AN IMPROVED ASSOCIATIVE CLASSIFICATION MODEL USING FUZZY PARAMETERIZED SOFT SET-BASED DECISION FOR TEXT CLASSIFICATION

DEDE ROHIDIN

A thesis submitted in
fulfilment of the requirement for the award of the
Doctor of Philosophy in Information Technology

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

MARCH 2023

ACKNOWLEDGEMENT

In the name of Allah, The Most Beneficent, The Most Merciful

All praises be to Allah, the Lord of the universe. May Allah bestow His mercy and grace upon His most beloved Prophet Muhammad PBUH, his family, and his friends. My deepest gratitude to the grace of Allah SWT as with his bounty and mercy, then I can successfully complete this PhD thesis. I would like to take this opportunity to acknowledge the guidance and support given by my supervisor, examiners, colleagues, friends, and family members during this study period.

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Noor Azah binti Samsudin for her continuous support, patience, encouragement, motivation, and enthusiasm. May Allah repay all your kindness, Insha' Allah.

I also would like to thank my parents Amin bin Darnuji and Djodjoh binti Karta (Almh), my mother-in-law Neneng Ulfah bin R. Luqman Djajasubrata, my lovely wife Anne Rahayu, my sweet children Andri-fina and faiz, Nanda and Iman, and Adinda for their prayers, love, and encouragement. Thanks to everybody who has contributed to this achievement either in direct or indirect way.

ABSTRACT

Text classification is applicable in various problem domains, including marketing, security, and biomedical. One of the potential text classifiers is the well-known associative classification approach. However, the existing associative classification approach is still prone to some limitations especially when dealing with the problem with too many rules in text classification problem. Some of the rules generated from the textual data may be irrelevant and redundant, result in low performance in imbalanced and class overlapping data. Therefore, this research has proposed an improved associative classification approach to enhance the performance and efficiency of the text classification by removing the irrelevant rules, reducing redundant rules, and handling the imbalanced and class overlapping issues in the textual data. The proposed associative classification approach consists of three stages: pre-processing, fuzzification and classification. In the classification stage primarily, this study proposed to integrating principles of fuzzy soft set theory into associative rules, therefore referred to as Class-Based Fuzzy Soft Associative (CBFSA) method. The experiments used 20 Newsgroup (balanced data) datasets and Reuter-25178 (imbalanced) to evaluate the proposed model. It shows that CBFSA is successful in removing irrelevant and reducing redundant rules. The CBFSA classifier applies smaller number of rules than Class Based Associative (CBA) and Class Based of Predictive Association Rule (CPAR). The CBFSA is also successful in dealing with imbalanced and class overlap data. The CBFSA performance is higher and faster than CBA and CPAR. Meanwhile, comparative analysis with some other non-associative based classifiers may achieve improved f1-measure between 6% to 32%. The processing time of CBFSA is faster than RNN and CNN but slightly slower than Decision Tree, k-NN, Naïve Bayes, Roccio, Bagging and Boosting.

ABSTRAK

Pengkelasan teks boleh diaplikasi dalam pelbagai domain masalah termasuklah pemasaran, keselamatan dan biomedical. Salah satu pengkelas teks yang diketahui berpotensi adalah pendekatan pengkelas kesatuan. Walau bagaimanapun, pendekatan pengkelas kesatuan sedia ada terdedah kepada beberapa kekangan terutamanya bila melibatkan bilangan peraturan yang banyak dalam masalah pengkelasan teks. Sebahagian peraturan mungkin tidak relevan dan bertindan, menghasilkan keputusan yang rendah dalam data yang tidak seimbang dan pertindanan data kelas. Oleh yang demikian, penyelidikan ini telah mencadangkan pendekatan pengkelasan kesatuan yang telah ditambahbaik untuk meningkatkan kemampuan dan kecekapan melabel teks dengan menyingkirkan peraturan tidak relevan, mengurangkan peraturan bertindan, dan menangani isu ketidakseimbangan dan pertindanan dalam data teks. Pendekatan pengkelasan kesatuan cadangan terdiri daripada tiga peringkat: pra-pemprosesan, pengkaburan and pengkelasan. Dalam peringkat pengkelasan terutamanya, kami telah mencadangkan integrasi prinsip teori *fuzzy soft set* ke dalam peraturan kesatuan dan dirujuk sebagai kaedah *Class-Based Fuzzy Soft Associative (CBFSA)*. Eksperimen dengan set data 20 Newsgroup (data seimbang) dan Reuter-25178 (data tidak seimbang) untuk menilai model cadangan. Ia menunjukkan CBFSA berjaya menyingkirkan data tidak relevan dan mengurangkan peraturan bertindan. CBFSA mengaplikasi bilangan peraturan lebih kecil berbanding *Class Based Associative (CBA)* dan *Class Based of Predictive Association Rule (CPAR)*. CBFSA juga berjaya dalam menghadapi data tidak seimbang dan bertindan kelas. CBFSA mencapai hasil yang lebih baik dan cepat berbanding dengan CBA dan CPAR. Analisa perbandingan dengan pengkelas bukan kesatuan, ukuran-f1 mencapai antara 6% hinggan 32%. Masa pemprosesan CBFSA lebih cepat daripada RNN dan CNN tetapi sedikit perlahan berbanding *Decision Tree, k-NN, Naïve Bayes, Roccio, Bagging and Boosting.*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AC	-	Associative Classification
ACC	-	Accuracy
CAR	-	Class Association Rules
CBA	-	Class Based Associative
CPAR	-	Classification Based on Predictive Association Rules
CBFSA	-	Class Based Fuzzy Soft Set Associative
DF	-	Document Frequency
ECOC	-	Error Correcting Output Coding
FCM	-	Fuzzy C-Means
FN	-	False Negative
FP	-	False Positive
FSSC	-	Fuzzy Soft Set Classifier
HFC	-	Hybrid Fuzzy Classifier
IDF	-	Inverse Document Frequency
KDD	-	Knowledge Discovery from Data
k-NN	-	K Nearest Neighbor
NB	-	Naïve Bayes
SSC	-	Soft Set Classifier
SVM	-	Support Vector Machine
TC	-	Text Classification

TDM	-	Term Document Matrix
TF	-	Term Frequency
TF-IDF-		Term Frequency – Inverse Document Frequency
TN	-	True Negative
TNR	-	True Negative Rate
TP	-	True Positive
TPR	-	True Positive Rate
WWW	-	World Wide Web
U	-	Initial universe.
P(U)	-	The set of all the soft sets over *U*
F(U)	-	The set of all the fuzzy sets over *U*
FPS (U)-		The set of all fs-sets over *U*
E	-	A set of parameters
F_E	-	A soft set over *U*.
$f_E(x)$	-	A soft set approximation function.
μ_x	-	A membership functions of x.
F_A	-	A fuzzy soft set.
$f_A(x)$	-	A fuzzy approximate function.
cF_A	-	A cardinal set of fuzzy soft set F_A .

LIST OF APPENDICES

LIST OF PUBLICATIONS

A fair amount of the materials presented in this thesis have been published in various refereed conference proceedings and journals.

Journals

1. Dede Rohidin, Noor A. Samsudin, Mohd.Ab. Aziz, and Shamsul K. A. Khalid (2019). "Using Fuzzy Association Rule Mining Approach to Identify the Student Skill Data Association", Journal of Theoretical and Applied Information Technology. (Scopus Indexed Q3). Little Lion Scientific Publisher; vol97 no 13, 2019 page 3691-3701. http://www.jatit.org/volumes/Vol97No13/12Vol97No13.pdf
2. Dede Rohidin, Noor A. Samsudin, Mustafa Mat Deris (2022). "Association Rules of Fuzzy Soft Set Based Classification for Text Classification Problem", Journal of King Saud University – Computer and Information Sciences. Elsevier, vol.34, no.3, 2022, pp 801-812 (Scopus Indexed Q1). King Saud University. https://doi.org/10.1016/j.jksuci.2020.03.014

Proceeding

1. Dede Rohidin, Noor A. Samsudin, and Tutut Herawan. "On Mining Association Rules of Real-valued Items using Fuzzy Soft Set". The Second International Conference on Soft Computing and Data Mining. SCDM 2016. Bandung, Indonesia, August 18-20. In: Herawan T., Ghazali R., Nawi N., Deris M. (eds) Advances in Intelligent Systems and Computing, vol 549. Springer, Cham. (pp. 517-528). (Scopus indexed; ISI indexed).

