

THE HYBRID FEATURE SELECTION
TECHNIQUE USING TERM FREQUENCY-
INVERSE DOCUMENT FREQUENCY AND
SUPPORT VECTOR MACHINE-RECURSIVE
FEATURE ELIMINATION FOR SENTIMENT
CLASSIFICATION

NUR SYAFIQAH BINTI MOHD NAFIS

DOCTOR OF PHILOSOPHY

UNIVERSITI MALAYSIA PAHANG



SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis, and, in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Doctor of Philosophy.

A handwritten signature in black ink, appearing to read 'Suryanti', is written above a horizontal line.

(Supervisor's Signature)

Full Name : TS. DR. SURYANTI BINTI AWANG

Position : SENIOR LECTURER

Date : 28 OCTOBER 2022



STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

A handwritten signature in black ink, consisting of a stylized 'S' followed by a flourish.

(Student's Signature)

Full Name : NUR SYAFIQAH BINTI MOHD NAFIS

ID Number : PCC17006

Date : 28 OCTOBER 2022

THE HYBRID FEATURE SELECTION TECHNIQUE USING TERM
FREQUENCY-INVERSE DOCUMENT FREQUENCY AND SUPPORT
VECTOR MACHINE-RECURSIVE FEATURE ELIMINATION FOR
SENTIMENT CLASSIFICATION

NUR SYAFIQAH BINTI MOHD NAFIS

Thesis submitted in fulfilment of the requirements
for the award of the degree of
Doctor of Philosophy

Faculty of Computing
UNIVERSITI MALAYSIA PAHANG

OCTOBER 2022

ACKNOWLEDGEMENTS

First and foremost, I praise to Almighty Allah for his blessings, mercies, and guidance during my years. In addition, I would like to thank everyone who played a role in the accomplishment of this thesis. I'd like to convey my extreme gratitude to my supervisor, Ts, Dr. Suryanti Awang, for her insightful and continuous guidance and advice. Her immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. Thank you for keeping me on track so that I didn't lose focus while working on this thesis.

Next, I'd like to express my heartfelt gratitude to my great husband, Khadir bin Mohamad for his unwavering support i.e., mentally, physically, and financially during my journey. Without his tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

Not to mention my wonderful mother, Anita binti Abdullah who has always encouraged me to continue my studies. Her prayers keep me strong to face all the challenges. I'd like to thank all of the scholars and authors included in this thesis, whose work and research have provided me with knowledge and assisted me in completing this project thesis. Not to forget my siblings who keep my days chill during ups and downs.

I'd like to express my gratitude to my post-graduate pals especially SPINT lab pals for their excellent suggestions and advice. Memories that will remain bloomed in my heart. Lastly, to my sponsor UMP and KPT. Without the scholarship offered, it may be difficult for me to further my studies in these hard years.

Thank you everybody for participating in my PhD adventure. I hope knowledge and experiences gained are beneficial for my future.

ABSTRAK

Pengkelasan sentimen semakin digunakan untuk mengenal pasti secara automatik sentimen positif atau negatif dalam dokumen teks pendapat, contohnya, maklum balas atau ulasan pelanggan. Pemilihan ciri sentiasa menjadi masalah kritikal dan mencabar dalam pengkelasan sentimen berasaskan pembelajaran mesin. Pemilihan ciri hibrid ialah teknik yang cekap dalam pengkelasan sentimen. Walau bagaimanapun, terdapat beberapa kelemahan yang boleh diselesaikan. Pertama, keupayaan untuk mengenal pasti kepentingan ciri dan mengurangkan beberapa ciri daripada dokumen teks pendapat. Kegagalan menangani isu ini akan mengakibatkan prestasi pengkelasan yang lemah. Oleh itu, penyelidikan ini bertujuan untuk menambah baik prestasi pengkelasan dengan mencadangkan Term Frequency-Inverse Document Frequency (TF-IDF) and Support Vector Machine-Recursive Feature Elimination (SVM-RFE) sebagai teknik pemilihan ciri hibrid. TF-IDF menilai kepentingan ciri, dan ambang berasaskan sisihan piawai digunakan untuk pengurangan ciri. Objektifnya adalah untuk menambah baik pendekatan konvensional untuk mengurangkan ciri daripada matriks ciri. Kemudian, SVM-RFE menilai semula dan meletakkan kedudukan ciri yang selebihnya daripada matriks ciri berasaskan TF-IDF. Hanya kumpulan ciri k -top daripada ciri kedudukan SVM-RFE digunakan untuk pengkelasan sentimen. Akhir sekali, pengkelas SVM digunakan untuk mengkelaskan dataset ulasan pelanggan bahasa Inggeris, iaitu, berlabel pendapat dan IMDb besar. Prestasi diukur menggunakan ketepatan, kejituan, ingat semula, ukuran-F, dan pengurangan saiz ciri. Keputusan eksperimen menunjukkan prestasi yang menjanjikan sehingga 95.06% dalam pengukuran prestasi, terutamanya daripada dataset IMDb yang besar dan dataset tambahan, ulasan hotel. Oleh itu, teknik yang dicadangkan boleh mengurangkan 31.80% hingga 64.00% ciri semasa pengkelasan. Kadar pengurangan ini penting dalam menggunakan sumber komputasi secara optimum sambil mengekalkan kecekapan prestasi pengkelasan.

ABSTRACT

Sentiment classification is increasingly used to automatically identify a positive or negative sentiment in the opinionated text document, for instance, customer feedback or review. Feature selection has always been a critical and challenging problem in machine learning-based sentiment classification. Hybrid feature selection is an efficient technique in sentiment classification. However, there are several disadvantages that can be solved. Firstly, the ability to identify feature importance and reduce some features from opinionated text documents. The failure to address this issue will result in poor classification performance. Therefore, this research aims to improve the classification performances by proposing term frequency-inverse document frequency (TF-IDF) and support vector machine-recursive feature elimination (SVM-RFE) as a hybrid feature selection technique. The TF-IDF evaluates the feature importance, and the standard deviation-based threshold is used for feature reduction. The objective is to improve the conventional approach of reducing features from feature matrix. Later, the SVM-RFE re-evaluates and ranks the remaining features from TF-IDF-based feature matrix. Only the k -top features group from the SVM-RFE ranked features were used for sentiment classification. Finally, the support vector machine (SVM) classifier is employed to classify the English customer review datasets, i.e., opinion-labelled, and large IMDb. The performance was measured using accuracy, precision, recall, F-measure, and feature size reduction. The experimental results present promising performances up to 95.06% in the performance measurements, especially from the large IMDb datasets and additional dataset, hotel review. Consequently, the proposed technique could minimise 31.80% to 64.00% of the features during classification. This reduction rate is significant in optimally utilising the computational resources while preserving the efficiency of the classification performance.

TABLE OF CONTENT

ACKNOWLEDGEMENTS	ii
ABSTRAK	iii
ABSTRACT	iv
TABLE OF CONTENT	v
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xi
LIST OF APPENDICES	xiii
CHAPTER 1 INTRODUCTION	1
1.1 Research Background	1
1.2 Problem Statement	7
1.3 Research Objectives	11
1.4 Research Scope and Limitation	12
1.5 Research Contributions	12
1.6 Thesis Outline	13
CHAPTER 2 LITERATURE REVIEW	14
2.1 Introduction	14
2.2 Sentiment Classification	15
2.2.1 Lexicon-based Sentiment Classification	20
2.2.2 The Hybrid Sentiment Classification (Lexicon + Machine Learning Sentiment Classification)	25
2.2.3 Machine Learning-based Sentiment Classification	27
2.3 Supervised Machine Learning-based Sentiment Classification Framework	29
2.3.1 Pre-Processing	30

2.3.2	Feature Selection	31
2.3.3	The opinionated text classifier	58
2.4	Discussion	63
CHAPTER 3 METHODOLOGY		65
3.1	Introduction	65
3.2	Data Acquisition Phase	66
3.3	Data Pre-processing Phase	71
3.3.1	Tokenisation	72
3.3.2	Stop-word removal	73
3.3.3	Short-word removal	74
3.3.4	Stemming	76
3.3.5	Term-Document Matrix (TDM) Conversion	78
3.4	Feature Selection Phase	79
3.4.1	Stage 1: Non-predictive Measure Stage	81
3.4.2	Stage 2: Predictive Measure Stage	87
3.5	Classification and Evaluation Phase	90
3.5.1	SVM Classifier	90
3.5.2	Performance Measures	92
3.6	Description of the Experiments	94
3.7	Summary	95
CHAPTER 4 RESULTS AND DISCUSSION		97
4.1	Introduction	97
4.2	Experiment 1: The Baseline Experiment using Term-Frequency	98
4.3	Experiment 2: Feature Reduction using Standard Deviation- based Threshold	101

4.4	Experiment 3: Sentiment Classification using The Proposed TF-IDF + SVM-RFE Feature Selection	105
4.5	Performance Comparison 1: The Baseline Experiment and The Proposed TF-IDF + SVM-RFE Feature Selection Technique	115
4.6	Experiment 4: Performances of The Proposed TF-IDF + SVM-RFE Feature Selection Technique using Other Datasets	120
4.7	Performance Comparison 2: The Proposed TF-IDF + SVM-RFE Feature Selection Technique with Other State-of-the-art Techniques	124
4.8	Conclusion	126
CHAPTER 5 CONCLUSION		129
5.1	Introduction	129
5.2	Summary of Contributions	130
5.3	Recommendations and Future Works	132
REFERENCES		133
LIST OF PUBLICATION		167
APPENDICES		168

REFERENCES

Bibliography

- Abd, D. H., Sadiq, A. T., & Abbas, A. R. (2020a). Classifying Political Arabic Articles Using Support Vector Machine with Different Feature Extraction. *Communications in Computer and Information Science, 1174 CCIS*, 79–94. https://doi.org/10.1007/978-3-030-38752-5_7
- Abd, D. H., Sadiq, A. T., & Abbas, A. R. (2020b). Political Articles Categorization Based on Different Naïve Bayes Models. In *Communications in Computer and Information Science: Vol. 1174 CCIS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-38752-5_23
- Abdulkhaliq, S. S., & Darwesh, A. M. (2020). Sentiment Analysis Using Hybrid Feature Selection Techniques. *UHD Journal of Science and Technology, 4*(1), 29. <https://doi.org/10.21928/uhdjst.v4n1y2020.pp29-40>
- Abdulla, N. A., Ahmed, N. A., Shehab, M. A., Al-Ayyoub, M., Al-Kabi, M. N., & Al-rifai, S. (2014). Towards improving the lexicon-based approach for arabic sentiment analysis. *International Journal of Information Technology and Web Engineering, 9*(3), 55–71. <https://doi.org/10.4018/ijitwe.2014070104>
- 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 (Switzerland), 10*(2), 1–29. <https://doi.org/10.3390/electronics10020101>
- Abualigah, L. M., Khader, A. T., Al-Betar, M. A., & Alomari, O. A. (2017). Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. *Expert Systems with Applications, 84*, 24–36. <https://doi.org/10.1016/j.eswa.2017.05.002>
- Adeleke, A., & Samsudin, N. (2018). A Hybrid Feature Selection Technique for Classification of Group-based Holy Quran Verses. *International Journal of Engineering & Technology, December*, 228–233. www.sciencepubco.com/index.php/IJET
- Agarwal, A., & Toshniwal, D. (2018). Application of Lexicon Based Approach in Sentiment Analysis for short Tweets. *Proceedings on 2018 International Conference on Advances in Computing and Communication Engineering, ICACCE 2018, June*, 189–193. <https://doi.org/10.1109/ICACCE.2018.8441696>
- Agarwal, B., & Mittal, N. (2016). Machine Learning Approach for Sentiment Analysis.

