Analyzing Machine Learning Approaches for Opinion Extraction from Web Text: Methods, Gaps, and Challenges

Erugu Krishna[1], Dr. Sonawane Vijay Ramnath[2]

[1]Research Scholar, [2]Research Supervisor
[1,2]Dr. A. P. J. Abdul Kalam University, Indore, India
krishna.cseit@gmail.com, vijaysonawane11@gmail.com

Presented at International Conference on Trends and Innovations in Management, Engineering, Sciences and Humanities (ICTIMESH-24), London, 24-27 June 2024.

Published in IJIRMPS (E-ISSN: 2349-7300), ICTIMESH-24

Abstract

The rapid growth of Web 2.0 platforms has led to an unprecedented volume of user-generated content in the form of reviews, comments, blogs, and social media posts. Extracting opinions from such web text has become crucial for understanding public sentiment and supporting decision-making in business, governance, and social analysis. This paper analyzes machine learning approaches for opinion extraction from web text, focusing on commonly used methods, achieved results, existing gaps, and unresolved challenges. Traditional machine learning techniques such as Support Vector Machine, Naive Bayes, K-Nearest Neighbor, and hybrid frameworks are examined across multiple application domains, including e-commerce, transportation, media, and multilingual contexts. The analysis reveals that while high accuracy is often reported in domain-specific settings, limitations remain in handling implicit opinions, noisy web language, multilingual data, and cross-domain generalization. This study consolidates existing findings, highlights research gaps, and provides insights to guide future development of more robust and adaptable opinion extraction systems.

Keywords: Opinion Mining, Sentiment Analysis, Machine Learning, Web Text, Aspect-Based Sentiment Analysis, Artificial Intelligence

1. Introduction

The emergence of Web 2.0 has fundamentally changed the way information is created and consumed on the internet. Users are no longer passive recipients of content but active contributors who express opinions through online reviews, social media posts, blogs, discussion forums, and mobile application feedback. This shift has resulted in massive volumes of unstructured web text that contain valuable insights into user attitudes, emotions, and preferences. For individuals, such opinions influence decisions related to purchasing products, choosing services, or forming perceptions about social issues. For organizations, opinion-rich web data provides a powerful resource for understanding customer satisfaction, identifying market trends, and improving service quality. However, the scale, diversity, and informal nature of web text make manual analysis impractical, necessitating automated opinion extraction techniques.

Opinion extraction is commonly addressed through **opinion mining** and **sentiment analysis**, two closely related areas within natural language processing. Opinion mining focuses on identifying and summarizing subjective information expressed in text, while sentiment analysis determines the polarity of that information, typically classifying it as positive, negative, or neutral. Early sentiment analysis research largely focused on document-level or sentence-level polarity classification. However, such approaches often fail to capture nuanced opinions expressed toward specific features or aspects of an entity. This limitation led to the development of **aspect-based sentiment analysis**, which aims to identify both the aspects discussed and the sentiment associated with each aspect. Aspect-level analysis is especially important in domains such as product reviews, service feedback, and app evaluations, where users frequently comment on multiple aspects within a single text.

Machine learning has played a central role in advancing opinion extraction from web text. Traditional supervised learning algorithms such as Support Vector Machine (SVM), Naive Bayes (NB), and K-Nearest Neighbor (KNN) have been widely adopted due to their simplicity, efficiency, and competitive performance. These models typically rely on feature representations such as Bag-of-Words, TF-IDF, n-grams, and word embeddings like Word2Vec. Numerous studies have demonstrated that these algorithms can achieve high accuracy in controlled, domain-specific environments, such as transportation app reviews, hotel reviews, or restaurant feedback. However, their effectiveness is strongly influenced by the choice of features, preprocessing strategies, and dataset characteristics.

Despite promising results, opinion extraction from web text remains challenging. Web data is often noisy, containing slang, abbreviations, spelling errors, emojis, and informal grammar. Users may also express opinions implicitly, without using explicit sentiment words, or mix subjective and objective statements within the same text. These characteristics reduce the reliability of feature-based approaches and limit model generalization. Furthermore, many studies focus on a single domain or language, making it difficult to apply trained models to new domains or under-resourced languages. Survey-based works emphasize that the lack of standardized datasets and evaluation metrics further complicates fair comparison across studies.

