science and Engineering (CASE)*, 972–977. <https://doi.org/10.1109/COASE.2016.7743509>

- Szell, M., Grauwin, S., & Ratti, C. (2014). Contraction of online response to major events. *PLoS One*, 9(2), 1–9. <https://doi.org/10.1371/journal.pone.0089052>
- Szilágyi, L., Medvés, L., & Szilágyi, S. M. (2010). A modified Markov clustering approach to unsupervised classification of protein sequences. *Neurocomputing*, 73(13–15), 2332–2345. <https://doi.org/10.1016/j.neucom.2010.02.023>
- Szilágyi, L., & Szilágyi, S. M. (2016). A modified two-stage Markov clustering algorithm for large and sparse networks. *Computer Methods and Programs in Biomedicine*, 135, 15–26. <https://doi.org/10.1016/j.cmpb.2016.07.007>
- Szilágyi, S. M., & Szilágyi, L. (2014). A fast hierarchical clustering algorithm for large-scale protein sequence data sets. *Computers in Biology and Medicine*, 48, 94–101. <https://doi.org/10.1016/j.compbiomed.2014.02.016>
- Taghian, S., Nadimi-Shahraki, M. H., & Zamani, H. (2018). Comparative Analysis of Transfer Function-based Binary Metaheuristic Algorithms for Feature Selection. *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1–6.
- Taha, A. M., Chen, S.-D., & Mustapha, A. (2015). Bat algorithm based hybrid filter-wrapper approach. *Advances in Operations Research*, 2015, 1–5. <https://doi.org/10.1155/2015/961494>
- Taha, A. M., Mustapha, A., & Chen, S.-D. (2013). Naive bayes-guided bat algorithm for feature selection. *The Scientific World Journal*, 2013. <https://doi.org/10.1155/2013/325973>
- Tang, D., Zhu, Q., & Yang, F. (2009). A comparative study of clustering algorithms for protein sequences. In *2009 Fourth International on Conference on Bio-Inspired Computing*, 1–5. <https://doi.org/10.1109/BICTA.2009.5338141>
- Tawhid, M. A., & Dsouza, K. B. (2018a). Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems. *Applied Computing and Informatics*, 16(1/2), 117–136. <https://doi.org/10.1016/j.aci.2018.04.001>
- Tawhid, M. A., & Dsouza, K. B. (2018b). Hybrid binary dragonfly enhanced particle swarm optimization algorithm for solving feature selection problems. *Mathematical Foundations of Computing*, 1(2), 181–200. <https://doi.org/10.3934/mfc.2018009>
- Temburnikar, S. D., & Patil, N. N. (2015). Topic detection using BNgram method and sentiment analysis on twitter dataset. *2015 4th International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2015*, 1–6. <https://doi.org/10.1109/ICRITO.2015.7359267>
- Titov, I., & McDonald, R. (2008). Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th International Conference on World Wide Web*, 111–120. <https://doi.org/10.1145/1367497.1367513>