- In *Prominent Feature Extraction for Sentiment Analysis* (pp. 21–45).
https://doi.org/10.1007/978-3-319-25343-5_3
- Agarwal, B., & Mittal, N. (2013). Sentiment Classification using Rough Set based Hybrid Feature Selection. *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2013), June*, 115–119. <http://www.aclweb.org/anthology/W/W13/W13-16.pdf#page=127>
- Ahmad, I. (2015). Feature Selection Using Particle Swarm Optimization. *JAISCR*, 5(4), 231–238.
- Ahmad, M., Aftab, S., & Ali, I. (2017). Sentiment Analysis of Tweets using SVM. *International Journal of Computer Applications*, 177(5), 25–29.
<https://doi.org/10.5120/ijca2017915758>
- Ahmad, S. R., Bakar, A. A., & Yaakub, M. R. (2015). Metaheuristic algorithms for feature selection in sentiment analysis. *Proceedings of the 2015 Science and Information Conference, SAI 2015*, 222–226.
<https://doi.org/10.1109/SAI.2015.7237148>
- Ahmad, S. R., Bakar, A. A., & Yaakub, M. R. (2019a). A review of feature selection techniques in sentiment analysis. *Intelligent Data Analysis*, 23(1), 159–189.
<https://doi.org/10.3233/IDA-173763>
- Ahmad, S. R., Bakar, A. A., & Yaakub, M. R. (2019b). Ant colony optimization for text feature selection in sentiment analysis. *Intelligent Data Analysis*, 23(1), 133–158.
<https://doi.org/10.3233/IDA-173740>
- Ahmad, S. R., Mohd Yusop, N. M., AbuBakar, A., & Yaakub, M. R. (2017). Statistical analysis for validating ACO-KNN algorithm as feature selection in sentiment analysis. *AIP Conference Proceedings*, 1891. <https://doi.org/10.1063/1.5005351>
- Ahuja, R., & Anand, W. (2018). Sentiment classification of movie reviews using dual training and dual prediction. *2017 4th International Conference on Image Information Processing, ICIIIP 2017, 2018-Janua*, 594–597.
<https://doi.org/10.1109/ICIIIP.2017.8313786>
- Al-Harbi, O. (2019). A comparative study of feature selection methods for dialectal arabic sentiment classification using support vector machine. *IJCSNS International Journal of Computer Science and Network Security*, 19(1), 167–176.
- Al-khaleefa, A. S., Awang, S., Tiun, S., AL-Saffar, A., & AL-Saiagh, W. (2018). Word sense disambiguation using hybrid swarm intelligence approach. *Plos One*, 13(12), e0208695. <https://doi.org/10.1371/journal.pone.0208695>

- Al-nasseri, A. A. (2016). *THE PREDICTIVE POWER OF STOCK MICRO-BLOGGING SENTIMENT IN FORECASTING STOCK MARKET BEHAVIOUR* (Issue February).
- Al Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis. *Procedia Computer Science*, 127, 511–520.
- Alatabi, H. A., & Abbas, A. R. (2020). Sentiment analysis in social media using machine learning techniques. *Iraqi Journal of Science*, 61(1), 193–201. <https://doi.org/10.24996/ij.s.2020.61.1.22>
- Alemneh, G. N., Rauber, A., & Atnafu, S. (2020). *Corpus based Amharic sentiment lexicon generation*. 1–3. <https://doi.org/10.18653/v1/2020.winlp-1.1>
- Alibeigi, M., Hashemi, S., & Hamzeh, A. (2012). DBFS: An effective Density Based Feature Selection scheme for small sample size and high dimensional imbalanced data sets. *Data and Knowledge Engineering*, 81–82, 67–103. <https://doi.org/10.1016/j.datak.2012.08.001>
- Almagrabi, H. (2019). *Predicting the Helpfulness of Product Reviews: A Sentence Classification Approach*.
- Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*, 10(2), 361–374. <https://doi.org/10.14569/ijacsa.2019.0100248>
- Alsaffar, A., & Omar, N. (2014). Study on feature selection and machine learning algorithms for Malay sentiment classification. *Conference Proceedings - 6th International Conference on Information Technology and Multimedia at UNITEN: Cultivating Creativity and Enabling Technology Through the Internet of Things, ICIMU 2014, Putrajaya, Malaysia*, 270–275.
- Alsaleem, S. (2011). Automated Arabic Text Categorization Using SVM and NB. *Int. Arab J. e-Technol.*, 2(2), 124–128.
- Alt, B., & Ganiz, M. C. (2018). Semantic text classification : A survey of past and recent advances. *Information Processing and Management*, 54(July), 1129–1153. <https://doi.org/10.1016/j.ipm.2018.08.001>
- Alwakid, G. (2020). *Sentiment Analysis of Dialectical Arabic Social Media Content Using A Hybrid Linguistic-Machine Learning Approach* (Issue January).
- Aman, S., & Szpakowicz, S. (2007). Identifying expressions of emotion in text. In *Text, Speech and Dialogue* (Vol. 4629, pp. 196–205). Springer.

https://doi.org/10.1007/978-3-540-74628-7_27

- Anand Joseph Daniel, D., & Janaki Meena, M. (2021). A NOVEL SENTIMENT ANALYSIS FOR AMAZON DATA WITH TSA BASED FEATURE SELECTION. *Scalable Computing: Practice and Experience*, 22, 53–66.
- Andrea, Z., Stuart E, M., Tokarchuk, N, L., Wang, & Xinyue. (2013). Social Media Text Mining and Network Analysis for Decision Support in Natural Crisis Management. *Proceedings of the 10th International ISCRAM Conference*, 840–845.
- Anjali, J. G. (2007). A Comparative Study of Stemming Algorithms. *October*, 2(2004), 1930–1938. <https://doi.org/10.1.1.642.7100>
- Ansari, G., Ahmad, T., & Doja, M. N. (2019a). Ensemble of feature ranking methods using hesitant fuzzy sets for sentiment classification. *International Journal of Machine Learning and Computing*, 9(5), 599–608. <https://doi.org/10.18178/ijmlc.2019.9.5.846>
- Ansari, G., Ahmad, T., & Doja, M. N. (2019b). Hybrid Filter–Wrapper Feature Selection Method for Sentiment Classification. *Arabian Journal for Science and Engineering*, 44(11), 9191–9208. <https://doi.org/10.1007/s13369-019-04064-6>
- Apolloni, J., Leguizamón, G., & Alba, E. (2016). Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. *Applied Soft Computing Journal*, 38, 922–932. <https://doi.org/10.1016/j.asoc.2015.10.037>
- Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, 108, 110–124. <https://doi.org/10.1016/j.knosys.2016.05.040>
- Arafat, H., Elawady, R. M., & Barakat, S. (2019). Different Feature Selection for Sentiment Classification. *International Journal of Information Science and Intelligent System*, 3(1), 137–150.
- Arif, M. H., Li, J., Iqbal, M., & Liu, K. (2018). Sentiment analysis and spam detection in short informal text using learning classifier systems. *Soft Computing*, 22(21), 7281–7291. <https://doi.org/10.1007/s00500-017-2729-x>
- Asghar, M. Z., Khan, A., Ahmad, S., & Kundi, F. M. (2014). A Review of Feature Extraction in Sentiment Analysis. *J. Basic. Appl. Sci. Res*, 4(3), 181–186.
- Asghar, M. Z., Khan, A., Ahmad, S., Qasim, M., & Khan, I. A. (2017). Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. *PLoS ONE*, 12(2), 1–22. <https://doi.org/10.1371/journal.pone.0171649>

- Auria, L., & Moro, R. A. (2008). *Support Vector Machines (SVM) as a Technique for Solvency Analysis* (Issue August).
- Avinash, M., & Sivasankar, E. (2019). A study of feature extraction techniques for sentiment analysis. In *Advances in Intelligent Systems and Computing* (pp. 1–11). Springer. <http://arxiv.org/abs/1906.01573v1>
- Bagheri, A., Saraee, M., & De Jong, F. (2013). Sentiment classification in Persian: Introducing a mutual information-based method for feature selection. *2013 21st Iranian Conference on Electrical Engineering, ICEE 2013*. <https://doi.org/10.1109/IranianCEE.2013.6599671>
- Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies, ICOT 2018*, 1–6. <https://doi.org/10.1109/ICOT.2018.8705796>
- Belazzoug, M., Touahria, M., Nouioua, F., & Brahimi, M. (2020). An improved sine cosine algorithm to select features for text categorization. *Journal of King Saud University - Computer and Information Sciences*, 32(4), 454–464. <https://doi.org/10.1016/j.jksuci.2019.07.003>
- Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, 28, 1–15. <https://doi.org/10.1016/j.inffus.2015.08.005>
- Bermejo, P., De La Ossa, L., Gámez, J. A., & Puerta, J. M. (2012). Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking. *Knowledge-Based Systems*, 25(1), 35–44. <https://doi.org/10.1016/j.knosys.2011.01.015>
- Bhamare, B. R., Jeyanthi, P., & Subhashini, R. (2019). Aspect category extraction for sentiment analysis using multivariate filter method of feature selection. *International Journal of Recent Technology and Engineering*, 8(3), 2138–2143. <https://doi.org/10.35940/ijrte.C4566.098319>
- Bhamare, B. R., & Prabhu, J. (2021). A supervised scheme for aspect extraction in sentiment analysis using the hybrid feature set of word dependency relations and lemmas. *PeerJ Comput. Sci.*, 1–22. <https://doi.org/10.7717/peerj-cs.347>
- 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>
- Bhattacharjee, U., Srijith, P. K., & Desarkar, M. S. (2019). Term Specific TF-IDF

