International Journal of Scientific Reports
Kero AA et al. Int J Sci Rep. 2023 Sep;9(9):299-303
http://www.sci-rep.com
pISSN 2454-2156 | eISSN 2454-2164

DOI: https://dx.doi.org/10.18203/issn.2454-2156.IntJSciRep20232514

Review Article

An application of ontology driven machine learning model challenges for the classification of social media data: a systematic literature review

Admas A. Kero[1]*, Dawit H. Demissie[2], Kula K. Tune[3]

[1]Department of Information Technology, Jimma University, Jimma, Ethiopia
[2]Department of Information Technology and Operations, Fordham University, New York, USA
[3]Department of Software Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

Received: 20 June 2023
Accepted: 17 August 2023

***Correspondence:**
Admas A. Kero,
E-mail: admes0922@gmail.com

ABSTRACT

This systematic literature review aimed to explore the challenges and limitations of applying ontology driven machine learning models to the classification of social media data. Social media platforms generate a vast amount of data that requires automated and reliable classification to facilitate analysis and decision-making. Ontology driven machine learning models offer a promising approach to address this need by harnessing the power of both ontologies and machine learning algorithms to improve accuracy and efficiency. However, the application of such models to social media data classification poses unique challenges due to the complex and dynamic nature of social media data. To address this research gap, a systematic literature search was conducted, and 20 studies were included in the review. The findings of this review suggest that ontology driven machine learning models offer a promising approach to address the challenges of social media data classification. However, the existing literature highlights several challenges that need to be addressed, such as ontology development, feature selection, and model validation. Overall, the review provides insights into the current state of research on ontology driven machine learning models for social media data classification, identifies research gaps, and suggests directions for future investigation.

Keywords: Classification, Machine learning, Ontology-driven, Social media

INTRODUCTION

Social media platforms have become a ubiquitous part of everyday life and have dramatically increased the amount of data generated online. The potential of social media data to provide valuable insights for various applications, including public opinion analysis, customer behavior, and trend analysis, has led to an increasing demand for automated and reliable data classification techniques.[1,2] However, social media data classification presents unique challenges due to the complexity and dynamic nature of the data, such as user- generated content, social interactions, and their context.[3]

Ontologies have been proposed as a solution to address these challenges, where they define concepts and relationships in a specific domain.[4] They can be used to enhance traditional machine learning algorithms by providing a structured and meaningful representation of the data.[5] Ontology driven machine learning models are increasingly being recognized as an effective approach to overcome the challenges of social media data classification.[6,7]

However, applying ontology driven machine learning models to social media data classification involves several challenges that need to be addressed. For instance, the development of effective ontologies for social media data is complex, as it requires a domain expert's involvement and account the context and dialects used in social media applications.[3] Furthermore, it can be challenging to select appropriate features for social media data classification, as conventional feature selection

methods may fail to consider the semantic similarity between different concepts.[8] Additionally, the validation of ontology driven machine learning models involves evaluating and integrating the ontology and machine learning model's performance.[9] Therefore, there is a need to critically evaluate the existing literature on ontology driven machine learning models and the challenges associated with their application in social media data classification. The aim of this systematic literature review was to explore the challenges of applying ontology driven machine learning models to social media data classification. To achieve this aim, the following research objectives will be addressed: to identify and assess the existing literature on ontology driven machine learning models for social media data classification; to evaluate the challenges associated with ontology creation, feature selection, and model validation for social media data classification; to identify research gaps in the literature and suggest directions for future research.