CHAPTER 1

INTRODUCTION

This chapter presents the research background, problem statement, objectives, scope, research significance, and thesis organization.

1.1 Research Background

Textual data classification is one of the challenging tasks in many applications including news categorization (Hadi *et al*., 2018), spam filtering (Subramaniam *et al*., 2010; Aski & Sourati, 2016; Dada *et al*., 2019), business review scrutinization (Fang & Zhan, 2015; Khan & Khalid, 2015; Prananda & Thalib, 2020; Rokade & Aruna Kumari, 2019), sentiment analysis (Pinto & Murari, 2008; Caetano *et al*., 2018; Ansari et al., 2020; Park *et al*., 2021), and medical records mining (Trieschnigg *et al*., 2009; Zhang *et al*., 2018). Indeed, in the presence of the Internet era, the growth of the textual data becomes very rapid (Kowsari *et al*., 2019). Textual data can easily be obtained from various resources such as electronic repositories, chat rooms, online news articles, digital libraries, online forums, email, and blog repositories. Thus, the availability of such resources facilitates collection, accessibility, and distribution of textual data for various purposes, including the possibility to developing of automated classification in text mining application.

One of the well-known automated classification approaches in text mining applications is the associative classification (AC) (Gupta & Lehal, 2009; Sheydaei *et al*., 2015). Besides its application in textual data labeling tasks, AC is also studied for other variety of data such as continuous data (Nakamura, 2012; Subbulakshmi, 2016), image (Deshmukh & Bhosle, 2016), Boolean dataset (Mlakar *et al*., 2017; Park & Lim, 2021). Different from other data classification approaches such as Bayesian classifiers (Jiang *et al*., 2016; Qu *et al*., 2018;), K-Nearest Neighbor classifiers (Samsudin & Bradley, 2010; Jiang *et al*., 2012), Decision Tree classifier (Zhou *et al*., 2017; Karabadji *et al*., 2017; Shaheen *et al*., 2020), Boosting classifier (Bickel *et al*., 2006; Bloehdorn & Hotho, 2006), Bagging Classifier (Geurts, 2000; Shinde *et al*., 2014), Convolutional Neural Network (CNN) (Jaderberg *et al*., 2016; Radhika *et al*., 2018; Alom *et al*., 2018; Junyi *et al*., 2020), Recurrent Neural Network (RNN) (Jing, 2019; El-Moneim *et al*., 2020; Caterini & Chang, 2018), AC is distinguishable in performing two major tasks in labeling the data: association rule mining (ARM) and AC (Chen *et al*., 2014). While the association rule mining task shall find relationships in the data to generating rules, the classification task then shall label the data upon learning the generated rules (Chen *et al*., 2014). The existing studies of AC highlight the benefits of data mining applications in exploring data space for establishing efficient rules and selecting most relevant set of rules for class label decision (Almasi & Saniee Abadeh, 2020). Besides, in some studies AC is also well-known for its capability to present simple output, achieve high accuracy, and maintain rules on the items in the data (Hadi *et al*., 2018).

The size of text data can be categorized into small size text data and large size text data. The small size text data consist of several words in text such as review products by the customer and short messages on social media like WhatsApp, Twitter, Facebook, and Instagram. In comparison, the large size text data has more than a hundred words in the document, such as news articles, short stories, and academic journals. On the other hand, the rules of associative-based text classifier are generated based on the number of words or terms presented in the textual data (Sheydaei *et al*., 2015). For instance, if there are a hundred different words in the textual data, the rules will consider the combination of a hundred words. Using the conventional associative classifiers such as Class-Base Associative (Liu *et al*., 1998), Classification based on Multiple-class Association Rule (CMAR) (Li *et al*., 2001), Classification based on Predictive Association Rules (CPAR) (Yin & Han, 2003) and its existing variations such as Fast Associative Classification Algorithm (FACA) (Hadi *et al*., 2016), Predictability-Based Class Collative Class Association Rules (PCAR) (Song & Lee, 2017), Weighted Classification Based on Association rules (WCBA) (Alwidian *et al*., 2018), Active Pruning Rules (APR) (Rajab, 2019), to label large size textual data is considerably difficult. The large size textual data classification task is challenging because using the existing associative classifiers will lead to high processing time and complexity (Alwidian *et al*., 2018). Besides, resulting large number rules but low accuracy that indicates the rules is inefficient. Consequently, the labeling output will present low classification accuracy. To automate the large size textual data classification, this study has proposed to improve existing AC approach by reducing number of generated rules and time complexity in the decision process, which eventually will increase the textual data labeling accuracy.

1.2 Problem Statement

Several earlier studies for text classification problem using the principles in AC approach can be found in (Liu *et al*., 1998; Yin & Han, 2003; Chen *et al*., 2005; Li *et al*., 2007; Mishra & Vishwakarma, 2017; Sheydaei *et al*., 2015; Yoon & Lee, 2007; Sokhangoee & Rezapour, 2022; Hadi *et al*., 2018). Note that, the existing AC for data labeling tasks consists of three main processes: generate rules, build classifier, and predict class label (Liu *et al*., 1998; Yin & Han, 2003; Thabtah *et al*., 2011; Sheydaei *et al*., 2015). Significant problems in the processes of the existing AC approach may arise when applying to labeling large size textual data, which are presented as follows:

a. In existing AC, too many rules will be generated if the size of data is large (Sheydaei *et al*., 2015; Mlakar *et al*., 2017; Son *et al*., 2018; Li & Sheu, 2021). Consequently, some rules may be redundant (Thabtah, 2007; Sheydaei *et al*., 2015; Son *et al*., 2018; Li & Sheu, 2021), while others may be irrelevant (Son *et al*., 2018; Li & Sheu, 2021). Indeed, such rules are considerably inappropriate for classifier building and class label prediction in the associative classification-based text classification approach.

b. Besides, building the classifiers and predicting the class label using the large number of rules will increase the time complexity. In the classifier building process, the set of rules produced from the rule generation process will be used. Some techniques CBA (Liu *et al*., 1998), FACA (Hadi *et al*., 2016), PCAR (Song & Lee, 2017), WCBA (Alwidian *et al*., 2018), and APR (Rajab, 2019) may be applicable to selecting relevant set of rules from the large number of generated rules using rules ranking and database coverage pruning. In the rules ranking technique, the parameters such as confidence, support, and cardinality were used. In the case of large data sets, there are possibilities where many rules will result in same confidence and support values. In this case, the rules are selected randomly. Such problems may arise when the generated rules have same cardinality or when there is a conflicting cardinality between general and specific rules (Thabtah, 2007; Hadi *et al*., 2016; Alwidian *et al*., 2018). That condition leads to incorrectly selecting rules. Obviously, using the selected set of rules is considerably impractical for classifier building and class label prediction process.

c. The text classification problems are often presented with randomly distributed data, due to imbalanced amount of data available between classes and occurrences of overlapping words with different frequencies in the classes (class overlap) (Li *et al*., 2010; Vuttipittayamongkol *et al*., 2021). In addition, the text classification is naturally an instance of multiclass classification problem (Pintas *et al*., 2021). Using the existing AC, text classification problem seem to be simple with if-then rules representation (Thabtah, 2007; Hadi *et al*., 2016) and manual weight assignment for the features as recommended by experts (Alwidian *et al*., 2018). However, the if-then rules application and such manual weight assignment seems to be practical when dealing with small size textual data. Constructing such rule-based representation and applying manual weight are more challenging when dealing with the multiclass and randomly distributed large size textual data classification problem.