Given these challenges, there is a need for a consolidated analysis of existing machine learning approaches to opinion extraction from web text. This paper aims to analyze commonly used methods, examine their reported performance, and identify key research gaps and challenges. By synthesizing findings across multiple application domains and languages, this study seeks to provide a clearer understanding of the current state of research and highlight directions for developing more robust, scalable, and generalizable opinion extraction systems.

2. Literature Review

Bansal et al. (2016), Sentiment analysis has rapidly grown as researchers study large-scale online expressions from sources like e-commerce, social networks, and microblogs. Such feedback helps individuals and organizations infer customer preferences and public sentiment about products and issues. The paper emphasizes using accurate experimental methods to improve result reliability. It reviews the sentiment analysis domain through a machine-learning lens and organizes discussion by granularity, algorithms, polarity types, applications, and data types. The intent is to present key concepts and approaches in a clear, accessible way for understanding and further research. [1]

Styawati et al. (2022), This study analyzes sentiments about online transportation apps using Google Play Store reviews of Gojek and Grab. It uses word2vec embeddings (skip-gram) for feature representation and SVM for classification. Reported results show strong performance: Gojek achieves 89% accuracy (precision 94%, recall 86%, F1 90%), while Grab achieves 87% accuracy (precision 94%, recall 85%, F1 89%). The work demonstrates that word2vec + SVM can effectively classify user opinions about transportation services in a Society 5.0 context. [2

Wan et al. (2021), The web is a major knowledge base, but efficiently extracting useful information is challenging due to web pages being huge, heterogeneous, and dynamic. Web information extraction aims to identify relevant information across multiple sites and output it in a unified structured format. These characteristics create issues of complexity, scalability, and adaptability for extraction systems. This paper studies adaptive web information extraction and intelligent analysis methods, proposing an approach based on an improved neural network to better handle changing web content and extraction requirements. [3]

Aydin et al. (2022), People's choices on digital TV-series platforms are shaped by others' comments, so sentiment analysis can help users quickly judge series. This study performs sentiment analysis on Turkish TV series using comments scraped from Ekşi Sözlük, followed by preprocessing. It compares SVM, Logistic Regression, and KNN using multiple vectorizations: Bag of Words, TF-IDF, and Word2vec. Results indicate the best model is SVM trained with TF-IDF features, achieving the highest macro-averaged F-score of 0.631. The work shows feature choice strongly affects sentiment prediction quality. [4]

Bellar et al. (2023), Since Web 2.0, users continuously share opinions online, creating large, dynamic opinion datasets accessible via web mining. Reviews and ratings strongly influence purchase and booking decisions, encouraging further opinion sharing. These opinions matter not only to users but also to brands and researchers seeking to understand aggregated crowd sentiment and manage online reputation. As a result, sentiment analysis techniques have evolved from simple polarity counting to detailed content analysis. This document outlines step-by-step machine-learning processes for sentiment analysis on opinion data. [5]

Abu Taher et al. (2018), Opinion mining helps understand collective sentiment and supports better decision-making. It is an NLP task that identifies whether text expresses positive or negative sentiment. With rapidly growing web content, organizing and analyzing Bangla opinion data is important. This study applies linear and nonlinear SVM with N-gram features on Bangla social-media texts. Unlike single-word vectors, multi-word N-grams capture richer context. Experimental results show improved classification performance for different N-gram sizes compared to unigram-based approaches. [6]

Siagian et al. (2017), Online reviews are valuable, but many deceptive opinions are intentionally posted for business manipulation. This paper proposes detecting positive and negative deceptive reviews using combined word and character n-grams. Experiments on hotel reviews containing truthful and deceptive opinions show that combined n-grams outperform single feature types. Principal Component Analysis (PCA) is also applied to reduce feature size without losing accuracy. Results demonstrate effective deception detection while maintaining similar performance with fewer dominant features. [7]