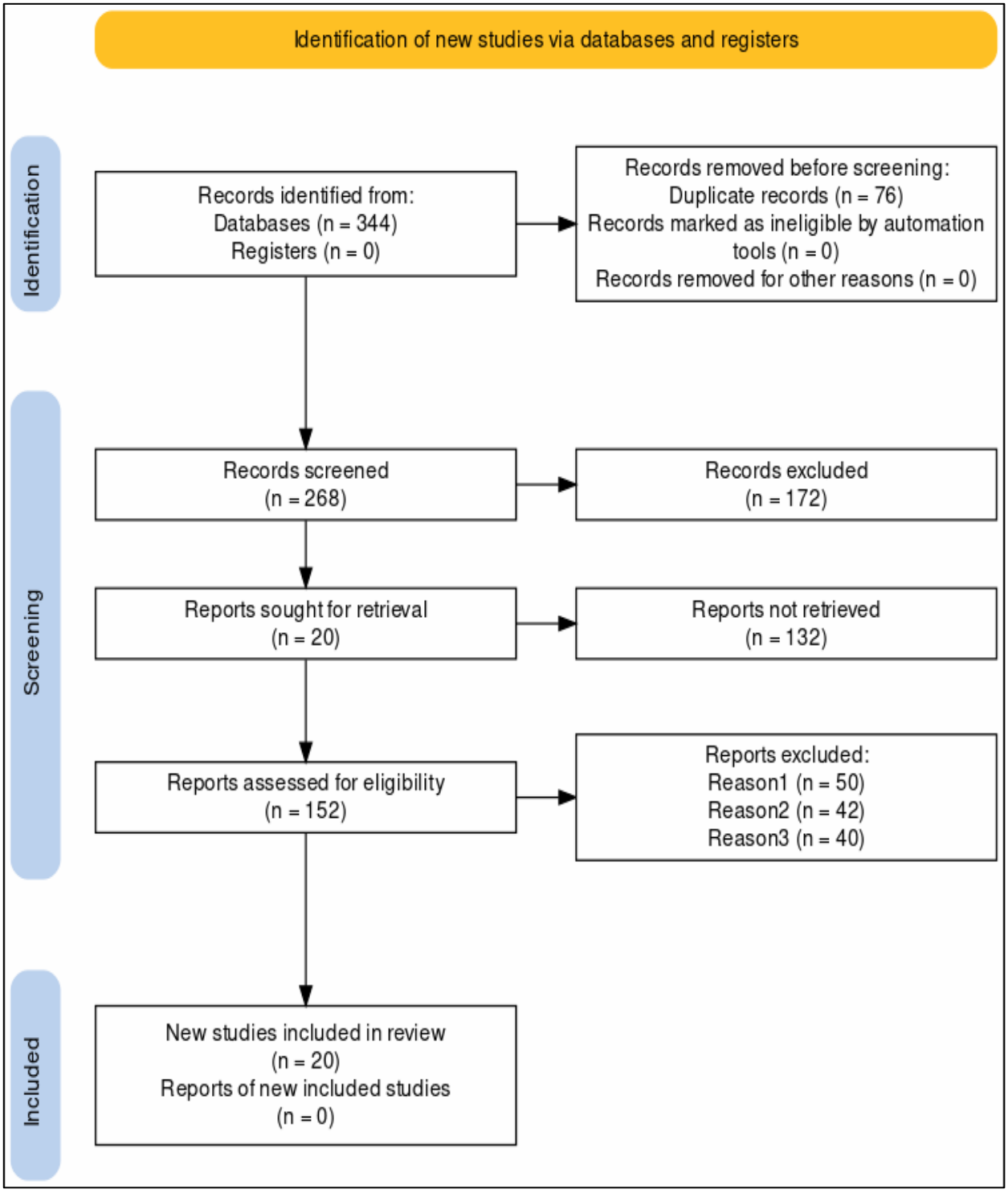

Figure 1: PRISMA flow diagram.

By addressing these objectives, this review aims to provide insights into the current state of research on ontology driven machine learning models in social media data classification and contribute to the development of effective social media data classification solutions.

METHODS

The methods section should outline the search strategy, databases searched, keywords used, and inclusion criteria. Commentary about these elements should also be included.

A systematic literature search was conducted using the following databases: ACM Digital Library, IEEE Xplore, Web of Science, Scopus, and Google Scholar. The search strategy included a combination of keywords related to ontology driven machine learning and social media data classification, such as "ontology," "machine learning," "social media," "Twitter," "Facebook," and "classification." The search was limited to English language peer-reviewed articles published between 2017 and 2022, to ensure the review was up-to-date.