Therefore, this study has proposed an improved AC approach by reducing the rules generated and processing time complexity in classifier building and increasing the class label prediction efficiency that is applicable for text classification problem. The proposed associative classification approach eventually shall increase accuracy of the text labeling results and exhibit lower processing time complexity.

1.3 Research Aim and Objectives

The research aims to improve the existing associative classifier by reducing number of generated rules and time complexity in labeling large size textual data set. To achieve the aim, the following objectives are to be satisfied:

a. To improve the associative classifier-based text classification approach with association rule mining technique and fuzzy soft set concept in generating reduced number of rules for the large size data.

b. To use the reduced number of rules generated in objective a) and implement fuzzy parameterized soft set-based decision in building the classifier.

c. To implement a new weighted parameter of fuzzy soft set and weighted rule techniques in predicting the textual data class membership.

1.4 Research Scope

This study focuses on improving text classification performance by extending an AC approach. The existing AC approach is well-known for various textual data labeling applications (Mishra &Vishwakarma, 2017; Sheydaei *et al*., 2015; Yoon & Lee, 2007; Sokhangoee & Rezapour, 2022; Hadi *et al*., 2018). However, in this study, an improved AC approach was proposed for labelling a large data. The proposed AC approach will be evaluated in experiments with benchmark data sets: the 20 Newsgroups and the Reuter-25178 datasets.

The performance of the proposed AC approach is compared to variations of existing textual data classifiers, including Class-Based Associative (CBA) (Liu *et al*., 1998), and classification based on Predictive Association Rules (CPAR) (Yin & Han, 2003). CBA and CPAR are chosen based on the research showing that CBA and CPAR exist as associative classifiers suitable for Big Data applications (Padillo *et al*., 2019). The methods for Big Data applications are also ideal for large textual datasets. The CBA has excellent accuracy and CPAR has the fastest processing time (Padillo *et al*., 2019). Besides, this research also compares the proposed AC approach against standard textual data classifiers namely Naïve Bayes, k-Nearest Neighbor (k-NN), and decision trees. Besides, bagging and boosting are also chosen in the comparative analysis to represent Ensemble Learning. Finally, this research also compares the performance of the proposed classifier with Deep Learning models: Convolutional Neural Network (CNN) (Alom *et al*., 2018) and Recurrent Neural Network (RNN) (Caterini & Chang, 2018).

1.5 Research Significance

This study establishes a new approach of association rule-based classifier to improve the performance of text classification. The improvement will consider the benefit of integrating a fuzzy soft set theory into associative classifier. The proposed approach produces the optimal classifier and less complexity (redundant rules may be eliminated, irrelevant rules may be removed). The proposed approach able to eliminate the randomly selection rule process, deal with imbalanced and class overlap data. In essence, the proposed approach could increase the performance of text classification problem. Therefore, it can use to classify a text document such as article, news document, the journal, text medical report, and academic report. The proposed approach able to handle large size textual data.

1.6 Thesis Organization

The discussion in this thesis is divided into five different chapters as follows:

Chapter 1 Describes the research background, problem statement, research aim, research objective, research scope, significance, and thesis organization.

Chapter 2 Presents a taxonomy of text classification including application of text classification, review of text classification method, Design of text classification and performance evaluation.

Chapter 3 Explains the proposed association-rule-based classifier that is applicable to text classification problem. The explanation includes association rule, fuzzy set-based decision-making problem, extended fuzzy soft set association rule, fuzzy soft set approach for associative classifier and the experiment methodology.

Chapter 4 Presents the performance of proposed method in the text classification problem against existing classifiers are presented: CBA, CPAR, Naïve Bayes, K-NN, Decision Tree, Class-Based Associative, Bagging, Boosting, Convolutional Neural Network, and Recurrent Neural Network.

Chapter 5 Presents the conclusions of the study and recommends some future works.

CHAPTER 2

A TAXONOMY OF TEXT CLASSIFICATION

This section discusses a text classification taxonomy. The discussions include text classification application, notion of text classification, review of text classification method, text classification design, and metric performance for text classification.

2.1 Applications of Text Classification

Textual data classification is one of important tasks in many applications including business, medical, politics, and spam filtering. Some of the applications are reviewed here to justify the importance to automating such classification tasks.

Business: A text classification application to track customer sentiment about the company, this application is called sentiment analysis. This application can inform customers' negative or positive impressions through online customer review analysis from social media, interaction with Call Center, or other data sources (Fang & Zhan, 2015; Khan & Khalid, 2015; Prananda & Thalib, 2020; Rokade & Aruna Kumari, 2019).

Medical: In the medical field, text classification is used to diagnose a patient's disease and medical conditions based on data reported from data mining results. Data mining is conducted on medical records, medicine receipts that patients and laboratory documents have consumed (Trieschnigg *et al*., 2009; Zhang *et al*., 2018; Banerjee *et al*., 2018).

Politics: During presidential election campaigns, thriving teams of candidates often use social media to rally support from voters. Voter support for a candidate can be analyzed using sentiment analysis. From comments on social media made by potential voters to a candidate, sentiment analysis can predict the number of voters who support him. Aside from presidential elections, the sentiment analysis is also used to see how much support for political parties or government policies (Caetano *et al*., 2018; Ansari *et al*., 2020; Park *et al*., 2021; Ankit & Saleena, 2018; Abdi *et al*., 2019).

Spam Filtering: Today, most people have an email account because email is an effective and efficient way of communicating. However, email owners often receive spam. One of the problems that spam brings is that it may be used as an entry point of virus. In this case, text classification can be implemented to construct active spam filtering (Subramaniam *et al*., 2010; Aski & Sourati, 2016; Dada *et al*., 2019).

2.2 Text Classification

Textual data set is presented with a collection of sentences, and a sentence is a collection of words. Therefore, textual data can be regarded as a collection of words. Formally, a textual data set can be written as a document set D = $\{d_1, d_2, \ldots, d_n\}$ where d_i refers to a textual data point (i.e., document, text segment), it has sentences, so that each sentence includes w_s words/terms with l_w letters. Each textual data point is labelled with a class value from a set of k different discrete value indices. Therefore, text classification can be defined as the process of assigning unseen textual data using predetermined class label. The feature of the textual data set is the terms/words that appear in textual documents. Considering that the number of words in a document is very large, the textual document needs to be converted into a vector space matrix (Handaga & Deris, 2013; Qureshi *et al*., 2015; Pintas *et al*., 2021). In this way, a document can be expressed as a vector word (word feature).

There are existing works on the use of machine learning techniques to automating the textual data classification tasks. Some common textual data classifiers are associative classifier (Liu *et al*., 1998; Li *et al*., 2001; Hadi *et al*., 2016; Sokhangoee & Rezapour, 2022), Naïve Bayes (Kolluri & Razia, 2020), K-Nears Neighbor (Heng *et al*., 2012; Bilal *et al*., 2016), Decision Tree (Wenlong Li & Xing,

REFERENCES

Abdelhamid, N., & Thabtah, F. (2014). Associative Classification Approaches: Review and Comparison. *Journal of Information and Knowledge Management*, *13*(3), 1–30. https://doi.org/10.1142/S0219649214500270

Abdi, A., Mariyam, S., Hasan, S., & Piran, J. (2019). Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. *Information Processing and Management*, *56*(4), 1245–1259. https://doi.org/10.1016/j.ipm.2019.02.018

Abdul, J., Subhani, O., & Varlamis, I. (2021). Fake news detection : A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, *1*(December 2020). https://doi.org/10.1016/j.jjimei.2020.100007

Abe, N., & Mamitsuka, H. (1998). Query learning strategies using boosting and bagging. *Proceedings of the 25th International Conference on Machine Learning*, *388*(January 1998), 1–9. http://webia.lip6.fr/~amini/RelatedWorks/Abe98.pdf

Agrawal, R. (1993). Mining Association Rules between Sets of Items in Large Databases. *Proceedings of ACM SIGMOD International Conference on Management of Data*, *May*, 207–216.