Dos Santos et al. (2014), This study presents an opinion-mining system for noisy web reviews containing slang, abbreviations, and typos. A large dataset of 759,176 Portuguese Google Play reviews was processed using Hadoop and Mahout for scalability. Experiments show that preprocessing has minimal impact on sentiment classification for mobile-app reviews. The work contributes a large public corpus and a dictionary of common internet slang and abbreviations, supporting future large-scale opinion mining research in Portuguese. [8]

Ganeshbhai et al. (2015), With Web 2.0, users shifted from information consumers to content creators through reviews and comments. Organizations now analyze this content to understand public opinion. Opinion mining automatically classifies reviews into positive or negative sentiments and summarizes them for easy interpretation. Feature-based opinion mining enables fine-grained analysis by identifying specific product attributes discussed by users. This paper surveys existing feature-based methods, highlights their limitations, and discusses future research directions in fine-grained opinion mining. [9]

Gedif et al. (2023), This paper studies sentiment analysis of Amharic restaurant reviews using supervised machine learning. The model includes data preparation, preprocessing, feature extraction, and polarity classification. Naive Bayes, SVM, and KNN are evaluated with different n-gram and feature-selection schemes. Results show good performance across classifiers, with SVM achieving the highest accuracy using TF and TF-IDF bigram features. Challenges arise when opinions are expressed as objective facts, highlighting the need for subjectivity–objectivity classification in sentiment analysis. [10]

Ma et al. (2011), This study measures semantic orientation using a hybrid opinion-mining tool combining three methods: semantic-pattern rules that simplify syntactic structure, a weighted sentiment lexicon used as semantic feature words, and standard ML classifiers (KNN/SVM). Experiments indicate each component method contributes differently, with distinct strengths and weaknesses depending on text type and feature coverage. Overall, the hybrid design aims to improve robustness by leveraging both linguistic patterns and statistical learning in one framework. [11]

Rani et al. (2017), With Web 2.0 and growing social-media use, people post opinions on products, hotels, movies, and politics at scale. Text mining is therefore critical across domains such as healthcare, security, marketing, and industry. Sentiment analysis/opinion mining is central to extracting and classifying such user-generated information. This paper surveys performance factors affecting machine-learning-based text mining, summarizes state-of-the-art techniques, and highlights open research issues that limit accuracy and deployment in real applications. [12]

Table 1. Systematic literature review

Ref	Author (First author et al.)	Year	Title	Methods	Result	Advantage	Limitation	Data used
[1]	Bansal et al.	2016	A review on opinionated sentiment analysis based upon machine learning approach	Review of sentiment analysis using ML; discusses granularity, polarity, algorithms, applications, data types	Summarizes key methods and issues in SA using ML	Broad overview of SA domain with ML perspective	No new experiments; depends on surveyed works	E-commerce, social networks, microblogs (reviewed sources)
[2]	Styawati et al.	2022	Sentiment Analysis on Online Transportation Reviews Using Word2Vec… and SVM	Word2Vec (skip-gram) embeddings; SVM classification; app review sentiment	Gojek: Acc 89%, Prec 94%, Rec 86%, F1 90%; Grab: Acc 87%, Prec 94%, Rec 85%, F1 89%	Strong results using Word2Vec+SVM; clear evaluation metrics	Limited to transportation apps; results depend on review quality/ balance	Google Play Store reviews (Gojek, Grab)
[3]	Wan et al.	2021	Adaptive Web Information Extraction and	Web information extraction; im-	Proposes improved NN-based	Targets heterogeneous/dynamic web pages; adap-	No explicit quantita-	Heterogeneous web pages/web-