- Too, J., & Mirjalili, S. (2021). A hyper learning binary dragonfly algorithm for feature selection: A covid-19 case study. *Knowledge-Based Systems*, 212, 106553. <https://doi.org/10.1016/j.knosys.2020.106553>
- Top 10 Famous News Channels in the World*. (2018). TenInsider. <http://www.teninsider.com/top-10-famous-news-channels-in-the-world/>
- Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9(1), 1–12. <https://doi.org/10.1038/s41598-019-41695-z>
- Trivedi, I. N., Pradeep, J., Narottam, J., Arvind, K., & Dilip, L. (2016). Novel adaptive whale optimization algorithm for global optimization. *Indian Journal of Science and Technology*, 9(38), 319–326. <https://doi.org/10.17485/ijst>
- Tsagkias, M., De Rijke, M., & Weerkamp, W. (2011). Linking online news and social media. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 565–574. <https://doi.org/10.1145/1935826.1935906>
- Tubishat, M., Abushariah, M. A. M., Idris, N., & Aljarah, I. (2019). Improved whale optimization algorithm for feature selection in Arabic sentiment analysis. *Applied Intelligence*, 49(5), 1688–1707. <https://doi.org/10.1007/s10489-018-1334-8>
- Tubishat, M., Alswaiti, M., Mirjalili, S., Al-Garadi, M. A., & Rana, T. A. (2020). Dynamic butterfly optimization algorithm for feature selection. *IEEE Access*, 8(2020), 194303–194314. <https://doi.org/10.1109/ACCESS.2020.3033757>
- Uğuz, H. (2011). A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. *Knowledge-Based Systems*, 24(7), 1024–1032. <https://doi.org/10.1016/j.knosys.2011.04.014>
- Ullah, S., & Islam, A. B. M. A. Al. (2019). A framework for extractive text summarization using semantic graph based approach. In *Proceedings of the 6th International Conference on Networking, Systems and Security*, 48–56. <https://doi.org/10.1145/3362966.3362971>
- Unankard, S., Li, X., & Sharaf, M. A. (2015). Emerging event detection in social networks with location sensitivity. *World Wide Web*, 18(5), 1393–1417. <https://doi.org/10.1007/s11280-014-0291-3>
- Unankard, S., Li, X., & Sharaf, M. A. (2013). Location-based emerging event detection in social networks. In *Asia-Pacific Web Conference*, 280–291. https://doi.org/10.1007/978-3-642-37401-2_29
- Unankard, S., Li, X., Sharaf, M., Zhong, J., & Li, X. (2014). Predicting elections from social networks based on sub-event detection and sentiment analysis. In *International Conference on Web Information Systems Engineering*, 1–16. https://doi.org/10.1007/978-3-319-11746-1_1

- Unler, A., Murat, A., & Chinnam, R. B. (2011). mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification. *Information Sciences*, 181(20), 4625–4641. <https://doi.org/10.1016/j.ins.2010.05.037>
- Upadhyaya, Y., Sachin, S., Tripathi, A., & Jain, R. (2018). Comparative Analysis of Feature Selection in Epilepsy Seizure Recognition Using Cuckoo, Gravitational Search and Bat Algorithm. *International Journal of Information Systems & Management Science*, 1(1).
- Vadiraja, P., Dengel, A., & Ishimaru, S. (2021). Text summary augmentation for intelligent reading assistant. In *Augmented Humans International Conference 2021 (AHs '21)*, 3. <https://doi.org/10.1145/3458709.3459002>
- Van Dongen, S. M. (2000). *Graph clustering by flow simulation*. University of Utrecht.
- Vavliakis, K. N., Symeonidis, A. L., & Mitkas, P. A. (2013). Event identification in web social media through named entity recognition and topic modeling. *Data & Knowledge Engineering*, 88, 1–24. <https://doi.org/10.1016/j.datak.2013.08.006>
- Verma, P., & Om, H. (2016). Extraction based text summarization methods on user's review data: A comparative study. In *International Conference on Smart Trends for Information Technology and Computer Communications*, 346–354. https://doi.org/10.1007/978-981-10-3433-6_42
- Verma, P., Pal, S., & Om, H. (2019). A comparative analysis on hindi and english extractive text summarization. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 18(3), 1–39. <https://doi.org/10.1145/3308754>
- Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining - An overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.
- Vijayarani, S., & Janani, R. (2016). Text mining: Open source tokenization tools—An analysis. *Advanced Computational Intelligence*, 3(1), 37–47. <https://doi.org/10.5121/acii.2016.3104>
- Wada, H. (2021). Assessing the social media user's credibility rating of shared content, and its utilization in decision making. *Emerging Science Journal*, 5(2), 191–199. <https://doi.org/10.28991/esj-2021-01269>
- Wan, X. (2007). TimedTextRank: Adding the temporal dimension to multi-document summarization. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 867–868. <https://doi.org/10.1145/1277741.1277949>
- Wang, C., Zhou, S., Gao, Y., & Liu, C. (2018). A self-adaptive bat algorithm for the truck and trailer routing problem. *Engineering Computations*, 35(1), 108–135.