Agrawal, R., & Srikant, R. (1994). Fast Algorithm For Mining Association Rules. In *proceeding of the 20th International Conference on Very Large Data Bases*.

Alayba, A. M., & Palade, V. (2021). Leveraging Arabic sentiment classification using an enhanced CNN-LSTM approach and effective Arabic text preparation. *Journal of King Saud University - Computer and Information Sciences*, *xxxx*. https://doi.org/10.1016/j.jksuci.2021.12.004

Albitar, S., Espinasse, B., & Fournier, S. (2012). Towards a supervised rocchio-based semantic classification of web pages. *Frontiers in Artificial Intelligence and Applications*, *243*(May 2014), 460–469. https://doi.org/10.3233/978-1-61499-105-2-460

Alcantud, J. C. R. (2016). A novel algorithm for fuzzy soft set based decision making from multiobserver input parameter data set. *Information Fusion*, *29*(May), 142–148. https://doi.org/10.1016/j.inffus.2015.08.007

Alkhazaleh, S. (2015). The Multi-Interval-Valued Fuzzy Soft Set with Application in Decision Making. *Applied Mathematics*, *July*, 1250–1262.

Almasi, M., & Saniee Abadeh, M. (2020). CARs-Lands: An associative classifier for large-scale datasets. *Pattern Recognition*, *100*. https://doi.org/10.1016/j.patcog.2019.107128

Alom, M. Z., Hasan, M., & Yakopcic, C. (2018). Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *Computer Vision and Pattern Recognition*, *arxiv.org*. https://arxiv.org/abs/1802.06955

Alwidian, J., Hammo, B. H., & Obeid, N. (2016). Enhanced CBA algorithm Based on Apriori optimization and statistical ranking measure. *Proceedings of the 28th International Business Information Management Association Conference - Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth*, *March 2017*, 4291–4306.

Alwidian, J., Hammo, B. H., & Obeid, N. (2018). WCBA: Weighted classification based on association rules algorithm for breast cancer disease. *Applied Soft Computing Journal*, *62*, 536–549. https://doi.org/10.1016/j.asoc.2017.11.013

Amado, N., Gama, J., & Silva, F. (2001). Parallel implementation of decision tree learning algorithms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2258 LNAI*(December), 6–13. https://doi.org/10.1007/3-540-45329-6_4

Ankit, & Saleena, N. (2018). An Ensemble Classification System for Twitter Sentiment Analysis. *Procedia Computer Science*, *132*(Iccids), 937–946. https://doi.org/10.1016/j.procs.2018.05.109

Ansari, M. Z., Aziz, M. B., Siddiqui, M. O., Mehra, H., & Singh, K. P. (2020). Analysis of Political Sentiment Orientations on Twitter. *Procedia Computer Science*, *167*, 1821–1828. https://doi.org/10.1016/j.procs.2020.03.201

Arar, Ö. F., & Ayan, K. (2017). A feature dependent Naive Bayes approach and its application to the software defect prediction problem. *Applied Soft Computing Journal*, *59*, 197–209. https://doi.org/10.1016/j.asoc.2017.05.043

Asha, T., Natarajan, S., & Murthy, K. (2011). A Study of Associative Classifiers with

Different Rule Evaluation Measures for Tuberculosis Prediction. *International Journal of Computer Applications*, 11–16. http://www.ijcaonline.org/specialissues/ait/number3/2839-220

Aski, A. S., & Sourati, N. K. (2016). Proposed efficient algorithm to filter spam using machine learning techniques. *Pacific Science Review A: Natural Science and Engineering*, *18*(2), 145–149. https://doi.org/10.1016/j.psra.2016.09.017

Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., Moradzadeh, N., Chapman, B., Amrhein, T., Mong, D., Rubin, D. L., Farri, O., & Lungren, M. P. (2018). Artificial Intelligence In Medicine Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. *Artificial Intelligence In Medicine*, *November*, 0–1. https://doi.org/10.1016/j.artmed.2018.11.004

Baralis, E., & Garza, P. (2002). A lazy approach to pruning classification rules. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 35–42. https://doi.org/10.1109/icdm.2002.1183883

Bauer, E., & Kohavi, R. (1999). Empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning*, *36*(1), 105–139. https://doi.org/10.1023/a:1007515423169

Bickel, P. J., Ritov, Y., & Zakai, A. (2006). Some theory for generalized boosting algorithms. *Journal of Machine Learning Research*, *7*(June 2014), 705–732.

Bilal, M., Israr, H., Shahid, M., & Khan, A. (2016). Sentiment classification of Roman-Urdu opinions using Naı ¨ ve Bayesian , Decision Tree and KNN classification techniques. *Journal of King Saud University - Computer and Information Sciences*, *28*(3), 330–344. https://doi.org/10.1016/j.jksuci.2015.11.003

Bloehdorn, S., & Hotho, A. (2006). Boosting for text classification with semantic features. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3932 LNAI*, 149–166. https://doi.org/10.1007/11899402_10

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://transacl.org/ojs/index.php/tacl/article/view/999

Breiman, L. (1996). Bagging predictions. *Machine Learning*, *24*(2), 123–140.

Breiman, Leo. (2001). Random forests. *Random Forests*, *Machines L*(45), 5–32. https://doi.org/10.1201/9780429469275-8

Caetano, J. A., Lima, H. S., Santos, M. F., & Marques-Neto, H. T. (2018). Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election. *Journal of Internet Services and Applications*, *9*(1). https://doi.org/10.1186/s13174-018-0089-0

Çağman, N., Çıtak, F., & Enginoğlu, S. (2011). FP-Soft Set Theory and Its Applications. *Annals of Fuzzy Mathematics and Informatics*, *2*(2), 219–226.

Çağman, N., & Deli, I. (2012). Products of FP-soft sets and their applications. *Hacettepe Journal of Mathematics and Statistics*, *41*(3), 365–374.

Cai, J., Durkin, J., & Cai, Q. (2005). CC4 . 5 : cost-sensitive decision tree pruning. *WIT Transactions on Information and Communication Technologies*, *35*, 239–245.

Caterini, A. L., & Chang, D. E. (2018). Recurrent neural networks. *SpringerBriefs in Computer Science*, *0*(9783319753034), 59–79. https://doi.org/10.1007/978-3-319-75304-1_5

Chen, F., Wang, Y., Li, M., Wu, H., & Tian, J. (2014). Principal association mining: An efficient classification approach. *Knowledge-Based Systems*, *67*, 16–25. https://doi.org/10.1016/j.knosys.2014.06.013

Chen, Jian, Yin, J., Zhang, J., & Huang, J. (2005). Associative classification in text categorization. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3644 LNCS*(60205007), 1035–1044. https://doi.org/10.1007/11538059_107

Chen, Junyi, Yan, S., & Wong, K. C. (2020). Verbal aggression detection on Twitter comments: convolutional neural network for short-text sentiment analysis. *Neural Computing and Applications*, *32*(15), 10809–10818. https://doi.org/10.1007/s00521-018-3442-0

Chen, Y. L., & Weng, C. H. (2008). Mining association rules from imprecise ordinal data. *Fuzzy Sets and Systems*, *159*(4), 460–474. https://doi.org/10.1016/j.fss.2007.10.005

Da Silva, N. F. F., Hruschka, E. R., & Hruschka, E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, *66*, 170–179. https://doi.org/10.1016/j.dss.2014.07.003

Dada, E. G., Bassi, J. S., Chiroma, H., Abdulhamid, S. M., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: review, approaches and open research problems. *Heliyon*, *5*(6). https://doi.org/10.1016/

j.heliyon.2019.e01802

Danjuma, S., Herawan, T., Ismail, M. A., Chiroma, H., Abubakar, A. I., & Zeki, A. M. (2017). A Review on Soft Set-Based Parameter Reduction and Decision Making. *IEEE Access*, *5*, 4671–4689. https://doi.org/10.1109/ACCESS.2017.2682231

Das P., K., Borgohain, R., & Pradesh, A. (2010). *An application of fuzzy soft set in medical diagnosis using fuzzy arithmetic operation on fuzzy number Created by Neevia Personal Converter trial version*. *05*, 107–116.