- Boosting for Detection of Rumours in Social Networks. *2019 11th International Conference on Communication Systems and Networks, COMSNETS 2019*, 726–731. <https://doi.org/10.1109/COMSNETS.2019.8711427>
- Bhavitha, B. K., Rodrigues, A. P., & Chiplunkar, N. N. (2017). Comparative study of machine learning techniques in sentimental analysis. *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017, Iccct*, 216–221. <https://doi.org/10.1109/ICICCT.2017.7975191>
- Bhuvaneswari, K., & Parimala, R. (2017). Sentiment Reviews Classification using Hybrid Feature Selection. *International Journal of Database Theory and Application*, 10(7), 1–12. <https://doi.org/10.14257/ijda.2017.10.7.01>
- Bi, J., Bennett, K. P., Embrechts, M., Breneman, C. M., & Song, M. (2003). Dimensionality Reduction via Sparse Support Vector Machines. *Journal of Machine Learning Research*, 3, 1229–1243. <https://doi.org/10.1162/153244303322753643>
- Bidi, N., & Elberrichi, Z. (2017). Feature selection for text classification using genetic algorithms. *Proceedings of 2016 8th International Conference on Modelling, Identification and Control, ICMIC 2016*, 806–810. <https://doi.org/10.1109/ICMIC.2016.7804223>
- Biltawi, M., Al-Naymat, G., & Tedmori, S. (2017). Arabic Sentiment Classification: A Hybrid Approach. *Proceedings - 2017 International Conference on New Trends in Computing Sciences, ICTCS 2017, 2018-Janua*, 104–108. <https://doi.org/10.1109/ICTCS.2017.24>
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. <https://doi.org/10.1016/j.knosys.2021.107134>
- Bodapati, J. D., Veeranjanyulu, N., & Shaik, S. (2019). Sentiment Analysis from Movie Reviews Using LSTMs. *International Information and Engineering Technology Association*, 24(1), 125–129.
- Bouaguel, W., Bel Mufti, G., & Limam, M. (2013). Rank aggregation for filter feature selection in credit scoring. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8284 LNAI(December), 7–15. https://doi.org/10.1007/978-3-319-03844-5_2
- Boudad, N., Faizi, R., Oulad Haj Thami, R., & Chiheb, R. (2018). Sentiment analysis in

- Arabic: A review of the literature. *Ain Shams Engineering Journal*, 9(4), 2479–2490. <https://doi.org/10.1016/j.asej.2017.04.007>
- Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300, 70–79. <https://doi.org/10.1016/j.neucom.2017.11.077>
- Cateni, S., Colla, V., & Vannucci, M. (2014). A hybrid feature selection method for classification purposes. *Proceedings - UKSim-AMSS 8th European Modelling Symposium on Computer Modelling and Simulation, EMS 2014, October 2014*, 39–44. <https://doi.org/10.1109/EMS.2014.44>
- Cekik, R., & Uysal, A. K. (2020). A novel filter feature selection method using rough set for short text data. *Expert Systems with Applications*, 160, 113691. <https://doi.org/10.1016/j.eswa.2020.113691>
- Chakriswaran, P., Vincent, D. R., Srinivasan, K., Sharma, V., Chang, C. Y., & Reina, D. G. (2019). Emotion AI-driven sentiment analysis: A survey, future research directions, and open issues. *Applied Sciences (Switzerland)*, 9(24). <https://doi.org/10.3390/app9245462>
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1), 16–28. <https://doi.org/10.1016/j.compeleceng.2013.11.024>
- Chang, J. R., Chen, M. Y., Chen, L. S., & Tseng, S. C. (2019). Why Customers Don't Revisit in Tourism and Hospitality Industry? *IEEE Access*, 7, 146588–146606. <https://doi.org/10.1109/ACCESS.2019.2946168>
- Chaoui, C., Webster, C., & Tan, L. P. (2017). SOCIAL MEDIA SENTIMENT ANALYSIS: LEXICON VERSUS MACHINE LEARNING. *Journal of Consumer Marketing*, 34(1), 1–5.
- Chen, C. W., Tsai, Y. H., Chang, F. R., & Lin, W. C. (2020). Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. *Expert Systems*, 37(5), 1–10. <https://doi.org/10.1111/exsy.12553>
- Chen, H., Hou, Q., Han, L., Hu, Z., Ye, Z., Zeng, J., & Yuan, J. (2019). Distributed text feature selection based on bat algorithm optimization. *Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2019, 1*, 75–80. <https://doi.org/10.1109/IDAACS.2019.8924308>
- Chen, J., Chen, C., & Liang, Y. (2016). Optimized TF-IDF Algorithm with the Adaptive Weight of Position of Word. *2nd International Conference on Artificial*

- Intelligence and Industrial Engineering (AIIE2016)*, 133, 114–117.
<https://doi.org/10.2991/aiie-16.2016.28>
- Chen, K., Zhang, Z., Long, J., & Zhang, H. (2016). Turning from TF-IDF to TF-IGM for term weighting in text classification. *Expert Systems with Applications*, 66, 1339–1351. <https://doi.org/10.1016/j.eswa.2016.09.009>
- Chen, T., Su, P., Shang, C., Hill, R., Zhang, H., & Shen, Q. (2019). Sentiment Classification of Drug Reviews Using Fuzzy-rough Feature Selection. *IEEE International Conference on Fuzzy Systems, 2019-June*.
<https://doi.org/10.1109/FUZZ-IEEE.2019.8858916>
- Cheng, I. (2013). Hybrid Methods for Feature Selection. In *Western Kentucky University* (Issue 02). <https://doi.org/10.1109/IACS.2015.7103228>
- Chi, X., Cambria, E., & Siew, T. P. (2017). Adaptive two-stage feature selection for sentiment classification. *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017, 2017-Janua*, 1238–1243.
<https://doi.org/10.1109/SMC.2017.8122782>
- Chinnalagu, A., & Durairaj, A. K. (2021). *Context-based sentiment analysis on customer reviews using machine learning linear models*.
<https://doi.org/10.7717/peerj-cs.813>
- Chong, W. Y., Selvaretnam, B., & Soon, L. K. (2014). Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets. *Proceedings - 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, ICAIET 2014*, 212–217.
<https://doi.org/10.1109/ICAIET.2014.43>
- Choudhary, S., Sandhu, A. S., & Pradhan, T. (2017). Enhanced Prediction of Online News Popularity. In *Intelligent Systems and Computing 556* (pp. 133–144).
<https://doi.org/10.1007/978-981-10-3874-7>
- Chua, S., & Kulathuramaiyer, N. (2004). Semantic feature selection using WordNet. *Proceedings - IEEE/WIC/ACM International Conference on Web Intelligence, WI 2004*, 166–172. <https://doi.org/10.1109/WI.2004.10115>
- Da Silva, N. F. F., Coletta, L. F. S., & Hruschka, E. R. (2016). A survey and comparative study of tweet sentiment analysis via semi-supervised learning. *ACM Computing Surveys*, 49(1), 1–26. <https://doi.org/10.1145/2932708>
- Daassi-Gnaba, H., & Oussary, Y. (2015). External vs. Internal SVM-RFE: The SVM-RFE method revisited and applied to emotion recognition. *Neural Network World*, 25(1), 75–91. <https://doi.org/10.14311/NNW.2015.25.004>

- Dadgar, S. M. H., Araghi, M. S., & Farahani, M. M. (2016). A novel text mining approach based on TF-IDF and support vector machine for news classification. *Proceedings of 2nd IEEE International Conference on Engineering and Technology, ICETECH 2016, March*, 112–116. <https://doi.org/10.1109/ICETECH.2016.7569223>
- Dai, W. (2018). Improvement and Implementation of Feature Weighting Algorithm TF-IDF in Text Classification. *Advances in Intelligent Systems Research*, 147, 583–587.
- Dang, N. C., Moreno-García, M. N., & de la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(483), 1–29.
- Das, B., & Chakraborty, S. (2018). An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation. *ArXiv*. <http://arxiv.org/abs/1806.06407>
- Das, P. K., Majumder, M. K., Kaushik, B. K., & Manhas, S. K. (2014). A Survey on Filter Techniques for Feature Selection in Text Mining. *Advances in Intelligent Systems and Computing*, 236(SocProS), 1117–1126. <https://doi.org/10.1007/978-81-322-1602-5>
- De Groot, R. (2012). *Data Mining for Tweet Sentiment Classification*.
- Deng, X., Li, Y., Weng, J., & Zhang, J. (2019a). Feature selection for text classification: A review. *Multimedia Tools and Applications*, 78(3), 3797–3816. <https://doi.org/10.1007/s11042-018-6083-5>
- Deng, X., Li, Y., Weng, J., & Zhang, J. (2019b). Feature selection for text classification: A review. *Multimed Tools Application*, 78, 3797–3816. <https://doi.org/10.1109/CCDC.2010.5498331>
- Desai, M., & Mehta, M. A. (2017). Techniques for sentiment analysis of Twitter data: A comprehensive survey. *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, 149–154. <https://doi.org/10.1109/CCAA.2016.7813707>
- Devi, D. V. N., Kumar, C. K., & Prasad, S. (2016). A Feature Based Approach for Sentiment Analysis by Using Support Vector Machine. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, 3–8. <https://doi.org/10.1109/IACC.2016.11>
- Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, 87, 44–49. <https://doi.org/10.1016/j.procs.2016.05.124>

- Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier. *International Journal of Information Engineering and Electronic Business*, 8(4), 54–62. <https://doi.org/10.5815/ijieeb.2016.04.07>
- Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*, 34(6), 480–488. <https://doi.org/10.1108/JCM-03-2017-2141>
- El Aboudi, N., & Benhlima, L. (2016). Review on wrapper feature selection approaches. *Proceedings - 2016 International Conference on Engineering and MIS, ICEMIS 2016*. <https://doi.org/10.1109/ICEMIS.2016.7745366>
- Elawady, R. M., Barakat, S., & Elrashidy, N. M. (2014). Different Feature Selection for Sentiment Classification. *International Journal of Information Science and Intelligent System*, 3(1), 137–150.
- Elshakankery, K., & Ahmed, M. F. (2019). HILATSA: A hybrid Incremental learning approach for Arabic tweets sentiment analysis. *Egyptian Informatics Journal*, 20(3), 163–171. <https://doi.org/10.1016/j.eij.2019.03.002>
- Endalie, D., & Haile, G. (2021). Hybrid Feature Selection for Amharic News Document Classification. *Mathematical Problems in Engineering*, 2021(March). <https://doi.org/10.1155/2021/5516262>
- Ernawati, S., Wati, R., Nuris, N., Marita, L. S., & Yulia, E. R. (2020). Comparison of Naïve Bayes Algorithm with Genetic Algorithm and Particle Swarm Optimization as Feature Selection for Sentiment Analysis Review of Digital Learning Application. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012040>
- Farisi, A. A., Sibaroni, Y., & Faraby, S. Al. (2019). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series*, 1192(1). <https://doi.org/10.1088/1742-6596/1192/1/012024>
- Garg, S., & Verma, S. (2019). A comparative study of evolutionary methods for feature selection in sentiment analysis. *IJCCI 2019 - Proceedings of the 11th International Joint Conference on Computational Intelligence, Ijcci*, 131–138. <https://doi.org/10.5220/0007948201310138>
- Garla, V., Taylor, C., & Brandt, C. (2013). Semi-supervised clinical text classification with Laplacian SVMs: An application to cancer case management. *Journal of Biomedical Informatics*, 46(5), 869–875. <https://doi.org/10.1016/j.jbi.2013.06.014>
- Geetharamani, R. (2016). Identification of Emotions in Text Articles through Data Pre-