<https://doi.org/10.1108/EC-11-2016-0408>

- Wang, F., Li, C., Wang, J., Xu, J., & Li, L. (2015). A two-stage feature selection method for text categorization by using category correlation degree and latent semantic indexing. *Journal of Shanghai Jiaotong University (Science)*, 20(1), 44–50. <https://doi.org/10.1007/s12204-015-1586-y>
- Wang, H., Xu, F., Hu, X., & Ohsawa, Y. (2013). Ideagraph: A graph-based algorithm of mining latent information for human cognition. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 952–957. <https://doi.org/10.1109/SMC.2013.167>
- Wang, Y., Sundaram, H., & Xie, L. (2012). Social event detection with interaction graph modeling. In *Proceedings of the 20th ACM International Conference on Multimedia*, 865–868. <https://doi.org/10.1145/2393347.2396332>
- Wei, Y., Singh, L., Buttler, D., & Gallagher, B. (2018). Using semantic graphs to detect overlapping target events and story lines from newspaper articles. *International Journal of Data Science and Analytics*, 5(1), 41–60. https://doi.org/10.1007/978-3-319-67056-0_19
- Weiler, A., Grossniklaus, M., & Scholl, M. H. (2016). Survey and experimental analysis of event detection techniques for twitter. *The Computer Journal*, 60(3), 329–346. <https://doi.org/10.1111/coin.12017>
- Weiler, A., Grossniklaus, M., & Scholl, M. H. (2014). Event identification and tracking in social media streaming data. In *Proceedings of the Workshops of the EDBT / ICDT 2014 Joint Conference: Multimodal Social Data Management (MSDM); Athens, Greece, March 28 Th, 2014 / Candan, K. Selcuk et Al. (Ed.). - (CEUR Workshop Proceedings; 1133).*, 282–287.
- Weng, J., & Lee, B.-S. (2011). Event detection in Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 401–406.
- Westling, M. (2007). Expanding the public sphere: The impact of Facebook on political communication. *The New Vernacular*, 18, 1–13.
- Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82. <https://doi.org/10.1109/4235.585893>
- Wu, F., & Huberman, B. A. (2004). Finding communities in linear time: A physics approach. *The European Physical Journal B, Condensed Matter Physics*, 38(2), 331–338. <https://doi.org/10.1140/epjb/e2004-00125-x>
- Wu, K., Shi, P., & Pan, D. (2015). An approach to automatic summarization for chinese text based on the combination of spectral clustering and LexRank. In *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 1350–1354. <https://doi.org/10.1109/FSKD.2015.7382140>

- X. Zhao, G. W. M. L. (2016). An improved bat algorithm with variable neighborhood search for global optimization. *2016 IEEE Congress on Evolutionary Computation (CEC)*, 1773–1778. <https://doi.org/10.1109/CEC.2016.7744003>
- Xie, X., Qin, X., Zhou, Q., Zhou, Y., Zhang, T., Janicki, R., & Zhao, W. (2019). A novel test-cost-sensitive attribute reduction approach using the binary bat algorithm. *Knowledge-Based Systems*, 186, 104938. <https://doi.org/10.1016/J.KNOSYS.2019.104938>
- Xiong, C., Li, X., Li, Y., & Liu, G. (2018). Multi-documents summarization based on TextRank and its application in online argumentation platform. *International Journal of Data Warehousing and Mining (IJDWM)*, 14(3), 69–89. <https://doi.org/10.4018/IJDWM.2018070104>
- Xu, H., Yu, S., Chen, J., & Zuo, X. (2018). An improved firefly algorithm for feature selection in classification. *Wireless Personal Communications*, 102(4), 2823–2834. <https://doi.org/10.1007/s11277-018-5309-1>
- Xu, K., Long, T., Qiao, S., Zheng, Y., & Han, N. (2017). A Global Flight Networks Analysis Approach Using Markov Clustering and PageRank. *2017 IEEE International Conference on Big Knowledge (ICBK)*, 96–102.
- Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4), 606–626. <https://doi.org/10.1109/TEVC.2015.2504420>
- Yadav, S. L., & Phogat, M. (2017). A review on bat algorithm. *International Journal of Computer Sciences and Engineering*, 5(7), 39–43. <https://doi.org/10.26438/ijcse/v5i7.3943>
- Yaghoubzadeh, R., Kamel, S. R., Barzgar, H., & Moshajeri San'ati, B. (2021). The Use of the Binary Bat Algorithm in Improving the Accuracy of Breast Cancer Diagnosis. *Multidisciplinary Cancer Investigation*, 5(1), 1–7.
- Yamanaka, T., Tanaka, Y., Hijikata, Y., & Nishida, S. (2010). A supporting system for situation assessment using text data with spatio-temporal information. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 22(6), 691–706. <https://doi.org/10.3156/jsoft.22.691>
- Yan, X., Guo, J., Lan, Y., Xu, J., & Cheng, X. (2015). A probabilistic model for bursty topic discovery in microblogs. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1), 353–359.
- Yang, B., Lu, Y., Zhu, K., Yang, G., Liu, J., & Yin, H. (2017). Feature selection based on modified bat algorithm. *Ieice Transactions on Information and Systems*, 100(8), 1860–1869. <https://doi.org/10.1587/transinf.2016EDP7471>
- Yang, S., Sun, Q., Zhou, H., Gong, Z., Zhou, Y., & Huang, J. (2018). A topic detection method based on keygraph and community partition. *In Proceedings of the 2018*