Deli, I., & Çağman, N. (2015). Intuitionistic fuzzy parameterized soft set theory and its decision making. *Applied Soft Computing Journal*, *28*, 109–113. https://doi.org/10.1016/j.asoc.2014.11.053

Deng, H., Runger, G., Tuv, E., & Bannister, W. (2014). CBC: An associative classifier with a small number of rules. *Decision Support Systems*, *59*(1), 163–170. https://doi.org/10.1016/j.dss.2013.11.004

Deng, J., Cheng, L., & Wang, Z. (2021). Computer Speech & Language Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classi fi cation. *Computer Speech & Language*, *68*, 101182. https://doi.org/10.1016/j.csl.2020.101182

Deshmukh, J., & Bhosle, U. (2016). Image Mining Using Association Rule for Medical Image Dataset. *Procedia Computer Science*, *85*(Cms), 117–124. https://doi.org/10.1016/j.procs.2016.05.196

Dhanabal, S., & Chandramathi, S. (2011). A Review of various k-Nearest Neighbor Query Processing Techniques. *International Journal of Computer Applications*, *31*(7), 14–22.

El-Moneim, S. A., Nassar, M. A., Dessouky, M. I., Ismail, N. A., El-Fishawy, A. S., & Abd El-Samie, F. E. (2020). Text-independent speaker recognition using LSTM-RNN and speech enhancement. *Multimedia Tools and Applications*, *79*(33–34), 24013–24028. https://doi.org/10.1007/s11042-019-08293-7

Elnagar, A., Al-debsi, R., & Einea, O. (2020). Arabic text classification using deep learning models. *Information Processing and Management*, *57*(1), 102121. https://doi.org/10.1016/j.ipm.2019.102121

Elsayed, S.A.M., Rajasekaran, S., Ammar, R.A. (2012). AC-CS: An Immune-Inspired Associative Classification Algorithm. In: Coello Coello, C.A., Greensmith, J., Krasnogor, N., Liò, P., Nicosia, G., Pavone, M. (eds) Artificial Immune Systems.

ICARIS 2012. Lecture Notes in Computer Science, vol 7597. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33757-4_11

Fang, W., Luo, H., Xu, S., Love, P. E. D., Lu, Z., & Ye, C. (2020). Advanced Engineering Informatics Automated text classification of near-misses from safety reports : An improved deep learning approach. *Advanced Engineering Informatics*, *44*(March 2019), 101060. https://doi.org/10.1016/j.aei.2020.101060

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, *2*(1). https://doi.org/10.1186/s40537-015-0015-2

Feng, F., Li, Y., Leoreanu, V., & Fotea. (2010). Application of level soft sets in decision making based on interval-valued fuzzy soft sets. *Computers & Mathematics with Applications*, *60*(6), 1756–1767. https://doi.org/https://doi.org/ 10.1016/j.camwa.2010.07.006

Feng, F., Cho, J., Pedrycz, W., Fujita, H., & Herawan, T. (2016). Soft set based association rule mining. *Knowledge-Based Systems*, *111*(August), 268–282. https://doi.org/10.1016/j.knosys.2016.08.020

Fersini, E., Messina, E., & Pozzi, F. A. (2014). Sentiment analysis: Bayesian Ensemble Learning. *Decision Support Systems*, *68*, 26–38. https://doi.org/10.1016/ j.dss.2014.10.004

Geurts, P. (2000). Some enhancements of decision tree bagging. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *1910*, 136–147. https://doi.org/10.1007/3-540-45372-5_14

Gulcehre, C., Cho, K., Pascanu, R., & Bengio, Y. (2014). Learned-norm pooling for deep neural networks. *Lecture Notes in Computer Science*, *8724*, 530–546. http://arxiv.org/abs/1311.1780

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, *1*(1), 60–76. https://doi.org/10.4304/jetwi.1.1.60-76

Hadi, W., Aburub, F., & Alhawari, S. (2016). A new fast associative classification algorithm for detecting phishing websites. *Applied Soft Computing Journal*, *48*, 729–734. https://doi.org/10.1016/j.asoc.2016.08.005

Hadi, W., Al-Radaideh, Q. A., & Alhawari, S. (2018). Integrating associative rule-based classification with Naïve Bayes for text classification. *Applied Soft*

Computing Journal, *69*, 344–356. https://doi.org/10.1016/j.asoc.2018.04.056

Han, J., Pei, J., Yin, Y., & Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, *8*(1), 53–87. https://doi.org/10.1023/B:DAMI.0000005258.31418.83

Handaga, B., & Deris, M. M. (2012). *A New classification technique based on hybrid fuzzy soft set and supervised fuzzy c-means*. Universiti Tun Hussein Onn Malaysia.

Handaga, B., Herawan, T., & Deris, M. M. (2012). An Algorithm for Classifying Numerical FSSC : *International Journal of Fuzzy System Applications*, *3*(December), 29–46.

Happawana, K. A., & Diamond, B. J. (2021). Association rule learning in neuropsychological data analysis for Alzheimer's disease. *Journal of Neuropsychology*, 1–15. https://doi.org/10.1111/jnp.12252

Hasnain, M., Ghani, I., Jeong, S. R., & Ali, A. (2022). Ensemble learning models for classification and selection of web services: A review. *Computer Systems Science and Engineering*, *40*(1), 327–339. https://doi.org/10.32604/CSSE.2022.018300

Heng, C., Hong, L., Rajkumar, R., & Isa, D. (2012). Expert Systems with Applications A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine. *Expert Systems With Applications*, *39*(15), 11880–11888. https://doi.org/10.1016/j.eswa.2012.02.068

Herawan, T., & Deris, M. M. (2010). Soft decision making for patients suspected influenza. *International Conference on Computational Science and Its Applications*, 405–418.

Herawan, T., & Deris, M. M. (2011). A soft set approach for association rules mining. In *Knowledge-Based Systems* (Vol. 24, Issue 1, pp. 186–195). https://doi.org/10.1016/j.knosys.2010.08.005

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jaderberg, M., Simonyan, K., Vedaldi, A., & Zisserman, A. (2016). Reading Text in the Wild with Convolutional Neural Networks. *International Journal of Computer Vision*, *116*(1), 1–20. https://doi.org/10.1007/s11263-015-0823-z

Jafreezal, C. C., Izzatdin, J., Aziz, A., Hilmi, M., & William, H. (2019). Algorithms for frequent itemset mining : a literature review. *Artificial Intelligence Review*, *52*(4), 2603–2621. https://doi.org/10.1007/s10462-018-9629-z

Jiang, L., Li, C., Wang, S., & Zhang, L. (2016). Deep feature weighting for naive Bayes and its application to text classification. *Engineering Applications of Artificial Intelligence*, *52*, 26–39. https://doi.org/10.1016/j.engappai.2016.02.002

Jiang, S., Pang, G., Wu, M., & Kuang, L. (2012). An improved K-nearest-neighbor algorithm for text categorization. *Expert Systems with Applications*, *39*(1), 1503–1509. https://doi.org/10.1016/j.eswa.2011.08.040

Jing, R. (2019). A Self-attention Based LSTM Network for Text Classification. *Journal of Physics: Conference Series*, *1207*(1). https://doi.org/10.1088/1742-6596/1207/1/012008

Jing, W., Huang, L., Luo, Y., Xu, W., & Yao, Y. (2006). An algorithm for privacy-preserving quantitative association rules mining. *Proceedings - 2nd IEEE International Symposium on Dependable, Autonomic and Secure Computing, DASC 2006*, 315–321. https://doi.org/10.1109/DASC.2006.18

Joyce, James (2003), "Bayes' Theorem", in Zalta, Edward N. (ed.), The Stanford Encyclopedia of Philosophy (Spring 2019 ed.), Metaphysics Research Lab, Stanford University, retrieved 2020-01-1

J. Rauch, Formal framework for data mining with association rules and domain knowledge–overview of an approach, Fundam. Inf. 137 (2) (2015) 171–217.