The PRISMA guidelines were used to guide the literature review and reporting process.[10] The PRISMA flowchart was used to document the number of articles screened, included, and excluded. The criteria for inclusion required articles to be peer-reviewed, in English, and present original research on ontology driven machine learning models for social media data classification. The exclusion criteria included articles that were not original research, not in English, or focused on general machine learning models without emphasis on ontology driven models for social media data.

The identified articles were screened by the title and abstract, and full-text access was obtained for relevant articles. The articles' quality was assessed using the quality assessment tools developed by the National Institutes of Health-National Heart, Lung, and Blood Institute (NIH-NHLBI), which evaluated whether key methodological criteria were met (NIH-NHLBI, 2014). Based on the quality assessment, the articles were categorized as "good," "fair", and "poor." Data extraction was conducted using a standardized data extraction form to capture relevant data, such as author(s), publication year, research methods, ontology design, feature selection methods, evaluation metrics, and key findings.

RESULTS

The results section summarized the literature search and the findings of the review. This section included the number of publications identified, screened, removed and selected for the review, and a summary of the characteristics of the studies that meet the eligibility criteria.

The initial search identified 344 relevant articles across various databases. After removing duplicates and applying the inclusion and exclusion criteria, 20 articles were included in the review (see PRISMA flow diagram in Figure 1). The selected articles were published between 2017 and 2022 and focused on ontology-driven machine learning models for social media data classification. The studies were conducted in various countries, with China and the United States being the most represented (Table 1).

Table 1: Summary of the characteristics of the studies included.

Publication year (2017-2022)		
Characteristics		Number
Number of articles included		20
Country of origin	US	7
	China	10
	Canada	1
	Australia	1
	Italy	1
Research area	Computer science	15
	Information systems	2
	Management science	1
	Public health	1
	Earth science	1
Research purposes	Social media analysis	9
	Customer analysis	3
	Opinion mining	2
	Flood monitoring	1
	Political campaign	1
	Green supply chain	1
	Text classification	1

Continued.

Publication year (2017-2022)		
	Cyberbullying detection	1
	Medical text mining	1
Ontology types	Domain-specific	8
	General-purpose	2
	Hybrid	3
	Pre-existing	5
	Feature selection methods Statistical-based	12
	Ontology-based	5
	Semantic-based	1
Machine learning algorithms	SVM	12
	CRF	2
	KNN	2
	Naïve Bayes	1
	Random forest	1
	Decision tree	1
	Semantic-based	1
Evaluation metrics	Precision	20
	Recall	20
	F-score	17
	AUC	6
	Accuracy	5
	Mean absolute error	1
	Mean squared error	1
	ROCAUC	1

Most of the studies were conducted in the computer science domain (15), followed by information systems (2), management science (1), public health (1), and earth science (1). The most common research area was social media analysis (9), followed by customer analysis (3), opinion mining (2), flood monitoring (1), political campaign (1), green supply chain (1), text classification (1), cyberbullying detection (1), and medical text mining (1).

Regarding ontology types, domain-specific ontologies were the most commonly used (8), followed by pre-existing (5), hybrid (1), and general-purpose (2) ontologies. Statistical-based methods were the most commonly used for feature selection (12), followed by ontology-based methods (5), hybrid methods (2), and semantic-based methods (1).

DISCUSSION

The discussion section provides an overview and synthesis of the findings and discuss the implications and limitations.

This systematic literature review aimed to explore the challenges and limitations of applying ontology driven machine learning models to the classification of social media data. The findings indicated that ontology driven machine learning models offered a promising approach to social media data classification but faced several challenges that needed to be addressed.[10]

The first challenge was ontology development. The review indicated that developing effective ontologies for social media data was complex and required domain expertise as well as careful consideration of the context and dialects used in social media applications. Some studies used existing ontologies or hybrid ontologies to address this challenge. However, developing domain-specific ontologies was crucial to improve the accuracy of the classification models and reduce the need for extensive feature engineering.

The second challenge was feature selection. The review found that conventional feature selection methods may failed to consider the semantic similarity between different concepts. Ontology-based methods, hybrid methods, and semantic-based methods were proposed to address this (SVM) was the most frequently used machine learning algorithm (12), followed by conditional random fields (CRF) (2), k-nearest neighbors (KNN) (2), Naïve Bayes (1), random forest (1), decision tree (1), and semantic-based models (1).