- International Conference on Computing and Artificial Intelligence*, 30–34. <https://doi.org/10.1145/3194452.3194474>
- Yang, X. S. (2010). A new metaheuristic bat-inspired algorithm. *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*, 65–74. https://doi.org/10.1007/978-3-642-12538-6_6
- Yang, X. S., & He, X. (2013). Bat algorithm: Literature review and applications. *International Journal of Bio-Inspired Computation*, 5(3), 141–149. <https://doi.org/10.1504/IJBIC.2013.055093>
- Yang, Y., Pierce, T., & Carbonell, J. (1998). A study of retrospective and on-line event detection. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 28–36. <https://doi.org/10.1145/290941.290953>
- Yao, J., Wan, X., & Xiao, J. (2017). Recent advances in document summarization. *Knowledge and Information Systems*, 53(2), 297–336. <https://doi.org/10.1007/s10115-017-1042-4>
- Yao, L., Pengzhou, Z., & Chi, Z. (2019). Research on news keyword extraction technology based on TF-IDF and TextRank. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, 452–455. <https://doi.org/10.1109/ICIS46139.2019.8940293>
- Yasaswini, V., & Baskaran, S. (2022). An optimization of feature selection for classification using modified bat algorithm. In M. Bianchini, V. Piuri, S. Das, & R. N. Shaw (Eds.), *Advanced Computing and Intelligent Technologies* (pp. 389–399). Springer Singapore. https://doi.org/10.1007/978-981-16-2164-2_31
- Ye, Z., Hou, X., Zhang, X., & Yang, J. (2018). Application of bat algorithm for texture image classification. *International Journal of Intelligent Systems and Applications*, 10(5), 42–50. <https://doi.org/10.5815/ijisa.2018.05.05>
- Yiğit, F., & Baykan, Ö. K. (2014). A new feature selection method for text categorization based on information gain and particle swarm optimization. In *2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems*, 523–529. <https://doi.org/10.1109/CCIS.2014.7175792>
- Yilmaz, S., & Küçüksille, E. U. (2015). A new modification approach on bat algorithm for solving optimization problems. *Applied Soft Computing*, 28, 259–275. <https://doi.org/10.1016/j.asoc.2014.11.029>
- Yousef, J., Youssef, A., & Keshk, A. (2021). A Hybrid Swarm Intelligence Based Feature Selection Algorithm for High Dimensional Datasets. *IJCI. International Journal of Computers and Information*, 8(1), 67–86.
- Yu, S., & Wu, B. (2018). Exploiting structured news information to improve event detection via dual-level clustering. *2018 IEEE Third International Conf. on Data*