Karabadji, N. E. I., Seridi, H., Bousetouane, F., Dhifli, W., & Aridhi, S. (2017). An evolutionary scheme for decision tree construction. *Knowledge-Based Systems*, *119*, 166–177. https://doi.org/10.1016/j.knosys.2016.12.011

Kataria, A., & Singh, M. D. (2013). A Review of Data Classification Using K-Nearest Neighbour Algorithm. *International Journal of Emerging Technology and Advanced Engineering*, *3*(6), 354–360.

Khan, M. T., & Khalid, S. (2015). Sentiment Analysis for Health Care. *International Journal of Privacy and Health Information Management*, *3*(2), 78–91. https://doi.org/10.4018/ijphim.2015070105

Kim, Y. H., Hahn, S. Y., & Zhang, B. T. (2000). Text filtering by boosting Naive Bayes classifiers. *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, 168–175. https://doi.org/10.1145/345508.345572

Kolluri, J., & Razia, S. (2020). Text classification using Naïve Bayes classifier. *Materials Today: Proceedings*, *xxxx*. https://doi.org/10.1016/j.matpr.2020.10.058

Kong, Z., Wang, L., & Wu, Z. (2011). Application of fuzzy soft set in decision making problems based on grey theory. *Journal of Computational and Applied Mathematics*, *236*(6), 1521–1530. https://doi.org/10.1016/j.cam.2011.09.016

Kowsari, K., Heidarysafa, M., Brown, D. E., Meimandi, K. J., & Barnes, L. E. (2018). RMDL: Random multimodel deep learning for classification. *ACM International Conference Proceeding Series*, 19–28. https://doi.org/10.1145/3206098.3206111

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information (Switzerland)*, *10*(4), 1–68. https://doi.org/10.3390/info10040150

Kundu, G., Islam, M. M., Munir, S., & Bari, M. F. (2008). ACN: An associative classifier with negative rules. *Proceedings - 2008 IEEE 11th International Conference on Computational Science and Engineering, CSE 2008*, 369–375. https://doi.org/10.1109/CSE.2008.48

Li, B., Sugandh, N., Garcia, E. V., & Ram, A. (2007). Adapting associative classification to text categorization. *DocEng'07: Proceedings of the 2007 ACM Symposium on Document Engineering*, *August 2007*, 205–208. https://doi.org/10.1145/1284420.1284470

Li, H., & Sheu, P. C. Y. (2021). A scalable association rule learning heuristic for large datasets. In *Journal of Big Data* (Vol. 8, Issue 1). Springer International Publishing. https://doi.org/10.1186/s40537-021-00473-3

Li, Wenlong, & Xing, C. (2010). Parallel decision tree algorithm based on combination. *Proceedings - 2010 International Forum on Information Technology and Applications, IFITA 2010*, *1*, 99–101. https://doi.org/10.1109/IFITA.2010.115

Li, Wenmin., Han, J., & Pei, J. (2001). CMAR: accurate and efficient classification based on multiple class-association rules. *IEEE International Conference on Data Mining (ICDM)*, 369–376.

Li, X., Qin, D., & Yu, C. (2008). ACCF: Associative classification based on closed frequent itemsets. *Proceedings - 5th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2008*, *2*, 380–384. https://doi.org/10.1109/FSKD.2008.396

Li, Y., Sun, G., & Zhu, Y. (2010). Data imbalance problem in text classification. *Proceedings - 3rd International Symposium on Information Processing, ISIP 2010*, 301–305. https://doi.org/10.1109/ISIP.2010.47

Liang, D., & Yi, B. (2021). Two-stage three-way enhanced technique for ensemble learning in inclusive policy text classification. *Information Sciences*, *547*, 271–288. https://doi.org/10.1016/j.ins.2020.08.051

Liu, B., Hsu, W., & Ma, Y. (1998). Integrating Classification and Association Rule Mining. *KDD-98*, *New York*, *NY*, *Aug*, 80–86.

Liu, Y., Loh, H. T., & Sun, A. (2009). Imbalanced text classification: A term weighting approach. *Expert Systems with Applications*, *36*(1), 690–701. https://doi.org/10.1016/j.eswa.2007.10.042

Mai, T., Vo, B., & Nguyen, L. T. T. (2017). A lattice-based approach for mining high utility association rules. Information Sciences, 399, pp.81–97. doi:10.1016/j.ins.2017.02.058

Maji, P. K., Biswas, R., & Roy, A. R. (2001). Fuzzy Soft Sets. *Journal of Fuzzy Mathematics*, *9*(3), 589–602.

Maji, P. K., Roy, A. R., & Biswas, R. (2002). An application of soft sets in a decision making problem. *Computers and Mathematics with Applications*, *44*(8–9), 1077–1083. https://doi.org/10.1016/S0898-1221(02)00216-X

Manning, C. ., Raghavan, P., & Schutze, H. (2009). *An Introduction to information retrieval(online edition)*. Cambridge University Press

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12.

Mishra, M., & Vishwakarma, S. (2017). Text Classification based on Association Rule Mining Technique. *International Journal of Computer Applications*, *169*(10), 46–50. https://doi.org/10.5120/ijca2017914905

Mlakar, U., Zorman, M., & Fister, I. (2017). Modified binary cuckoo search for association rule mining. *Journal of Intelligent and Fuzzy Systems*, *32*(6), 4319–4330. https://doi.org/10.3233/JIFS-16963

Molodtsov, D. (1999). Soft Set Theory: First Results. *Computers & Mathematics with Applications*, *37*, 19–31.

Myaeng, S. H., Han, K. S., & Rim, H. C. (2006a). Some effective techniques for naive bayes text classification. *IEEE Transactions on Knowledge and Data Engineering*, *18*(11), 1457–1466. https://doi.org/10.1109/TKDE.2006.180

Myaeng, S. H., Han, K. S., & Rim, H. C. (2006b). Some effective techniques for naive bayes text classification. *IEEE Transactions on Knowledge and Data Engineering*, *18*(11), 1457–1466. https://doi.org/10.1109/TKDE.2006.180

Nakamura, Y. R. . (2012). A Binary Bat Algorithm for Feature Selection. *In XXV SIBGRAPI Conference on Graphics, Patterns and Images*, 291–297.

Nasef, A. A., & Jafari, S. (2018). *Another Application of Fuzzy Soft Sets in Real Life Problems*. *July*.

Nguyen.L.T,Vo.B, Hong, Thanh.H.C (2012), Classification based on association rules: A lattice-based approach, Expert Syst. Appl. 39 (13) (2012) 357–366

N.S. Nithya, K. Duraiswamy(2015), Correlated gain ratio based fuzzy weighted association rule mining classifier for diagnosis health care data, J. Intell. Fuzzy Syst. 29 (4),1453–1464.