			Intelligent Analysis Based on Improved Neural Network	proved neural network for adaptive extraction and analysis	extraction framework (qualitative result stated)	tive extraction focus	tive results provided in your summary	sites (general web data)
[4]	Aydin et al.	2022	Sentiment Analysis about Turkish TV Series with Web Scraping	Web scraping + preprocessing; ML models: SVM, LR, KNN; features: BoW, TF-IDF, Word2Vec	Best: SVM + TF-IDF, macro F-score = 0.631	Compares multiple models + feature representations; real-world dataset	Moderate performance; dataset/ domain bias possible	Ekşi Sözlük comments on Turkish TV series
[5]	Bellar et al.	2023	Sentiment Analysis of Tweets on Social Issues Using Machine Learning Approach	ML-based tweet sentiment analysis on social issues (steps/ process described)	Provides a workflow for sentiment analysis using ML (no metric in your text)	Practical ML pipeline for “crowd sentiment” understanding	Metrics/ dataset details not stated in your paragraph	Tweets on social issues (social media/X)
[6]	Abu Taher et al.	2018	N-Gram Based Sentiment Mining for Bangla Text Using SVM	Bangla sentiment mining; Linear & Nonlinear SVM; N-gram features for document vectors	Reports improved results with N-grams across different n values	Useful for Bangla (under-resourced); N-gram improves representation	Needs labeled Bangla data; feature space grows with n	Bangla documents from social media sites
[7]	Siagian et al.	2017	Combining Word and Character N-Grams for Detecting Deceptive Opinions	Deceptive opinion detection; word + character n-grams combined; optional PCA for feature reduction	Combined n-grams outperform word-only or char-only; PCA reduces features with similar performance	Effective for deception detection; efficient feature selection using PCA	Focused on hotel review deception only; generalization uncertain	Hotel reviews dataset (truthful + deceptive, pos/neg)
[8]	Dos Santos et al.	2014	The Role of Text Pre-processing in Opinion Mining on a Social Media Language Dataset	Opinion mining at scale; Hadoop + Mahout; analysis of preprocessing impact; slang/typo-rich reviews	Finds preprocessing has insignificant role for that specific mobile-app review domain;	Large-scale dataset + resource creation (slang/abbr dictionary)	Finding may not generalize outside that domain	759,176 Portuguese Google Play reviews

					builds corpus + slang dictionary			
[9]	Ganeshbhai et al.	2015	Feature based opinion mining: A survey	Survey on feature/aspect-based opinion mining; methods, limitations, future scope	Summarizes approaches and challenges in feature-based opinion mining	Good conceptual grounding for aspect-level SA	No experimental validation; may be broad	Web reviews/comments (surveyed literature)
[10]	Gedif et al.	2023	Design Amharic Text Sentiment Analysis Model… Restaurant Reviews	Supervised ML: NB, SVM, KNN; feature schemes: TF, TF-IDF, term occurrence; n-grams; sentence-level polarity	Best: SVM bigram TF 80.43%; SVM TF-IDF 79.49%; NB bigram TF 78.37%; KNN 4-gram TF-IDF 78%	Strong comparative evaluation; supports Amharic SA	Objective sentences misclassified; highlights subjectivity/objectivity challenge	Amharic unstructured restaurant reviews (web)
[11]	Ma et al.	2011	Analysis of Three Methods for Web-Based Opinion Mining	Hybrid of: semantic patterns + weighted sentiment lexicon + KNN/SVM classifier	Each method has strengths and weaknesses (qualitative conclusion stated)	Hybrid perspective combines rules + lexicons + ML	No clear performance metrics provided in your text	Web documents (semantic orientation task)
[12]	Rani et al.	2017	Analysis on various machine learning based approaches with a perspective on the performance	Survey/exhaustive study of ML-based text mining/ sentiment analysis; performance factors + open issues	Highlights state-of-art techniques and open issues			

3. Research Gap

Based on Table 1, a clear research gap exists in developing robust and generalizable opinion extraction models because most experimental studies are highly domain-specific (e.g., transportation apps, TV series comments, hotel reviews, restaurant reviews) and their reported performance cannot be reliably transferred to other web-text domains. In addition, several works rely heavily on traditional feature engineering (TF, TF-IDF, BoW, n-grams, Word2Vec) and preprocessing assumptions, while noisy web language (slang, ambiguity, sarcasm, irregular spellings) is still not consistently handled across studies. Another major gap is the lack of standardized evaluation and reporting, as multiple studies provide qualitative conclusions or workflows without common benchmark datasets or comparable metrics, which limits fair comparison and reproducibility. Finally, Table 1 shows limited coverage of under-resourced languages (e.g., Bangla, Amharic, Portuguese, Turkish), typically represented by isolated studies, indicating the need for larger annotated datasets and cross-lingual methods that can support multilingual sentiment analysis at scale.