Science in Cyberspace (DSC), 873–880. <https://doi.org/10.1109/dsc.2018.00140>

- Yusof, M. M., & Wahidi, N. (2018). A comparative study of feature selection techniques for bat algorithm in various applications. *In MATEC Web of Conferences*, 150, 1–6. <https://doi.org/10.1051/mateconf/201815006006>
- Zarrinkalam, F., & Bagheri, E. (2017). Event identification in social networks. *Encyclopedia with Semantic Computing and Robotic Intelligence*, 1(01), 21–27. <https://doi.org/10.1142/S2425038416300020>
- Zaware, S., Patadiya, D., Gaikwad, A., Gulhane, S., & Thakare, A. (2021). Text summarization using TF-IDF and TextRank algorithm. *In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1399–1407. <https://doi.org/10.1109/ICOEI51242.2021.9453071>
- Zhang, B., Yuan, H., Sun, L., Shi, J., Ma, Z., & Zhou, L. (2017). A two-stage framework for bat algorithm. *Neural Computing and Applications*, 28(9), 2605–2619. <https://doi.org/10.1007/s00521-016-2192-0>
- Zhang, C., Wang, H., Wang, W., Ma, C., Li, J., Wang, Y., & Xu, F. (2016). EventPanorama: A framework for event detection and visualization from online news. *In 2016 49th Hawaii International Conference on System Sciences (HICSS)*, 3739–3748. <https://doi.org/10.1109/HICSS.2016.466>
- Zhang, C., Wang, H., Wang, W., & Xu, F. (2014). An improved ideagraph algorithm for discovering important rare events. *In 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3290–3295.
- Zhang, C., Wang, H., Xu, F., & Hu, X. (2013). Ideagraph plus: A topic-based algorithm for perceiving unnoticed events. *In 2013 IEEE 13th International Conference on Data Mining Workshops*, 735–741. <https://doi.org/10.1109/ICDMW.2013.16>
- Zhang, J., Chen, W.-N., Zhan, Z.-H., Yu, W.-J., Li, Y.-L., Chen, N., & Zhou, Q. (2012). A survey on algorithm adaptation in evolutionary computation. *Frontiers of Electrical and Electronic Engineering*, 7(1), 16–31. <https://doi.org/10.1007/s11460-012-0192-0>
- Zhang, L., Mistry, K., Lim, C. P., & Neoh, S. C. (2018). Feature selection using firefly optimization for classification and regression models. *Decision Support Systems*, 106, 64–85. <https://doi.org/10.1016/j.dss.2017.12.001>
- Zhang, Y., Szabo, C., & Sheng, Q. Z. (2015). Sense and focus: Towards effective location inference and event detection on Twitter. *In International Conference on Web Information Systems Engineering*, 463–477. https://doi.org/10.1007/978-3-319-26190-4_31
- Zhao, D., & He, Y. (2015). Chaotic binary bat algorithm for analog test point selection. *Analog Integrated Circuits and Signal Processing*, 84(2), 201–214. <https://doi.org/10.1007/s10470-015-0548-5>

- Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. *In European Conference on Information Retrieval*, 338–349. https://doi.org/10.1007/978-3-642-20161-5_34
- Zhou, X., Cao, J., Jin, Z., Xie, F., Su, Y., Chu, D., Cao, X., & Zhang, J. (2015). Real-time news certification system on sina weibo. *In Proceedings of the 24th International Conference on World Wide Web*, 983–988. <https://doi.org/10.1145/2740908.2742571>



APPENDIX A

List of Publications

This thesis is based on the following publications:

- Al-Dyani, W. Z., Yahya, A. H., & Ahmad, F. K. (2018). Challenges of event detection from social media streams. *International Journal of Engineering & Technology*, 7(2.15), 72–75. <https://doi.org/10.14419/ijet.v7i2.15.11217>.
- Al-Dyani, W. Z., Ahmad, F. K., & Kamaruddin, S. S. (2019). Event detection model for Facebook news posts. *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 98–102. <https://doi.org/10.35940/ijitee.A3930.119119>.
- AL-Dyani, W. Z., Ahmad, F. K., & Kamaruddin, S. S. (2020). A Survey on event detection models for text data streams. *Journal of Computer Science*, 16(7 SE-Review Article). <https://doi.org/10.3844/jcssp.2020.916.935>.
- AL-Dyani, W. Z., Ahmad, F. K., & Kamaruddin, S. S. (2022). Binary Bat Algorithm for Text Feature Selection in News Events Detection Model Using Markov Clustering. *Cogent Engineering*, 9(1). <https://doi.org/10.1080/23311916.2021.2010923>.
- Al-Dyani, W. Z., Ahmad, F. K., & Kamaruddin, S. S. (2022). Improvements of bat algorithm for optimal feature selection: A systematic literature review. *Intelligent Data Analysis*, 26, 5–31. <https://doi.org/10.3233/IDA-205455>.