Onan, A., Korukoğlu, S., & Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications*, *57*, 232–247. https://doi.org/10.1016/j.eswa.2016.03.045

Öztürk, M. A., & Inan, E. (2012). Fuzzy soft subnear-rings and (∈, ∈ ∨ q) -fuzzy soft subnear-rings. In *Computers and Mathematics with Applications* (Vol. 63, Issue 3, pp. 617–628). https://doi.org/10.1016/j.camwa.2011.11.008

Padillo, F., Luna, J. M., & Ventura, S. (2019). Evaluating associative classification algorithms for Big Data. *Big Data Analytics*, *4*(1), 1–27. https://doi.org/10.1186/s41044-018-0039-7

Park, H. Y., & Lim, D. J. (2021). A design failure pre-alarming system using score- and vote-based associative classification. *Expert Systems with Applications*, *164*(May 2020), 113950. https://doi.org/10.1016/j.eswa.2020.113950

Park, S., Strover, S., Choi, J., & Schnell, M. (2021). Mind games: A temporal sentiment analysis of the political messages of the Internet Research Agency on Facebook and Twitter. *New Media & Society*, 146144482110143. https://doi.org/10.1177/14614448211014355

Pennington, J., Richard, S., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceeding of the 2014 Conference on Empirical Method in Natural Language Processing (EMNLP)*, 1532–1543.

Pintas, J. T., Fernandes, L. A. F., & Garcia, A. C. B. (2021). Feature selection methods for text classification: a systematic literature review. In *Artificial Intelligence Review* (Vol. 54, Issue 8). Springer Netherlands. https://doi.org/10.1007/s10462-021-09970-6

Pinto, J. P., & Murari, V. (2008). Real Time Sentiment Analysis of Political Twitter Data Using Machine Learning Approach. *International Research Journal of*

Engineering and Technology, 4124. www.irjet.net

Prananda, A. R., & Thalib, I. (2020). Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion. *Journal of Information Systems Engineering and Business Intelligence*, *6*(1), 1. https://doi.org/10.20473/jisebi.6.1.1-8

Qin, H., Ma, X., Zain, J. M., & Herawan, T. (2012). A novel soft set approach in selecting clustering attribute. *Knowledge-Based Systems*, *36*, 139–145. https://doi.org/10.1016/j.knosys.2012.06.001

Qu, Z., Song, X., Zheng, S., Wang, X., Song, X., & Li, Z. (2018). Improved Bayes Method Based on TF-IDF Feature and Grade Factor Feature for Chinese Information Classification. *Proceedings - 2018 IEEE International Conference on Big Data and Smart Computing, BigComp 2018*, 677–680. https://doi.org/10.1109/BigComp.2018.00124

Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, *27*(3), 221–234. https://doi.org/https://doi.org/10.1016/S0020-7373(87)80053-6.

Qureshi, M. N., Section, E. E., Faisal, H., Aldheleai, H., & Tamandani, Y. K. (2015). An Improved Documents Classification Technique Using Association Rules Mining. *IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 460–465.

Radhika, K., Bindu, K. R., & Parameswaran, L. (2018). A Text Classification Model Using Convolution Neural Network and Recurrent Neural Network. *International Journal of Pure and Applied Mathematics*, *119*(15), 1549–1554.

Rajab, K. D. (2019). New Associative Classification Method Based on Rule Pruning for Classification of Datasets. *IEEE Access*, *7*, 157783–157795. https://doi.org/10.1109/ACCESS.2019.2950374

Rokade, P. P., & Aruna Kumari, D. (2019). Business intelligence analytics using sentiment analysis-a survey. *International Journal of Electrical and Computer Engineering*, *9*(1), 613–620. https://doi.org/10.11591/ijece.v9i1.pp613-620

Rotjanasom, C., Inbunleu, C., & Suebsan, P. (2021). Applications of Fuzzy Parameterized Relative Soft Sets in Decision-Making Problems. *IAENG International Journal of Applied Mathematics*, *51*(3), 0–5.

Roy, A. R., & Maji, P. K. (2007). *A fuzzy soft set theoretic approach to decision making problems*. *203*, 412–418. https://doi.org/10.1016/j.cam.2006.04.008

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary*

Reviews: Data Mining and Knowledge Discovery, *8*(4), 1–18. https://doi.org/10.1002/widm.1249

Sahgal, D., & Parida, M. (2014). Object recognition using Gabor wavelet features with various classification techniques. *Advances in Intelligent Systems and Computing*, *258*, 793–804. https://doi.org/10.1007/978-81-322-1771-8_69

Sahgal, D., & Ramesh, A. (2015). On Road Vehicle detection using gabor wavwlwt with various classification technique.pdf. *International Journal of Electrical and Electronic Engineering & Telecomunications*, *1*(special issue), 302–312.

Salton, G., Wong, A., & Yang, C. S. (1975). *AVector Space Model for Automatic Indexing*. *18*(11).

Salton G, Buckley C (1988), Term-weighting approaches in automatic text retrieval,Information Processing & Management, Volume 24, Issue 5,Pp. 513-523,https://doi.org/10.1016/0306-4573(88)90021-0.

Samsudin, N. A., & Bradley, A. P. (2010). Nearest neighbour group-based classification. *Pattern Recognition*, *43*(10), 3458–3467. https://doi.org/10.1016/j.patcog.2010.05.010

SathiyaPriya, K., Sudha Sadasivam, G., & B. Karthikeyan, V. (2012). A New Method for preserving privacy in Quantitative Association Rules using Genetic Algorithm. *International Journal of Computer Applications*, *60*(12), 12–19. https://doi.org/10.5120/9743-4295

Schapire, R. E. (1990). The Strength of Weak Learnability. *Machine Learning*, *5*(2), 197–227. https://doi.org/10.1023/A:1022648800760

Schapire, R. E., & Singer, Y. (2000). BoosTexter: a boosting-based system for text categorization. *Machine Learning*, *39*(2), 135–168. https://doi.org/10.1023/a:1007649029923

Selvi, S. T., Karthikeyan, P., Vincent, A., Abinaya, V., Neeraja, G., & Deepika, R. (2017). Text categorization using Rocchio algorithm and random forest algorithm. *2016 8th International Conference on Advanced Computing, ICoAC 2016*, 7–12. https://doi.org/10.1109/ICoAC.2017.7951736

Severyn, A., & Moschittiy, A. (2015). Learning to rank short text pairs with convolutional deep neural networks. *SIGIR 2015 - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, *August*, 373–382. https://doi.org/10.1145/2766462.2767738

Shaheen, M., Zafar, T., & Ali Khan, S. (2020). Decision tree classification: Ranking journals using IGIDI. *Journal of Information Science*, *46*(3), 325–339. https://doi.org/10.1177/0165551519837176

Sharma, H., & Kumar, S. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. *International Journal of Science and Research (IJSR)*, *5*(4), 2094–2097. https://doi.org/10.21275/v5i4.nov162954

Sheydaei, N., Saraee, M., & Shahgholian, A. (2015). A novel feature selection method for text classification using association rules and clustering. *Journal of Information Science*, *41*(1), 3–15. https://doi.org/10.1177/0165551514550143

Shinde, A., Sahu, A., Apley, D., & Runger, G. (2014). Preimages for variation patterns from kernel PCA and bagging. *IIE Transactions (Institute of Industrial Engineers)*, *46*(5), 429–456. https://doi.org/10.1080/0740817X.2013.849836

Sokhangoee, Z. F., & Rezapour, A. (2022). A novel approach for spam detection based on association rule mining and genetic algorithm. *Computers & Electrical Engineering*, *97*(January), 107655. https://doi.org/10.1016/j.compeleceng.2021.107655

Son, L. H., Chiclana, F., Kumar, R., Mittal, M., Khari, M., Chatterjee, J. M., & Baik, S. W. (2018). ARM–AMO: An efficient association rule mining algorithm based on animal migration optimization. *Knowledge-Based Systems*, *154*(September 2017), 68–80. https://doi.org/10.1016/j.knosys.2018.04.038

Song, K., & Lee, K. (2017). Predictability-based collective class association rule mining. *Expert Systems with Applications*, *79*, 1–7. https://doi.org/10.1016/j.eswa.2017.02.024

Song. A, Ding.X, Chen, M. Li, W. Cao, K. Pu,(2016) Multi-objective association rule mining with binary bat algorithm, Intell. Data Anal. 20 (1), 105–112

Sowmya, B. J., Chetan, & Srinivasa, K. G. (2016). Large scale multi-label text classification of a hierarchical dataset using Rocchio algorithm. *2016 International Conference on Computation System and Information Technology for Sustainable Solutions, CSITSS 2016*, 291–296. https://doi.org/10.1109/CSITSS.2016.7779373

Srikant, R., & Agrawal, R. (1996). Mining Quantitative Association Rules in Large Relational Tables. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, *25*(2), 1–12. https://doi.org/10.1145/235968.233311

Subbulakshmi, B. (2016). Analysis on Prediction of Network Traffic through

Association Rule Mining. *International Journal of Communication and Networking System*, *5*(2), 126–131. https://doi.org/10.20894/ijcnes.103.005.002.008

Subramaniam, T., Jalab, H. A., & Taqa, A. Y. (2010). Overview of textual anti-spam filtering techniques. *International Journal of Physical Sciences*, *5*(12), 1869–1882.