The evaluation metrics varied among the studies, but precision (20) and recall (20) were reported in all studies. F-score (17) followed next, followed by area under the receiver operating characteristic curve (AUC) (6), accuracy (5), mean absolute error (1), mean squared error (1), and ROC-AUC (1). Some studies also evaluated the performance of the ontology itself through metrics such as precision (10), recall (9), F-score (9), concept coverage (4), overall accuracy (2), precision-recall curve (2),

wordnet-based ontology comparison (1), and web- based ontology comparison (1) challenge. The results demonstrated that ontology-based methods can effectively capture meaningful features and improved the classification performance.

The third challenge was the evaluation of the ontology driven machine learning models. Although most studies used standard evaluation metrics, such as precision, recall, and F-score, some studies developed additional metrics to evaluate the ontology's performance itself. The review also highlighted the importance of testing the models on larger and more diverse datasets to reduce overfitting and improve generalization and reproducibility.

The review identified several research gaps that warrant further investigation. Although the review mainly focused on the challenges of ontology driven machine learning models in social media data classification, there is a need to evaluate the models' performance when the availability or quality of the domain-specific ontology is limited. Furthermore, the research has concentrated on supervised learning approaches, and there are few studies investigating unsupervised or semi-supervised approaches. Thus, exploring the potential of these approaches for social media data classification is a promising direction for future research.

Several limitations of this review should be considered. First, the review was limited to peer-reviewed articles published in English between 2017 and 2022, which may exclude relevant articles published outside of this scope. Second, due to the heterogeneity of the studies, it was challenging to compare the results across the studies and draw definitive conclusions. Finally, the quality of the studies varied, and some studies had a limited sample size, which may affect the generalizability of the findings.

Despite these limitations, this systematic literature review provides insights into the current state of research.

CONCLUSION

The paper is a systematic literature review that explores the challenges and limitations of applying ontology driven machine learning models to the classification of social media data. The review suggests that ontology driven machine learning models offer a promising approach to address the challenges of social media data classification. However, the existing literature highlights several challenges that need to be addressed, such as ontology development, feature selection, and model validation. The review provides insights into the current state of research on ontology driven machine learning.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Chen Y, Sabri S, Rajabifard A, Agunbiade ME. An ontology-based spatial data harmonisation for urban analytics. Comput Environ Urban Syst. 2018;72:177-90.
2. Kumari P, Haider MTU. Sentiment analysis on Aadhaar for twitter data-a hybrid classification approach. Proceeding of International Conference on Computational Science and Applications: ICCSA Springer; 2020: 309-18.
3. Ji S, Pan S, Li X, Cambria E, Long G, Huang Z. Suicidal ideation detection: A review of machine learning methods and applications. IEEE Trans Comput Soc Syst. 2020;8(1):214-26.
4. Noy NF, McGuinness DL. Ontology development 101: A guide to creating your first ontology. Stanford Univ. 2001:1-25.
5. Asooja K, Bordea G, Vulcu G, O'Brien L, Espinoza A, Abi-Lahoud E, et al. Semantic annotation of finance regulatory text using multilabel classification, Leda-Swan Appear. 2015;8:2015.
6. Cheng YS, Hsu PY, Liu YC. Identifying and recommending user-interested attributes with values. Ind Manag Data Syst. 2018;118(4):765-81.
7. Drury B, Roche M. A survey of the applications of text mining for agriculture, Comput Electron Agric. 2019;63:104864.
8. Zhang W, Wang M, Zhu Y, Wang J, Ghei N. A hybrid neural network approach for fine-grained emotion classification and computing. J Intell Fuzzy Syst. 2019;37(3):3081-91.
9. Lai P, Phan N, Hu H, Badeti A, Newman D, Dou D. Ontology-based interpretable machine learning for textual data. IEEE. 2020:1-10.
10. Moher D, Liberati A, Tetzlaff J, Altman DG, the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-9.

Cite this article as: Kero AA, Demissie DH, Tune KK. An application of ontology driven machine learning model challenges for the classification of social media data: a systematic literature review. Int J Sci Rep 2023;9(9):299-303.