- Processing and Data Mining Techniques. *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, 978, 611–615.
- Ghiassi, M., & Lee, S. (2018). A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Systems with Applications*, *106*, 197–216. <https://doi.org/10.1016/j.eswa.2018.04.006>
- Ghorpade, M. B., & Navale, V. (2021). an Approach for Aspect Based Sentiment Classification Using Machine Learning Algorithms. *International Journal*, *6*(3), 130–134.
http://www.ijasret.com/VolumeArticles/FullTextPDF/735_21.AN_APPROACH_FOR_OR_ASPECT_BASED_SENTIMENT_CLASSIFICATION_USING_MACHINE_LEARNING_ALGORITHMS.pdf
- Ghosh, M., Guha, R., Sarkar, R., & Abraham, A. (2020). A wrapper-filter feature selection technique based on ant colony optimization. *Neural Computing and Applications*, *32*(12), 7839–7857. <https://doi.org/10.1007/s00521-019-04171-3>
- Gokalp, O., Tasci, E., & Ugur, A. (2020). A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. *Expert Systems with Applications*, *146*, 113176. <https://doi.org/10.1016/j.eswa.2020.113176>
- Goudjil, M., Koudil, M., Bedda, M., & Ghoggali, N. (2018). A Novel Active Learning Method Using SVM for Text Classification. *International Journal of Automation and Computing*, *15*(3), 290–298. <https://doi.org/10.1007/s11633-015-0912-z>
- Gupta, I., & Joshi, N. (2020). Enhanced twitter sentiment analysis using hybrid approach and by accounting local contextual semantic. *Journal of Intelligent Systems*, *29*(1), 1611–1625. <https://doi.org/10.1515/jisys-2019-0106>
- Gustafsson, M. (2018). *Sentiment Analysis for Tweets*.
http://www2.aueb.gr/users/ion/docs/korovesis_msc_thesis.pdf
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1–3), 389–422. <https://doi.org/10.1023/A:1012487302797>
- HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE*, *15*(5), 1–22. <https://doi.org/10.1371/journal.pone.0232525>
- Hai, N. T., Nghia, N. H., Le, T. D., & Nguyen, V. T. (2016). A Hybrid Feature Selection Method for Vietnamese Text Classification. *Proceedings - 2015 IEEE International Conference on Knowledge and Systems Engineering, KSE 2015*, 91–

96. <https://doi.org/10.1109/KSE.2015.25>
- Hailong, Z., Wenyan, G., & Bo, J. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. *Proceedings - 11th Web Information System and Application Conference, WISA 2014*, 262–265.
<https://doi.org/10.1109/WISA.2014.55>
- Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. (2014). Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach. *Proceedings - 2014 6th International Conference on Information Technology and Electrical Engineering: Leveraging Research and Technology Through University-Industry Collaboration, ICITEE 2014*, 0–3.
<https://doi.org/10.1109/ICITEED.2014.7007894>
- Haque, T. U., Saber, N. N., & Shah, F. muhammad. (2018). Sentiment Analysis on Large Scale Amazon Product Reviews. *2018 IEEE International Conference on Innovative Research and Development (ICIRD), May*, 85–88.
- Harer, S., & Kadam, S. (2014). Sentiment Classification and Feature based Summarization of Movie Reviews in Mobile Environment. *International Journal of Computer Applications*, 100(1), 30–35. <https://doi.org/10.5120/17490-8008>
- Harish Rao M, S. D. . (2017). Automatic Product Review Sentiment Analysis Using Vader and Feature Visulaization. *International Journal of Computer Science Engineering and Information Technology Research*, 7(4), 53–66.
<https://doi.org/10.24247/ijcseitraug20178>
- Hassan, P. A. (2013). A Hybrid Feature Selection approach of ensemble multiple Filter methods and wrapper method for Improving the Classification Accuracy of Microarray Data Set. *IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS)*, 3(2), 185–190.
- Hassan, S.-U., Saleem, A., Soroya, S. H., Safder, I., Iqbal, S., Jamil, S., Bukhari, F., Aljohani, N. R., & Nawaz, R. (2020). Sentiment analysis of tweets through Altmetrics: A machine learning approach. *Journal of Information Science*, 016555152093091. <https://doi.org/10.1177/0165551520930917>
- Hassler, M., & Fliedl, G. (2006). Text preparation through extended tokenization. *WIT Transactions on Information and Communication Technologies*, 37, 13–21.
<https://doi.org/10.2495/DATA060021>
- Hassonah, M. A., Al-Sayyed, R., Rodan, A., Al-Zoubi, A. M., Aljarah, I., & Faris, H. (2020). An efficient hybrid filter and evolutionary wrapper approach for sentiment

- analysis of various topics on Twitter. *Knowledge-Based Systems*, 192, 105353. <https://doi.org/10.1016/j.knosys.2019.105353>
- Hathaway, N. (2018). *Generalizing Sentiment Analysis Techniques Across Sub-Categories of IMDB Movie Reviews*. <https://pdfs.semanticscholar.org/4f23/ec6b9f17bb115b2e4b8f98fa3bc884a0ee3ff>
- He, R., & McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. *25th International World Wide Web Conference, WWW 2016*, 507–517. <https://doi.org/10.1145/2872427.2883037>
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining : A case study in the pizza industry. *International Journal of Information Management*, 33, 464–472. <https://doi.org/10.1016/j.ijinfomgt.2013.01.001>
- Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, 52(3), 1495–1545. <https://doi.org/10.1007/s10462-017-9599-6>
- Hira, Z. M., & Gillies, D. F. (2015). A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Advances in Bioinformatics*, 2015(1), 1–14. <https://doi.org/10.1155/2015/198363>
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. <https://doi.org/10.1145/1014052.1014073>
- Hu, Y. H., Chen, Y. L., & Chou, H. L. (2017). Opinion mining from online hotel reviews – A text summarization approach. *Information Processing and Management*, 53(2), 436–449. <https://doi.org/10.1016/j.ipm.2016.12.002>
- Huang, M. L., Hung, Y. H., Lee, W. M., Li, R. K., & Jiang, B. R. (2014). SVM-RFE based feature selection and taguchi parameters optimization for multiclass SVM Classifier. *Scientific World Journal*, 2014. <https://doi.org/10.1155/2014/795624>
- Hung, L. P., Alfred, R., & Hijazi, M. H. A. (2015). A review on feature selection methods for sentiment analysis. *Advanced Science Letters*, 21(10), 2952–2956. <https://doi.org/10.1166/asl.2015.6475>
- Hussain, I., Ghous, H., & Rizvi, T. Z. (2021). A Novel Hybrid Framework For Twitter Sentiment Analysis Using Feature Selection Method Isomap and Machine Learning Methods. *INTERNATIONAL JOURNAL OF SCIENTIFIC PROGRESS AND RESEARCH (IJSPR)*, 79(179), 72–80.

- Hussein, M., & Özyurt, F. (2021). A New Technique for Sentiment Analysis System Based on Deep Learning Using Chi- Square Feature Selection Method. *BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING*, 9(4), 320–326. <https://doi.org/10.17694/bajece.887339>
- Ibrahim, R., Yousefpour, A., Nuzly, H., Hamed, A., & Hajmohammadi, M. S. (2014). A Comparative Study on Sentiment Analysis. *AENSI Journals Adv. Environ. Biol*, 8(13), 53–68.
- Iqbal, F., Hashmi, J. M., Fung, B. C. M., Batool, R., Khattak, A. M., Aleem, S., & Hung, P. C. K. (2019). A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction. *IEEE Access*, 7(c), 14637–14652. <https://doi.org/10.1109/ACCESS.2019.2892852>
- Isa, D., Lee, L. H., Kallimani, V. P., & RajKumar, R. (2008). Text Document Preprocessing with the Bayes Formula for Classification Using the Support Vector Machine. *IEEE Trans. on Knowl. and Data Eng.*, 20(9), 1264–1272. <https://doi.org/http://dx.doi.org/10.1109/TKDE.2008.76>
- Isa, N., Puteh, M., & Kamarudin, R. M. H. R. M. H. R. (2013). Sentiment classification of malay newspaper using immune network (SCIN). *Proceedings of the World Congress on Engineering 2013*, 3, 1543–1548. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84887882006&partnerID=40&md5=652fdc713458c4dfedcbc4e3a0b736b6%5Cnh> <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84887882006&partnerID=40&md5=652fdc713458c4dfedcbc4e3a0b736b6>
- Isabelle, G., Maharani, W., & Asror, I. (2019). Analysis on Opinion Mining Using Combining Lexicon-Based Method and Multinomial Naïve Bayes. *International Conference on Industrial Enterprise and System Engineering (IcoIESE 2018)*, 2(IcoIESE 2018), 214–219. <https://doi.org/10.2991/icoiese-18.2019.38>
- Islam, M., Anjum, A., Ahsan, T., & Wang, L. (2019). Dimensionality Reduction for Sentiment Classification using Machine Learning Classifiers. *2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019, Mi*, 3097–3103. <https://doi.org/10.1109/SSCI44817.2019.9002967>
- Islam, M. S., Md Jubayer, F. E., & Ahmed, S. I. (2017). A support vector machine mixed with TF-IDF algorithm to categorize Bengali document. *ECCE 2017 - International Conference on Electrical, Computer and Communication Engineering*, 191–196. <https://doi.org/10.1109/ECACE.2017.7912904>
- Jabri, S., Dahbi, A., Gadi, T., & Bassir, A. (2018). Ranking of text documents using TF-IDF weighting and association rules mining. *Proceedings of the 2018*

- International Conference on Optimization and Applications, ICOA 2018*, 1–6.
<https://doi.org/10.1109/ICOA.2018.8370597>
- Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2017). Sentiment Analysis of Events from Twitter Using Open Source Tool. *International Journal of Computer Science and Mobile Computing*, May 2016.
- Jain, D., Vardhan, B. H., & Kandasamy, S. (2019). Sentiment analysis of product reviews – A survey. *International Journal of Scientific and Technology Research*, 8(12), 2452–2459.
- Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, 41, 100413.
<https://doi.org/10.1016/j.cosrev.2021.100413>
- 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>
- Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep Convolution Neural Networks for Twitter Sentiment Analysis. *IEEE Access*, 6(c), 23253–23260.
<https://doi.org/10.1109/ACCESS.2017.2776930>
- Jing, L.-P., Huang, H.-K., & Shi, H.-B. (2002). IMPROVED FEATURE SELECTION APPROACH TFIDF IN TEXT MINING. *Proceedings of the First International Conference on Machine Learning and Cybernetics, November*, 4–5.
- Jotheeswaran, J., & Koteeswaran, S. (2016). Feature selection using random forest method for sentiment analysis. *Indian Journal of Science and Technology*, 9(3), 1–7. <https://doi.org/10.17485/ijst/2016/v9i3/86387>
- 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>
- Kadhun, M., Manaseer, S., & Dalhoum, A. L. A. (2021). Evaluation feature selection technique on classification by using evolutionary ELM wrapper method with features priorities. *Journal of Advances in Information Technology*, 12(1), 21–28.
<https://doi.org/10.12720/jait.12.1.21-28>
- Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2010a). Comparative Study of Attribute Selection Using Gain Ratio and Correlation Based Feature Selection. *International Journal of Information Technology and Knowledge Management*,

2(2), 271–277.

- Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2010b). Feature Subset Selection Problem using Wrapper Approach in Supervised Learning. *International Journal of Computer Applications*, 1(7), 13–17. <https://doi.org/10.5120/169-295>
- Kaur, F., & Bhatia, R. (2016). Sentiment Analyzing by Dictionary based Approach. *International Journal of Computer Applications*, 152(5), 32–34. <https://doi.org/10.5120/ijca2016911814>
- Kaur, H., Mangat, V., & Nidhi. (2017). A survey of sentiment analysis techniques. *Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2017*, 921–925. <https://doi.org/10.1109/I-SMAC.2017.8058315>
- Khader, M., Awajan, A., & Al-Naymat, G. (2019). The Effects of Natural Language Processing on Big Data Analysis: Sentiment Analysis Case Study. *ACIT 2018 - 19th International Arab Conference on Information Technology*, 1–7. <https://doi.org/10.1109/ACIT.2018.8672697>
- Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM : Twitter opinion mining framework using hybrid classification scheme. *Decision Supp. Syst.*, 57(January), 245–257.
- Khan, F. H., Qamar, U., & Bashir, S. (2016). SWIMS: Semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis. *Knowledge-Based Systems*, 100, 97–111. <https://doi.org/10.1016/j.knosys.2016.02.011>
- Kharde, V. A. (2016). Sentiment Analysis of Twitter Data : A Survey of Techniques. *International Journal of Computer Applications*, 139(11), 5–15.
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text Mining in Organizational Research. In *Organizational Research Methods* (Vol. 21, Issue 3). <https://doi.org/10.1177/1094428117722619>
- Kolchyna, O., Souza, T. T. P., Treleaven, P., & Aste, T. (2015). Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination. In *Handbook of Sentiment Analysis in Finance*. <http://arxiv.org/abs/1507.00955>
- Kolekar, N. V., Rao, G., Dey, S., Mane, M., Jadhav, V., & Patil, S. (2016). Sentiment analysis and classification using lexicon-based approach and addressing polarity shift problem. *Journal of Theoretical and Applied Information Technology*, 90(1), 118–125.
- Koppel, M., & Schler, J. (2006). The importance of neutral examples for learning

- sentiment. *Computational Intelligence*, 22(2), 100–109.
<https://doi.org/10.1111/j.1467-8640.2006.00276.x>
- Korovkinas, K., Danėnas, P., & Garšva, G. (2019). SVM and k-Means Hybrid Method for Textual Data Sentiment Analysis. *Baltic Journal of Modern Computing*, 7(1).
<https://doi.org/10.22364/bjmc.2018.7.1.04>
- Kotzias, D., Denil, M., De Freitas, N., & Smyth, P. (2015). From group to individual labels using deep features. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015-Augus*, 597–606.
<https://doi.org/10.1145/2783258.2783380>
- Kumar, H. M. K., & Harish, B. S. (2020). A New Feature Selection Method for Sentiment Analysis in Short Text. *Journal of Intelligent Systems*, 29(1), 1122–1134. <https://doi.org/10.1515/jisys-2018-0171>
- Kumar, K., Harish, B. S., & Darshan, H. K. (2018). Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(5), 109.
<https://doi.org/10.9781/ijimai.2018.12.005>
- Kumar, S., Gahalawat, M., Roy, P. P., Dogra, D. P., & Kim, B. G. (2020). Exploring impact of age and gender on sentiment analysis using machine learning. *Electronics (Switzerland)*, 9(2), 1–14. <https://doi.org/10.3390/electronics9020374>
- Kumar, S. S., & Rajini, A. (2019). Extensive Survey on Feature Extraction and Feature Selection Techniques for Sentiment Classification in Social Media. *10th ICCCNT 2019*, 6–11.
- Kumar, V. (2014). Feature Selection: A literature Review. *The Smart Computing Review*, 4(3). <https://doi.org/10.6029/smarterc.2014.03.007>
- Kurniawati, I., & Pardede, H. F. (2018). Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis. *2018 International Conference on Information Technology Systems and Innovation, ICITSI 2018 - Proceedings*, 1–5.
<https://doi.org/10.1109/ICITSI.2018.8695953>
- Larasati, U. I., Muslim, M. A., Arifudin, R., & Alamsyah, A. (2019). Improve the Accuracy of Support Vector Machine Using Chi Square Statistic and Term Frequency Inverse Document Frequency on Movie Review Sentiment Analysis. *Scientific Journal of Informatics*, 6(1), 138–149.
<https://doi.org/10.15294/sji.v6i1.14244>
- Lee, C., & Lee, C. (2017). An Information-Theoretic Filter Approach for Bayes

- Reference : Weighted Classification Learning in Naive Bayes. *Data & Knowledge Engineering*. <https://doi.org/10.1016/j.datak.2017.11.002>
- Li, R., & Zhang, Y. (2012). Study on the method of feature selection based on hybrid model for text classification. *Advanced Materials Research*, 433–440, 2881–2886. <https://doi.org/10.4028/www.scientific.net/AMR.433-440.2881>
- Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. In *Artificial Intelligence Review* (Vol. 54, Issue 7). Springer Netherlands. <https://doi.org/10.1007/s10462-021-09973-3>
- Lin, X., Li, C., Zhang, Y., Su, B., Fan, M., & Wei, H. (2018). Selecting Feature Subsets Based on SVM-RFE and the Overlapping Ratio with Applications in Bioinformatics. *Molecules*. <https://doi.org/10.3390/molecules23010052>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. In *SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES*.
- Liu, L., Kang, J., Yu, J., & Wang, Z. (2005). A comparative study on unsupervised feature selection methods for text clustering. *Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering, IEEE NLP-KE'05, 2005*, 597–601. <https://doi.org/10.1109/NLPKE.2005.1598807>
- Liu, R., & Gillies, D. F. (2016). Overfitting in linear feature extraction for classification of high-dimensional image data. *Pattern Recognition*, 53, 73–86. <https://doi.org/10.1016/j.patcog.2015.11.015>
- Liu, Y., Bi, J. W., & Fan, Z. P. (2017). Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms. *Expert Systems with Applications*, 80, 323–339. <https://doi.org/10.1016/j.eswa.2017.03.042>
- Luhn, H. P. (1957). A Statistical Approach to Mechanized Encoding and Searching of Literary Information. *IBM Journal of Research and Development*, 1(4), 309–317. <https://doi.org/10.1147/rd.14.0309>
- Luo, F., Li, C., & Cao, Z. (2016). Affective-feature-based sentiment analysis using SVM classifier. *Proceedings of the 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2016*, 276–281. <https://doi.org/10.1109/CSCWD.2016.7566001>
- Luo, M., & Luo, Li. (2010). Feature Selection for Text Classification using OR+SVM-RFE. *2010 Chinese Control and Decision Conference, Xuzhou, China*, 16, 257–274. <https://doi.org/10.1177/058310248902100419>

- M, M., & Mehla, S. (2019). Sentiment Analysis of Movie Reviews using Machine Learning Classifiers. *International Journal of Computer Applications*, 182(50), 25–28. <https://doi.org/10.5120/ijca2019918756>
- Ma, L. (2017). *A Multi-label Text Classification Framework : Using Supervised and Unsupervised Feature Selection Strategy*.
- Madasu, A., & Elango, S. (2019). Efficient feature selection techniques for sentiment analysis. *Multimedia Tools and Applications*, 79, 6313–6335. <https://doi.org/https://doi.org/10.1007/s11042-019-08409-z> Efficient
- Mafarja, M., & Mirjalili, S. (2018). Whale optimization approaches for wrapper feature selection. *Applied Soft Computing*, 62, 441–453. <https://doi.org/10.1016/j.asoc.2017.11.006>
- Mahek Fatima, Q., & Mustakim, S. (2020). Sentiment Analysis and Opinion Mining on Products, Movies, Tourism Reviews Using Machine Learning Techniques. *International Journal of Engineering Applied Sciences and Technology*, 5(6), 148–153. <https://doi.org/10.33564/ijeast.2020.v05i06.020>
- Majeed, A. (2019). Improving Time Complexity and Accuracy of the Machine Learning Algorithms Through Selection of Highly Weighted Top k Features from Complex Datasets. *Annals of Data Science*, 6(4), 599–621. <https://doi.org/10.1007/s40745-019-00217-4>
- Maldonado, S., & Weber, R. (2009). A wrapper method for feature selection using Support Vector Machines. *Information Sciences*, 179(13), 2208–2217. <https://doi.org/10.1016/j.ins.2009.02.014>
- Manda, K. R. (2019). *Sentiment Analysis of Twitter Data Using Machine Learning and Deep Learning Methods* (Issue June). <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1335995>
- Manek, A. S., Shenoy, P. D., Mohan, M. C., & Venugopal, K. R. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World Wide Web*, 20(2), 135–154. <https://doi.org/10.1007/s11280-015-0381-x>
- Matsuo, R., & Ho, T. B. (2018). Semantic Term Weighting for Clinical Texts. *Expert Systems With Applications*. <https://doi.org/10.1016/j.eswa.2018.08.028>
- Méndez, J. R., Cotos-Yañez, T. R., & Ruano-Ordás, D. (2019). A new semantic-based feature selection method for spam filtering. *Applied Soft Computing Journal*, 76, 89–104. <https://doi.org/10.1016/j.asoc.2018.12.008>

- Mhatre, M., Phondekar, D., Kadam, P., Chawathe, A., & Ghag, K. (2018). Dimensionality reduction for sentiment analysis using pre-processing techniques. In *Proceedings of the International Conference on Computing Methodologies and Communication, ICCMC 2017* (Vols. 2018-Janua, Issue Iccmc, pp. 16–21). <https://doi.org/10.1109/ICCMC.2017.8282676>
- Mlambo, N, W. C. (2016). A survey and comparative study of filter and wrapper feature selection techniques. *The International Journal Of Engineering And Science (IJES)*, 5(8), 57–67. <https://pdfs.semanticscholar.org/fe2c/a244dd5473270ce0a8e2a387349b9831cf47.pdf>
- Mohammad, S. (2018). *Corpus-based Construction of Sentiment Lexicon to Identify Extreme Opinions by Supervised and Unsupervised Machine learning Methods*. <https://citius.usc.es/sites/default/files/tesis/Thesis.pdf>
- Mohan, S., & Sunitha, R. (2020). Survey On Aspect Based Sentiment Analysis Using Machine Learning Techniques. *European Journal of Molecular & Clinical Medicine*, 07(10).
- Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(2), 621–633. <https://doi.org/10.1016/j.eswa.2012.07.059>
- Mountassir, A., Benbrahim, H., & Berrada, I. (2012). A cross-study of sentiment classification on arabic corpora. *Res. and Dev. in Intelligent Syst. XXIX: Incorporating Applications and Innovations in Intel. Sys. XX - AI 2012, 32nd SGAI Int. Conf. on Innovative Techniques and Applications of Artificial Intel.*, 259–272. <https://doi.org/10.1007/978-1-4471-4739-8-21>
- Mtetwa, N., Awukam, A. O., & Yousefi, M. (2018). Feature Extraction and Classification of Movie Reviews. *5th International Conference on Soft Computing and Machine Intelligence, ISCFMI 2018*, 67–71. <https://doi.org/10.1109/ISCFMI.2018.8703235>
- Muhammad, A. (2016). Contextual lexicon-based sentiment analysis for social media. *PQDT - UK & Ireland*.
- Mukhtar, N., Khan, M. A., & Chiragh, N. (2018). Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains. *Telematics and Informatics*, 35(8), 2173–2183. <https://doi.org/10.1016/j.tele.2018.08.003>
- Muthia, D. A. (2014). Sentiment Analysis of Hotel Review Using Naïve Bayes