Swain, P. H., & Hauska, H. (1977). Decision Tree Classifier: Design and Potential. *IEEE Trans Geosci Electron*, *GE-15*(3), 142–147. https://doi.org/10.1109/tge.1977.6498972

Taba, D. A., Hasankhani, A., & Bolurian, M. (2012). Soft nexuses. *Computers and Mathematics with Applications*, *64*(6), 1812–1821. https://doi.org/10.1016/j.camwa.2012.02.048

Tang, Z., & Liao, Q. (2007). A New Class Based Associative Classification Algorithm. *International Journal of Applied Mathematics*, *May*, 1–5. https://doi.org/10.1111/tpj.12039

Taunk, K. (2019). A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019*, *Iciccs*, 1255–1260.

Thabtah, F. (2007). A review of associative classification mining. *Knowledge Engineering Review*, *22*(1), 37–65. https://doi.org/10.1017/S0269888907001026

Thabtah, F., Cowling, P., & Peng, Y. (2005). MCAR: Multi-class Classification based on Association Rule. *3rd ACS/IEEE International Conference on Computer Systems and Applications, 2005*, *2005*, 127–133. https://doi.org/10.1109/AICCSA.2005.1387030

Thabtah, F., Hadi, W., Abdelhamid, N., & Issa, A. (2011). Prediction phase in associative classification mining. *International Journal of Software Engineering and Knowledge Engineering*, *21*(6), 855–876. https://doi.org/10.1142/S0218194011005463

Thangaraj, M., & Sivakami, M. (2018). Text Classification Techniques: A Literature Review. *Interdisciplinary Journal Of Impormation, Knowledge, and Management*, *13*, 117–135.

Trieschnigg, D., Pezik, P., Lee, V., de Jong, F., Kraaij, W., & Rebholz-Schuhmann, D. (2009). MeSH Up: Effective MeSH text classification for improved document retrieval. *Bioinformatics*, *25*(11), 1412–1418. https://doi.org/10.1093/

bioinformatics/btp249

Türkmen, E., & Pancar, A. (2013). On some new operations in soft module theory. *Neural Computing and Applications*, *22*(6), 1233–1237. https://doi.org/10.1007/s00521-012-0893-6

Vanwinckelen, G. (2019). On Estimating Model Accuracy with Repeated Cross-Validation. In *lirias.kuleuven*.

Vuttipittayamongkol, P., Elyan, E., & Petrovski, A. (2021). On the class overlap problem in imbalanced data classification. *Knowledge-Based Systems*, *212*, 106631. https://doi.org/10.1016/j.knosys.2020.106631

Vyas, R., Sharma, L. K., Vyas, O. P., & Scheider, S. (2008). Associative classifiers for predictive analytics: Comparative performance study. *Proceedings - EMS 2008, European Modelling Symposium, 2nd UKSim European Symposium on Computer Modelling and Simulation*, *September 2014*, 289–294. https://doi.org/10.1109/EMS.2008.29

Wang, X, K Yue, W Niu and Z Shi (2011). An approach for adaptive associative classification. Expert Systems with Applications: An International Journal, 38(9), 11873–11883

Wang, Y., Liao, W., & Chang, Y. (2018). Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies*, *11*(8), 1–14. https://doi.org/10.3390/en11082163

Wedyan, Suzan., & Wedyan, Fadi. (2013). An Associative Classification Data Mining Approach for Detecting Phishing Websites. *Journal of Emerging Trends in Computing and Information Sciences*, *4*(12), 888–899.

Wicaksono, D., Jambak, M. I., & Saputra, D. M. (2020). The Comparison of Apriori Algorithm with Preprocessing and FP-Growth Algorithm for Finding Frequent Data Pattern in Association Rule. In *Advances in Intelligent Systems Research* (Vol. 172, Issue Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019), pp. 315–319). https://doi.org/10.2991/aisr.k.200424.047

Xiao, Z., Chen, W., & Li, L. (2012). An integrated FCM and fuzzy soft set for supplier selection problem based on risk evaluation. *Applied Mathematical Modelling*, *36*(4), 1444–1454. https://doi.org/10.1016/j.apm.2011.09.038

Xiao, Z., Xia, S., Gong, K., & Li, D. (2012). The trapezoidal fuzzy soft set and its application in MCDM. *Applied Mathematical Modelling*, *36*(12), 5844–5855.

https://doi.org/10.1016/j.apm.2012.01.036

Yang, Y., & Miller, J. (1997). Rules over Interval Data. *ACM SIGMOD International Conference on Management of Data*, 452–461.

Yin, X., & Han, J. (2003). CPAR: Classification based on Predictive Association Rules. In *Proceedings of the 2003 SIAM International Conference on Data Mining* (pp. 331–335). https://doi.org/10.1137/1.9781611972733.40

Yoon, Y., & Lee, G. G. (2007). Efficient implementation of associative classifiers for document classification. *Information Processing and Management*, *43*(2), 393–405. https://doi.org/10.1016/j.ipm.2006.07.012

Zhang, J., Kowsari, K., Harrison, J. H., Lobo, J. M., & Barnes, L. E. (2018). Patient2Vec: A Personalized Interpretable Deep Representation of the Longitudinal Electronic Health Record. *IEEE Access*, *6*(c), 65333–65346. https://doi.org/10.1109/ACCESS.2018.2875677

Zhang, W., & Gao, F. (2011). An Improvement to Naive Bayes for Text Classification. *Procedia Engineering*, *15*, 2160–2164. https://doi.org/10.1016/j.proeng.2011.08.404

Zhou, Y., Ji, Z., & Wang, K. (2017). *A Parallel Decision Tree Based Algorithm on MPI for Multi-label Classification Learning*. *134*(Caai), 366–369. https://doi.org/10.2991/caai-17.2017.83

VITA

The author was born in Bandung, Indonesia, on December 20, 1967. He attended the SMAN 4 Bandung, Indonesia, for his senior high school. He pursued his degree at the Bandung Institute of Technology, Indonesia, and graduated with a B.Sc. in applied mathematics for computation in 1991. After graduating, he worked as a lecturer at Telkom University, Indonesia, Department of Informatics Engineering. Apart from that, he also works as a consultant in information technology implementation. He is a consultant for the information technology master plan of the Bandung district government, the Sumedang district government, and the Tangerang city government. Consultant of human resource development for information technology in communication and information service of West Java. Consultant for financial budget information systems for Indonesian Embassies in Singapore, Kuala Lumpur, Penang, Kinabalu, Bandar Sri Begawan, Manila, Tokyo, New Delhi, Hong Kong, and Rome. He then enrolled at the Bandung Institute of Technology, Indonesia, in 1996, where he was awarded an M. Eng. in software engineering in 1999. In 2015, Mr. Dede Rohidin attended the Graduate School of University Tun Hussein Onn Malaysia and was accepted into the Ph.D. program in computer science and information technology. During this time, he was the author of three papers in text classification. He is currently a member of the Institute of Electrical and Electronics Engineering (IEEE).