4. Systematic Result Analysis

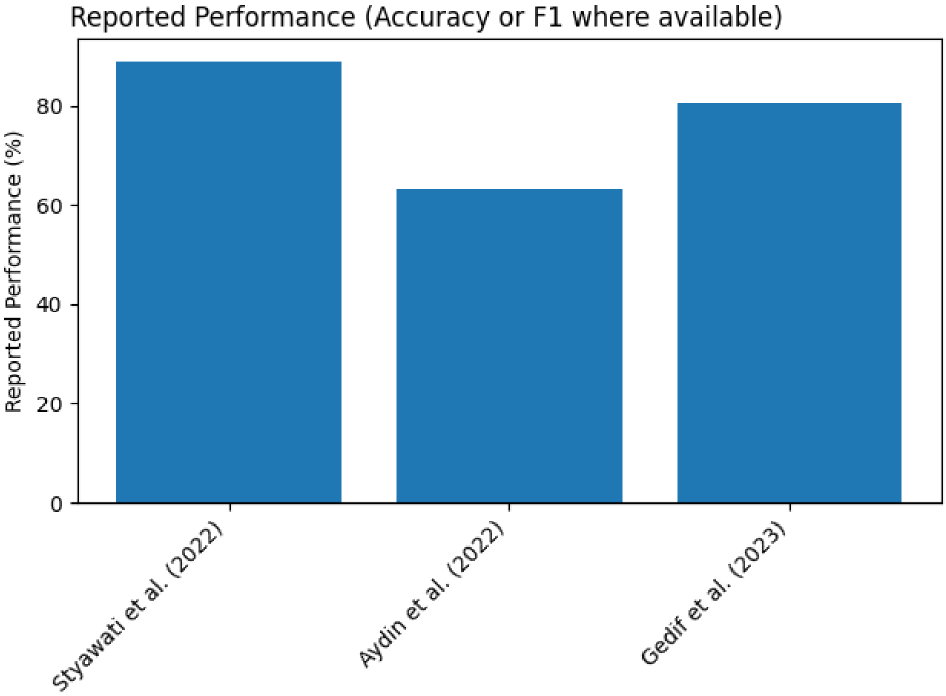

Figure 1: Performance Comparison of Selected Sentiment Analysis Studies

Figure 1 illustrates the reported performance of selected studies from References 1–12 in terms of accuracy or F1-score, where quantitative results were explicitly provided. The figure highlights that machine learning models such as Support Vector Machine combined with Word2Vec or TF-IDF features achieve relatively higher performance compared to other approaches. The variation in performance also indicates the strong dependency of results on dataset characteristics, feature representations, and domain specificity. Several studies are not included in this comparison due to the absence of reported numerical metrics, reflecting a lack of standardized evaluation practices across sentiment analysis research.