- Algorithm and Integration of Information Gain and Genetic Algorithm As Feature Selection Methods. *International Seminar on Scientific Issues and Trends (ISSIT)*, 25–30.
- Muthia, D. A., Putri, D. A., Rachmi, H., & Surniandari, A. (2019). Implementation of Text Mining in Predicting Consumer Interest on Digital Camera Products. *2018 6th International Conference on Cyber and IT Service Management, CITSM 2018, Citsm*, 1–7. <https://doi.org/10.1109/CITSM.2018.8674063>
- Mutinda, J. (2021). Lexicon-pointed hybrid N-gram Features Extraction Model (LeNFEM) for sentence level sentiment analysis. *Wiley, September 2020*, 1–17. <https://doi.org/10.1002/eng2.12374>
- Nam, L. N. H., & Quoc, H. B. (2017). The Hybrid Filter Feature Selection Methods for Improving High-Dimensional Text Categorization. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 25(2), 235–265. <https://doi.org/10.1142/S021848851750009X>
- Naqvi, G. (2012). A Hybrid Filter-Wrapper Approach for Feature Selection. In *Örebro University*.
- Nasser, A., Dincer, K., & Sever, H. (2016). Investigation of the Feature Selection Problem for Sentiment Analysis in Arabic Language. *Research in Computing Science*, 110(1), 41–54. <https://doi.org/10.13053/rcs-110-1-4>
- Navaney, P., & Rana, A. (2018). SMS Spam Filtering using Supervised Machine Learning Algorithms. *2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 43–48.
- Naz, S., Sharan, A., & Malik, N. (2019). Sentiment Classification on Twitter Data Using Support Vector Machine. *Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018*, 676–679. <https://doi.org/10.1109/WI.2018.00-13>
- Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2010). Recognition of affect, judgment, and appreciation in text. *Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference*, 2(May), 806–814.
- Nguyen, H., Al, R., & Academy, K. (2018). Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches. *SMU Data Science Review*, 1(4).
- Nie, F., Huang, H., Cai, X., & Ding, C. H. (2010). Efficient and robust feature selection via joint ℓ_2 , 1-norms minimization. *Advances in Neural Information Processing Systems*, 23, 1813–1821.

- Nohh, N. H., Megat, N., & Zainuddin, M. (2019). Sentiment Analysis towards Hotel Reviews. *Open International Journal Of Informatics(OIJI)*, 7(2), 1–19.
- Omar, N., Albared, M., Al-Moslmi, T., & Al-Shabi, A. (2014). A comparative study of feature selection and machine learning algorithms for arabic sentiment classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8870, 429–443. https://doi.org/10.1007/978-3-319-12844-3_37
- Omar, N., Albared, M., Al-Moslmi, T., & Al-Shabi, A. Q. (2013). Ensemble of Classification Algorithms for Subjectivity and Sentiment Analysis of Arabic Customers' Reviews Natural Language Processing View project Cross-domain sentiment analysis View project Ensemble of Classification Algorithms for Subjectivity and Sen. *Article in International Journal of Advancements in Computing Technology*, 5(October), 77–85. <https://www.researchgate.net/publication/258218164>
- Palad, E. B. B., Kyaw, K. S., Sattayaraksa, T., Tepsongkroh, P., & Sasha, F. (2021). Hybrid Feature Selection-Based Sentiment Analysis Concept : Case-Study of E-Commerce Opinion Mining Model. *The 12th Hatyai National and International Conference*, 1897–1912.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.3748/wjg.v22.i45.9898>
- Pant, S., & Jain, K. (2017). Sentiment Analysis Using Feature Selection and Classification Algorithms - A Survey. *International Journal of Innovations in Engineering Research and Technology (IJIERT)*, 4(3), 109–113.
- Papacharissi, Z., Curty, R. G., Zhang, P., Verma, J. P., Patel, B., Patel, A., Wong, W. P.-M., Lo, M.-C., Ramayah, T., Limón, M., Mason, L., Parhami, B., Sa, S., Cheok, A. D., Goh, K. H., Liu, W., Farbiz, F., Fong, S. W., Teo, S. L., ... Khan, A. (2013). Sentiment Analysis of Roman Urdu/Hindi using supervised methods. *Ain Shams Engineering Journal*, 2(3), 1093–1113.
- Parveen, R., Shrivastava, N., & Tripathi, P. (2020). Sentiment classification of movie reviews by supervised machine learning approaches using ensemble learning voted algorithm. *2nd International Conference on Data, Engineering and Applications, IDEA 2020*, 4(4), 285–292. <https://doi.org/10.1109/IDEA49133.2020.9170684>
- Patil, L. H., & Atique, M. (2013). A novel approach for feature selection method TF-IDF in document clustering. *Proceedings of the 2013 3rd IEEE International Advance Computing Conference, IACC, Ghaziabad, India*, 858–862.

- Paudel, S., Prasad, P. W. C., Alsadoon, A., Islam, M. R., & Elchouemi, A. (2019). Feature selection approach for twitter sentiment analysis and text classification based on chi-square and naïve bayes. *Advances in Intelligent Systems and Computing*, 842, 281–298. https://doi.org/10.1007/978-3-319-98776-7_30
- Paul, A. K., & Shill, P. C. (2017). Sentiment mining from Bangla data using mutual information. *ICECTE 2016 - 2nd International Conference on Electrical, Computer and Telecommunication Engineering, December*, 8–10. <https://doi.org/10.1109/ICECTE.2016.7879569>
- Pawar, K. K., Shrishrimal, P. P., & Deshmukh, R. R. (2015). Twitter Sentiment Analysis: A Review. *International Journal of Scientific & Engineering Research*, 6(4), 957–964.
- Pes, B., Dessì, N., & Angioni, M. (2017). Exploiting the ensemble paradigm for stable feature selection: A case study on high-dimensional genomic data. *Information Fusion*, 35, 132–147. <https://doi.org/10.1016/j.inffus.2016.10.001>
- Porter, M. . (1980). An algorithm for suffix stripping. *Program: Electronic Library and Information System*, 14(3), 313–316. <https://doi.org/https://doi.org/10.1108/eb046814>
- Pradha, S., Halgamuge, M. N., & Tran Quoc Vinh, N. (2019). Effective text data preprocessing technique for sentiment analysis in social media data. *Proceedings of 2019 11th International Conference on Knowledge and Systems Engineering, KSE 2019*, 1–8. <https://doi.org/10.1109/KSE.2019.8919368>
- Prakash, N. (2019). A Comparative study of Lexicon based and Machine learning based classifications in Sentiment analysis. *International Journal of Data Mining Techniques and Applications, December*, 8–13. <https://doi.org/10.20894/IJDMTA.102.008.001.005>
- Pratiwi, A. I., & Adiwijaya. (2018). On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis. *Applied Computational Intelligence and Soft Computing, 2018*. <https://doi.org/10.1155/2018/1407817>
- Price, T. (2018). *Evaluation of Feature Selection Methods for Machine Learning Classification of Breast Cancer*. *Evaluation of Feature Selection Methods for Machine Learning Classification of Breast Cancer*.
- Prusa, J. D., Khoshgoftaar, T. M., & Dittman, D. J. (2015). Impact of feature selection techniques for tweet sentiment classification. *Proceedings of the 28th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2015*, 299–304.

- Pujari, C., Aiswarya, & Shetty, N. P. (2018). Comparison of classification techniques for feature oriented sentiment analysis of product review data. *Advances in Intelligent Systems and Computing*, 542, 149–158. https://doi.org/10.1007/978-981-10-3223-3_14
- Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-Based Sentiment Analysis of Teachers' Evaluation. *Applied Computational Intelligence and Soft Computing*, 2016, 1–12. <https://doi.org/10.1155/2016/2385429>
- Rajput, R., & Solanki, A. K. (2016). Review of Sentimental Analysis Methods using Lexicon Based Approach. *International Journal of Computer Science and Mobile Computing*, 5(2), 159–166. www.ijcsmc.com
- Ramesh, B., & Sathiaselvan, J. G. R. (2015). An Advanced Multi Class Instance Selection based Support Vector Machine for Text Classification. *Procedia Computer Science*, 57, 1124–1130. <https://doi.org/10.1016/j.procs.2015.07.400>
- Rana, S., & Singh, A. (2017). Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques. *Proceedings on 2016 2nd International Conference on Next Generation Computing Technologies, NGCT 2016, October*, 106–111. <https://doi.org/10.1109/NGCT.2016.7877399>
- Rani, S., Singh Gill, N., & Gulia, P. (2021). Analyzing impact of number of features on efficiency of hybrid model of lexicon and stack based ensemble classifier for twitter sentiment analysis using WEKA tool. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 1041. <https://doi.org/10.11591/ijeecs.v22.i2.pp1041-1051>
- Rasjid, Z. E., & Setiawan, R. (2017). Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques. *Procedia Computer Science*, 116, 107–112. <https://doi.org/10.1016/j.procs.2017.10.017>
- Rasool, A., Tao, R., Kamyab, M., & Hayat, S. (2020). GAWA—A Feature Selection Method for Hybrid Sentiment Classification. *IEEE Access*, 8, 191850–191861. <https://doi.org/10.1109/access.2020.3030642>
- Raza, H., Faizan, M., Hamza, A., Mushtaq, A., & Akhtar, N. (2019). Scientific text sentiment analysis using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 10(12), 157–165. <https://doi.org/10.14569/ijacsa.2019.0101222>
- Rehman, A., Javed, K., & Babri, H. A. (2017). Feature selection based on a normalized difference measure for text classification. *Information Processing & Management*,

- 53(2), 473–489. <https://doi.org/10.1016/j.ipm.2016.12.004>
- Rehman, A. U., Malik, A. K., Raza, B., & Ali, W. (2019). A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. *Multimedia Tools and Applications*, 78(18), 26597–26613. <https://doi.org/10.1007/s11042-019-07788-7>
- Ren, F., & Sohrab, M. G. (2013). Class-indexing-based term weighting for automatic text classification. *Information Sciences*, 236, 109–125. <https://doi.org/10.1016/j.ins.2013.02.029>
- Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1), 1–29. <https://doi.org/10.1140/epjds/s13688-016-0085-1>
- Rosita Kamala, F., & Ranjit Jeba Thangaiah, P. (2019). An improved hybrid feature selection method for huge dimensional datasets. *IAES International Journal of Artificial Intelligence*, 8(1), 77–86. <https://doi.org/10.11591/ijai.v8.i1.pp77-86>
- Rouby, I., Badawy, M., Nour, M., & Hegazi, N. (2018). Performance evaluation of an adopted sentiment analysis model for arabic comments from the facebook. *Journal of Theoretical and Applied Information Technology*, 96(21), 7098–7112.
- Ruangkanokmas, P., Achalakul, T., Akkarajitsakul, K., Sharma, S., Jain, A., Uysal, A. K., & Murphey, Y. L. (2017). Sentiment Classification: Feature Selection Based Approaches Versus Deep Learning. *IEEE CIT 2017 - 17th IEEE International Conference on Computer and Information Technology*, 0(2), 23–30. <https://doi.org/10.4018/ijirr.2020040103>
- Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS ONE*, 16(2), 1–23. <https://doi.org/10.1371/journal.pone.0245909>
- Sabbah, T., Ayyash, M., & Ashraf, M. (2018). Hybrid support vector machine based feature selection method for text classification. *International Arab Journal of Information Technology*, 15(3A Special Issue), 599–609.
- Saberi, B., & Saad, S. (2017). Sentiment analysis or opinion mining: A review. *International Journal on Advanced Science, Engineering and Information Technology*, 7(5), 1660–1666. <https://doi.org/10.18517/ijaseit.7.5.2137>
- Sadia, A., Khan, F., & Bashir, F. (2018). An overview of lexicon-based approach for sentiment analysis. *International Electrical Engineering Conference, IEEEC*, 1–6.

- Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), 2507–2517. <https://doi.org/10.1093/bioinformatics/btm344>
- Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), 2507–2517. <https://doi.org/10.1093/bioinformatics/btm344>
- Sahu, T. P., & Ahuja, S. (2016). Sentiment analysis of movie reviews: A study on feature selection and classification algorithms. *2016 International Conference on Microelectronics, Computing and Communications (MicroCom), Durgapur, India*, 1–6.
- Saif, H., Fernandez, M., He, Y., & Alani, H. (2014). On stopwords, filtering and data sparsity for sentiment analysis of twitter. *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, 810–817.
- Salas-Zárate, M. D. P., Paredes-Valverde, M. A., Limon-Romero, J., Tlapa, D., & Baez-Lopez, Y. (2016). Sentiment classification of spanish reviews: An approach based on feature selection and machine learning methods. *Journal of Universal Computer Science*, 22(5), 691–708.
- Samal, B. R., Behera, A. K., & Panda, M. (2017). Performance analysis of supervised machine learning techniques for sentiment analysis. *Proceedings of 2017 3rd IEEE International Conference on Sensing, Signal Processing and Security, ICSSS 2017*, 128–133. <https://doi.org/10.1109/SSPS.2017.8071579>
- Saraee, M., & Bagheri, A. (2013). Feature selection methods in persian sentiment analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7934 LNCS, 303–308. https://doi.org/10.1007/978-3-642-38824-8_29
- Sarvakar, K., & Kuchara, U. K. (2018). Sentiment analysis of movie reviews: A new feature-based sentiment classification. *International Journal of Scientific Research in Computer Science and Engineering*, 6(3), 8–12. <https://doi.org/10.1109/iMac4s.2013.6526500>
- Sasikala, P., & Mary Immaculate Sheela, L. (2020). Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00308-7>
- Selvi, C., Ahuja, C., & Sivasankar, E. (2015). A comparative study of feature selection and machine learning methods for sentiment classification on movie data set. *Advances in Intelligent Systems and Computing*, 343, 367–379.

https://doi.org/10.1007/978-81-322-2268-2_39

- Shamantha Rai, B., Shetty, S. M., & Rai, P. (2019). Sentiment analysis using Machine learning classifiers: Evaluation of performance. *2019 IEEE 4th International Conference on Computer and Communication Systems, ICCCS 2019*, 21–25. <https://doi.org/10.1109/CCOMS.2019.8821650>
- Shang, C., Li, M., Feng, S., Jiang, Q., & Fan, J. (2013). Feature selection via maximizing global information gain for text classification. *Knowledge-Based Systems*, *54*, 298–309. <https://doi.org/10.1016/j.knosys.2013.09.019>
- Sharma, S., & Jain, A. (2020). Hybrid Ensemble Learning With Feature Selection for Sentiment Classification in Social Media. *International Journal of Information Retrieval Research*, *10*(2), 40–58. <https://doi.org/10.4018/ijirr.2020040103>
- Shaukat, Z., Zulfiqar, A. A., Xiao, C., Azeem, M., & Mahmood, T. (2020). Sentiment analysis on IMDB using lexicon and neural networks. *SN Applied Sciences*, *2*(2), 1–10. <https://doi.org/10.1007/s42452-019-1926-x>
- Sheeba, J. I., & Vivekanandan, K. (2013). Low frequency keyword extraction with sentiment classification and cyberbully detection using fuzzy logic technique. *2013 IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC 2013, December*. <https://doi.org/10.1109/ICCIC.2013.6724124>
- Shein, K. P. P., & Nyunt, T. T. S. (2010). Sentiment classification based on ontology and SVM classifier. *2nd International Conference on Communication Software and Networks, ICCSN 2010*, 169–172. <https://doi.org/10.1109/ICCSN.2010.35>
- Shubhangi Patil, S. D., & Deshmukh Professor, R. R. (2014). Review of Twitter sentiment analysis. *International Journal of Scientific & Engineering Research*, *5*(10), 1616–1623. <http://www.ijser.org>
- Siddiqui, M. A. (2016). An empirical evaluation of text classification and feature selection methods. *Artificial Intelligence Research*, *5*(2), 70. <https://doi.org/10.5430/air.v5n2p70>
- Singh, A., Agarwal, A., & Dimri, P. (2018). Comparative Study of Machine Learning Approaches for Amazon Reviews. *Procedia Computer Science*, *132*, 1552–1561. <https://doi.org/10.1016/j.procs.2018.05.119>
- Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human-Centric Computing and Information Sciences*, *7*(1). <https://doi.org/10.1186/s13673-017-0116-3>

- Siregar, A., Faisal, S., Tukino, T., Puspabhuana, A., & Simarangkir, M. (2019). Comparison Study Of Term Weighting Optimally With SVM In Sentiment Analysis. *ICASI 2019*, 1–9. <https://doi.org/10.4108/eai.18-7-2019.2288508>
- Soumya, S., & Pramod, K. . (2020). Sentiment analysis of malayalam tweets using machine learning techniques. *ICT Express*, 6(4), 300–305. <https://doi.org/10.1016/j.ict.2020.04.003>
- Spärck Jones, K. (1972). A Statistical Interpretation of Term Specificity and its Retrieval. *Journal of Documentation*, 28(1), 11–21. <https://doi.org/10.1108/eb026526>
- Srinivasan, R., & Subalalitha, C. N. (2021). Sentimental analysis from imbalanced code-mixed data using machine learning approaches. *Distributed and Parallel Databases*, 0123456789. <https://doi.org/10.1007/s10619-021-07331-4>
- Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment analysis of twitter data: A hybrid approach. *International Journal of Healthcare Information Systems and Informatics*, 14(2), 1–16. <https://doi.org/10.4018/IJHISI.2019040101>
- Stephen, K. V. K., Al-Harthy, F. R. A., & Shaikh, M. T. (2020). An efficient feature selection system for automotive sentiment classification in hadoop framework using naive bayes classifier. *ARNP Journal of Engineering and Applied Sciences*, 15(6), 841–846.
- Suchetha, N. K., Nikhil, A., & Hrudya, P. (2019). Comparing the wrapper feature selection evaluators on twitter sentiment classification. *Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019), Chennai, India*, 1–6. <https://doi.org/10.1109/ICCIDS.2019.8862033>
- Sureja, N. D. (2017). Using Sentimental Analysis Approach Review on Classification of Movie Script. *International Journal of Engineering Development and Research*, 4(2), 616–620.
- Suthaharan, S. (2014). Big data classification : problems and challenges in network intrusion prediction with machine learning. *Performance Evaluation Review*, 41(4), 70–73. <https://doi.org/10.1145/2627534.2627557>
- Syamala, M., & Nalini, N. J. (2020). A filter based improved decision tree sentiment classification model for real-time amazon product review data. *International Journal of Intelligent Engineering and Systems*, 13(1), 191–202. <https://doi.org/10.22266/ijies2020.0229.18>
- Symeonidis, S., Effrosynidis, D., & Arampatzis, A. (2018). A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis.

- Expert Systems With Applications*, 110, 298–310.
<https://doi.org/10.1016/j.eswa.2018.06.022>
- Taj, S., Shaikh, B. B., & Fatemah Meghji, A. (2019). Sentiment analysis of news articles: A lexicon based approach. *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies, ICoMET 2019, February*. <https://doi.org/10.1109/ICOMET.2019.8673428>
- Tang, B., Kay, S., & He, H. (2016). Toward Optimal Feature Selection in Naive Bayes for Text Categorization. *IEEE Transactions on Knowledge and Data Engineering*, 28(9), 2508–2521. <https://doi.org/10.1109/TKDE.2016.2563436>
- Tang, Y., Zhang, Y.-Q., & Huang, 'zhen. (2007). Development of Two-Stage SVM-RFE Gene Selection Strategy for Microarray Expression Data Analysis. *IEEE Transaction on Computational Biology and Bioinformatics*, 4(3), 365–381. <https://doi.org/10.1109/TCBB.2007.1028>
- Tao, J., & Fang, X. (2020). Toward multi-label sentiment analysis: a transfer learning based approach. *Journal of Big Data*, 7(1), 1–26. <https://doi.org/10.1186/s40537-019-0278-0>
- Taylor, P., Imani, M. B., Keyvanpour, M. R., & Azmi, R. (2013). A Novel Embedded Feature Selection Method : A Comparative Study in The Application of Text Categorization. *Applied Artificial Intelligence*, 27(June), 408–427. <https://doi.org/10.1080/08839514.2013.774211>
- Tedmori, S., & Awajan, A. (2019). Sentiment analysis main tasks and applications: A survey. *Journal of Information Processing Systems*, 15(3), 500–519. <https://doi.org/10.3745/JIPS.04.0120>
- Thomas, M., Student, Mt., & Professor, A. (2015). Comparison of Lexicon based and Naïve Bayes Classifier in Sentiment Analysis. *IJSRD-International Journal for Scientific Research & Development*, 3(04), 2321–0613. www.ijssrd.com
- Timmaraju, A., & Khanna, V. (2015). Sentiment Analysis on Movie Reviews using Recursive and Recurrent Neural Network Architectures. *CS224N Projects*, 1(10), 2–7.
- Tripathi, G., & S, N. (2015). Feature Selection and Classification Approach for Sentiment Analysis. *Machine Learning and Applications: An International Journal*, 2(2), 01–16. <https://doi.org/10.5121/mlaij.2015.2201>
- Tu, S., & Yang, B. (2021). Research on sentiment classification of micro-blog short text based on topic clustering. *Journal of Physics: Conference Series*, 1827(1). <https://doi.org/10.1088/1742-6596/1827/1/012160>