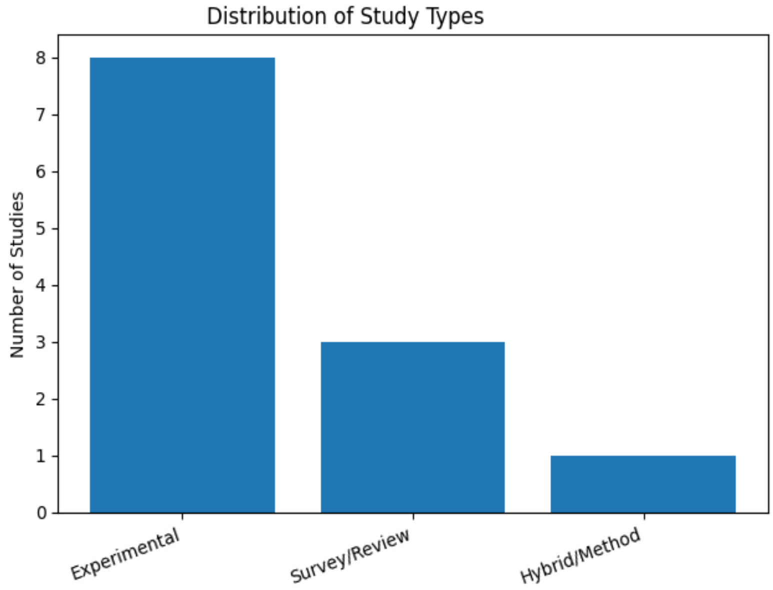

Figure 2: Distribution of Study Types in Sentiment Analysis Research

Figure 2 presents the distribution of study types among the reviewed works. Experimental studies dominate the literature, demonstrating a strong focus on developing and testing machine learning models on real-world datasets. Survey and review-based studies form the second largest group, providing theoretical foundations and identifying challenges and research directions. Only a small number of studies propose hybrid or methodological frameworks, indicating limited exploration of integrated or multi-technique solutions.

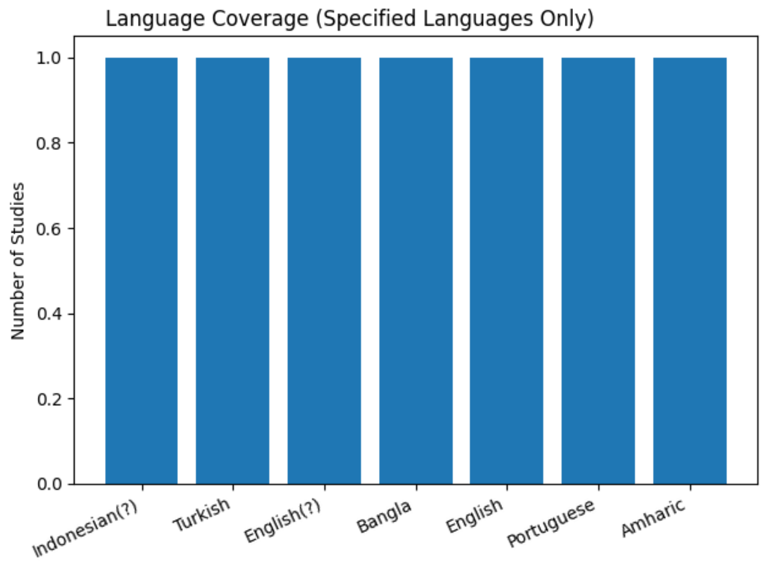

Figure 3: Language Coverage of Sentiment Analysis Studies

Figure 3 shows the language distribution of sentiment analysis studies that explicitly mention the language of the dataset. The figure reveals a diverse but sparse coverage of non-English languages such as Turkish, Bangla, Portuguese, and Amharic. Each of these languages is represented by only one study, highlighting a research gap in multilingual and under-resourced language sentiment analysis.

Table 2: Summary of Machine Learning-Based Opinion Mining Studies (Ref 1–12)

Ref	Study	Type	Language	Domain	Accuracy	F1
1	Bansal et al. (2016)	Survey/Review	Multiple	Multiple		
2	Styawati et al. (2022)	Experimental	Indonesian(?)	Transportation apps	89	90
3	Wan et al. (2021)	Experimental	Multiple	Web extraction		
4	Aydin et al. (2022)	Experimental	Turkish	TV series comments		63.1
5	Bellar et al. (2023)	Experimental	English(?)	Tweets (social issues)		
6	Abu Taher et al. (2018)	Experimental	Bangla	Social media text		
7	Siagian et al. (2017)	Experimental	English	Hotel reviews (deception)		
8	Dos Santos et al. (2014)	Experimental	Portuguese	Google Play reviews		
9	Ganeshbhai et al. (2015)	Survey/Review	Multiple	Feature-based OM		
10	Gedif et al. (2023)	Experimental	Amharic	Restaurant reviews	80.43	
11	Ma et al. (2011)	Hybrid/Method	Multiple	Web documents		
12	Rani et al. (2017)	Survey/Review	Multiple	Multiple		

Table 1 summarizes the key characteristics of sentiment analysis and opinion mining studies reviewed in this work. It includes author information, publication year, title, methods used, datasets, reported results, advantages, and limitations. The table shows that traditional machine learning algorithms particularly SVM, Naive Bayes, and KNN are the most used techniques. While high accuracy is often reported in domain-specific settings, limitations such as poor generalization, reliance on handcrafted features, and difficulty handling implicit sentiments remain prevalent.

5. Conclusion

This study analyzed machine learning approaches for opinion extraction from web text, focusing on methods, achieved results, advantages, limitations, and open challenges. The analysis shows that traditional machine learning models such as SVM, Naive Bayes, and KNN remain dominant due to their simplicity and effectiveness in domain-specific sentiment classification tasks. Hybrid approaches and feature engineering techniques have further improved performance in controlled settings. However, the review also reveals significant limitations, including strong domain dependency, reliance on handcrafted features, insufficient handling of implicit opinions, and limited support for multilingual and under-resourced languages. In addition, the lack of standardized datasets and consistent evaluation metrics restricts reproducibility and fair comparison across studies. Overall, while current machine learning approaches provide a solid foundation for opinion extraction, they are not yet robust enough to handle the complexity and diversity of real-world web text at scale. Future research should focus on developing domain-adaptive, aspect-aware, and multilingual opinion extraction models that can robustly handle noisy web language and implicit sentiment expressions.

References

1. N. Bansal and A. Singh, "A review on opinionated sentiment analysis based upon machine learning approach," *2016 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2016, pp. 1-6
2. S. Styawati, A. Nurkholis, A. A. Aldino, S. Samsugi, E. Suryati and R. P. Cahyono, "Sentiment Analysis on Online Transportation Reviews Using Word2Vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm," *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, Jakarta, Indonesia, 2022, pp. 163-167
3. H. Wan and F. Zhang, "Adaptive Web Information Extraction and Intelligent Analysis Based on Improved Neural Network," *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2021, pp. 1-4

4. Z. Ergul Aydin, "Sentiment Analysis about Turkish TV Series with Web Scraping," *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Ankara, Turkey, 2022, pp. 1-4
5. O. Bellar, A. Baina and M. Bellafkih, "Sentiment Analysis of Tweets on Social Issues Using Machine Learning Approach," *2023 International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA)*, Casablanca, Morocco, 2023, pp. 126-131
6. S. Abu Taher, K. Afsana Akhter and K. M. Azharul Hasan, "N-Gram Based Sentiment Mining for Bangla Text Using Support Vector Machine," *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, Sylhet, Bangladesh, 2018, pp. 1-5
7. A. H. A. M. Siagian and M. Aritsugi, "Combining Word and Character N-Grams for Detecting Deceptive Opinions," *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, Turin, Italy, 2017, pp. 828-833
8. F. L. Dos Santos and M. Ladeira, "The Role of Text Pre-processing in Opinion Mining on a Social Media Language Dataset," *2014 Brazilian Conference on Intelligent Systems*, Sao Paulo, Brazil, 2014, pp. 50-54
9. S. Y. Ganeshbhai and B. K. Shah, "Feature based opinion mining: A survey," *2015 IEEE International Advance Computing Conference (IACC)*, Banglore, India, 2015, pp. 919-923
10. B. Gedif, A. Alemu, Y. Assefa and S. Nibret, "Design Amharic Text Sentiment Analysis Model Using Machine Learning Techniques. In Case of Restaurant Reviews," *2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, Bahir Dar, Ethiopia, 2023, pp. 150-154
11. H. -B. Ma, Y. -B. Geng and J. -R. Qiu, "ANALYSIS OF THREE METHODS FOR WEB-BASED OPINION MINING," *2011 International Conference on Machine Learning and Cybernetics*, Guilin, China, 2011, pp. 915-919
12. M. S. Rani and S. Sumathy, "Analysis on various machine learning based approaches with a perspective on the performance," *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vellore, India, 2017, pp. 1-7