- Turney, P. D. (2006). *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*. 8.
- Ulfa, M. A., Irmawati, B., & Husodo, A. Y. (2018). Twitter Sentiment Analysis using Naïve Bayes Classifier with Mutual Information Feature Selection. *Journal of Computer Science and Informatics Engineering (J-Cosine)*, 2(2), 106–111.
- Uysal, A. K., & Murphey, Y. L. (2017). Sentiment Classification: Feature Selection Based Approaches Versus Deep Learning. *IEEE CIT 2017 - 17th IEEE International Conference on Computer and Information Technology*, 23–30. <https://doi.org/10.1109/CIT.2017.53>
- Varela, P. L., Martins, A. F., Aguiar, P. M., & Figueiredo, M. A. (2013). An empirical study of feature selection for sentiment analysis. *9th Conference on Telecommunications, Conftele, Castelo Branco, December*.
- Venkatesh, B., & Anuradha, J. (2019). A review of Feature Selection and its methods. *Cybernetics and Information Technologies*, 19(1), 3–26. <https://doi.org/10.2478/CAIT-2019-0001>
- Verma, B., & Thakur, R. S. (2018). Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches: A Survey. In *Lecture Notes in Networks and Systems* (Vol. 34). Springer Singapore. https://doi.org/10.1007/978-981-10-8198-9_46
- Vijayarani, S., Ilamathi, J., & Nithya, M. (2015). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16. <http://www.ijcscn.com/Documents/Volumes/vol5issue1/ijcscn2015050102.pdf>
- Vijayarani, S., & Janani, R. (2016). Text Mining: open Source Tokenization Tools – An Analysis. *Advanced Computational Intelligence: An International Journal (ACII)*, 3(1), 37–47. <https://doi.org/10.5121/acii.2016.3104>
- Vu, L., & Le, T. (2017). A lexicon-based method for Sentiment Analysis using social network data. *Int'l Conf. Information and Knowledge Engineering, September*, 10–16.
- Wagh, R., & Punde, P. (2018). Survey on Sentiment Analysis using Twitter Dataset. *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018, Iceca*, 208–211. <https://doi.org/10.1109/ICECA.2018.8474783>
- Waheeb, S. A., Khan, N. A., Chen, B., & Shang, X. (2020). Machine learning based sentiment text classification for evaluating treatment quality of discharge summary. *Information (Switzerland)*, 11(5).

<https://doi.org/10.3390/INFO11050281>

- Wan, C. H., Lee, L. H., Rajkumar, R., & Isa, D. (2012). 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>
- Wan, X. (2008). Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. *EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL, October*, 553–561. <https://doi.org/10.3115/1613715.1613783>
- Wan, Y. (2019). Short-Text Sentiment Classification Based on Graph-LSTM. *Proceedings - 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2019*, 35–38. <https://doi.org/10.1109/AIAM48774.2019.00014>
- Wang, H., & Hong, M. (2017). Probability and Variance Score : an Efficient Supervised Feature Selection Method for Text Classification. *Journal of Residuals Science & Technology (JRST)*, 14(3), 218–232. <https://doi.org/10.14355/jrst.2017.1403.023>
- Wang, J. H., Liu, T. W., Luo, X., & Wang, L. (2018). An LSTM approach to short text sentiment classification with word embeddings. *Proceedings of the 30th Conference on Computational Linguistics and Speech Processing, ROCLING 2018*, 214–223.
- Wang, J., Shan, G., Duan, X., & Wen, B. (2011). *Improved SVM-RFE Feature Selection Method for Multi-SVM Classifier*. 1592–1595.
- Wang, W., Wang, H., & Song, Y. (2017). Ranking product aspects through sentiment analysis of online reviews. *Journal of Experimental and Theoretical Artificial Intelligence*, 29(2), 227–246. <https://doi.org/10.1080/0952813X.2015.1132270>
- Wang, Z., & Lin, Z. (2019). Optimal Feature Selection for Learning-Based Algorithms for Sentiment Classification. *Cognitive Computation*, 12(1), 238–248. <https://doi.org/10.1007/s12559-019-09669-5>
- Westerski, A. (2007). Sentiment Analysis: Introduction and the State of the Art overview. *Universidad Politecnica de Madrid, España*, 211–218. http://www.adamwesterski.com/wp-content/files/docsCursos/sentimentA_doc_TLAW.pdf
- Wibowo Haryanto, A., Kholid Mawardi, E., & Muljono. (2018). Influence of Word Normalization and Chi-Squared Feature Selection on Support Vector Machine

- (SVM) Text Classification. *Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, ISEmantic ,Semarang, Indonesia*, 229–233.
- Widodo Wijayanto, U., & Sarno, R. (2018). An Experimental Study of Supervised Sentiment Analysis Using Gaussian Naïve Bayes. *Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, ISEmantic 2018*, 476–481. <https://doi.org/10.1109/ISEMANTIC.2018.8549788>
- Widya Sihwi, S., Prasetya Jati, I., & Anggrainingsih, R. (2018). Twitter Sentiment Analysis of Movie Reviews Using Information Gain and Naïve Bayes Classifier. *Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, ISEmantic 2018*, 190–195. <https://doi.org/10.1109/ISEMANTIC.2018.8549757>
- Wu, L., Wang, Y., Zhang, S., & Zhang, Y. (2017). Fusing Gini Index and Term Frequency for Text Feature Selection. *Proceedings - 2017 IEEE 3rd International Conference on Multimedia Big Data, BigMM 2017*, 280–283. <https://doi.org/10.1109/BigMM.2017.65>
- Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, 44(1), 48–59. <https://doi.org/10.1177/0165551516677946>
- Xue, B., Zhang, M., Member, S., & Browne, W. N. (2015). A Survey on Evolutionary Computation Approaches to Feature Selection. *IEEE Transactions on Evolutionary Computation*, September. <https://doi.org/10.1109/TEVC.2015.2504420>
- Yadla, H. K., & Rao, P. V. R. D. P. (2020). Machine learning based text classifier centered on TF-IDF vectoriser. *International Journal of Scientific and Technology Research*, 9(3), 583–586.
- Yang, P., & Chen, Y. (2018). A survey on sentiment analysis by using machine learning methods. *Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2017, 2018-Janua*, 117–121. <https://doi.org/10.1109/ITNEC.2017.8284920>
- Yazdani, S. F., Murad, M. A. A., Sharef, N. M., Singh, Y. P., & Latiff, A. R. A. (2017). Sentiment classification of financial news using statistical features. *International Journal of Pattern Recognition and Artificial Intelligence*, 31(3). <https://doi.org/10.1142/S0218001417500069>
- Yenter, A., & Verma, A. (2017). Deep CNN-LSTM with combined kernels from

- multiple branches for IMDb review sentiment analysis. *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2017, 2018-Janua*, 540–546.
<https://doi.org/10.1109/UEMCON.2017.8249013>
- Yiming, Y., & Jan O., P. (1997). A Comparative Study on Feature Selection in Text Categorization. *Proceeding ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning*, 53(9), 412–420.
- Yong, Y., Jian, X. H., Hua, D. X., & Xiao, L. (2012). Comparative study on feature selection in Uighur text categorization. *Advances in Information Sciences and Service Sciences*, 4(3), 19–26. <https://doi.org/10.4156/AISS.vol4.issue3.3>
- Yordanova, S., & Kabakchieva, D. (2017). Sentiment Classification of Hotel Reviews in Social Media with Decision Tree Learning. *International Journal of Computer Applications*, 158(5), 1–7. <https://doi.org/10.5120/ijca2017912806>
- Yousefpour, A., Ibrahim, R., & Hamed, H. N. A. (2017). Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis. *Expert Systems with Applications*, 75, 80–93. <https://doi.org/10.1016/j.eswa.2017.01.009>
- Yousefpour, A., Ibrahim, R., Nuzly, H., & Hamed, A. (2014). A Novel Feature Reduction Method in Sentiment Analysis. *International Journal of ...*, 4(1), 34–40. <http://se.cs.utm.my/ijic/index.php/ijic/article/view/81>
- Yu, B., Zhou, J., Zhang, Y., & Cao, Y. (2017). Identifying restaurant features via sentiment analysis on yelp reviews. *ArXiv*, 1–6.
- Yu, Y. (2008). SVM-RFE Algorithm for Gene Feature Selection. *Computer Engineering*.
- Yuchun, T., Yan-Qing, Z., & Zhen, H. (2005). FCM-SVM-RFE Gene Feature Selection Algorithm for Leukemia Classification from Microarray Gene Expression Data. *Fuzzy Systems, 2005. FUZZ '05. The 14th IEEE International Conference On*, 97–101. <https://doi.org/10.1109/FUZZY.2005.1452375>
- Yusof, Y., & Mustaffa, Z. (2011). Dengue Outbreak Prediction : A Least Squares Support Vector Machines Approach. *International Journal of Computer Theory and Engineering*, 3(4), 489–493. <https://doi.org/10.7763/IJCTE.2011.V3.355>
- Zhai, Y., Song, W., Liu, X., Liu, L., & Zhao, X. (2019). A Chi-Square Statistics Based Feature Selection Method in Text Classification. *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, 2018-Novem*, 160–163. <https://doi.org/10.1109/ICSESS.2018.8663882>

- Zhang, J., Hu, X., Li, P., He, W., Zhang, Y., & Li, H. (2014). A hybrid feature selection approach by correlation-based filters and SVM-RFE. *Proceedings - International Conference on Pattern Recognition*, 3684–3689. <https://doi.org/10.1109/ICPR.2014.633>
- Zhang, W., Yoshida, T., & Tang, X. (2011). A comparative study of TF*IDF, LSI and multi-words for text classification. *Expert Systems with Applications*, 38(3), 2758–2765. <https://doi.org/10.1016/j.eswa.2010.08.066>
- Zhen, Z., Wang, H., Han, L., & Shi, Z. (2011). Categorical document frequency based feature selection for text categorization. *Proceedings - 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, ICM 2011*, 2, 65–68. <https://doi.org/10.1109/ICM.2011.365>
- Zhou, F., Jiao, J. R., Yang, X. J., & Lei, B. (2017). Augmenting feature model through customer preference mining by hybrid sentiment analysis. *Expert Systems with Applications*, 89, 306–317. <https://doi.org/10.1016/j.eswa.2017.07.021>
- Zhou, H., Han, S., & Liu, Y. (2018). A Novel Feature Selection Approach Based on Document Frequency of Segmented Term Frequency. *IEEE Access*, 6, 53811–53821. <https://doi.org/10.1109/ACCESS.2018.2871109>
- Zhou, P., & El-Gohary, N. (2014). Semantic-based text classification of environmental regulatory documents for supporting automated environmental compliance checking in construction. *Construction Research Congress 2014: Construction in a Global Network - Proceedings of the 2014 Construction Research Congress*, 897–906. <https://doi.org/10.1061/9780784413517.0092>
- Zvarevashe, K. (2018). A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews. *2018 Conference on Information Communications Technology and Society (ICTAS)*, 0–3.