**Literature Survey:**

Prepare below table after reading and analysing IEEE Papers:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No** | **Title of Paper** | **Name of Authors** | **Published Year** | **Remarks / Findings** |
| 1 | Anomaly Detection through Enhanced Sentiment Analysis on Social Media Data | Zhaoxia WANG, Victor Joo Chuan TONG, Xin XIN | 2014 | Methodology: The methodology employed in this research involves a comprehensive approach to anomaly detection through enhanced sentiment analysis on social media data, particularly focusing on Twitter. To address the challenges posed by the diversity and size of social media data, the study first surveys existing anomaly analysis and sentiment analysis methods, highlighting their limitations and challenges. In response, an enhanced sentiment classification method is proposed, incorporating features such as object consideration, negation handling, emoticon interpretation, and special lexicon treatment. The proposed method is tested for applicability and robustness through sentiment analysis on tweet data. The results showcase the method's capabilities in identifying anomaly sentiment patterns, providing meaningful insights into this research area. The evaluation involves comparing the proposed method with human annotators, demonstrating a high level of agreement and validating its effectiveness in sentiment pattern classification. The research concludes by emphasizing the practical utility of the method for businesses, government organizations, and other entities seeking to understand and respond to sentiment anomalies in social media data. Future work is suggested to extend the method, including the detection and classification of detailed subcategories of sentiment emotions and addressing practical problems in various domains.  Algorithms: the proposed enhancement sentiment classification algorithm  Advantages:  1.Early Anomaly Detection: The enhanced sentiment analysis method provides a means to detect anomalies in sentiment patterns early. This early detection allows businesses and government organizations to intervene promptly or adopt appropriate strategies.  2. Real-time Monitoring: Social media data, especially from platforms like Twitter, provides a vast source of real-time information. The proposed method leverages this data to monitor sentiment patterns and identify abnormal events promptly.  3. Applicability to Various Domains: The research demonstrates the applicability of the proposed method to different domains, such as business and government. This versatility makes it a valuable tool for a wide range of applications.  4.Validation against Human Annotations: The method's classification results were compared to annotations made by human experts, showing a high level of agreement. This suggests that the proposed method is comparable to human annotators in sentiment pattern classification.  5.Insights for Risk Management: By analysing sentiment patterns and changes, the method provides insights into potential risks. For example, shifts in user attitudes detected through sentiment analysis may indicate impending crises or risks, allowing for proactive risk management.  Disadvantages:  1.Dependency on Data Quality: Like many machine learning methods, the proposed enhanced sentiment analysis method relies on the quality of training data. If the training data is not representative or of good quality, it may affect the accuracy and reliability of the anomaly detection.  2. Diversity and Size of Social Media Data: The diversity and size of social media data, especially on platforms like Twitter, pose challenges. Handling colossal amounts of data requires robust algorithms and computing resources, and ensuring diversity in training datasets is challenging.  3.Semantic Ambiguity in Lexicon-Based Approaches: While lexicon-based sentiment analysis methods are easily applicable to different datasets, they may suffer from semantic ambiguity. The accuracy of lexicon-based approaches is limited in dealing with the nuanced meanings of words in different contexts.  4.Resource Intensive Machine Learning Methods: Machine-learning-based sentiment analysis methods, while effective, can be resource-intensive. They require large training datasets and domain experts to clean up the data, making them costly and impractical for new domains.  5. Limited Generalizability: The machine-learning-based methods' generalizability is limited by the need for large and domain-specific training datasets. This limitation makes them less directly applicable when training datasets are not readily available.  **Applications:**   * Business Reputation Management * Customer Feedback Analysis * Market Research * Brand Management * Crisis Management |
| 2 | Distantly Supervised Lifelong Learning for Large-Scale Social Media Sentiment Analysis | Rui Xia, Jie Jiang, and Huihui He | 2017 | Methodology: The proposed methodology introduces a distantly supervised lifelong learning framework for large-scale social media sentiment analysis. Given the challenges of continuously increasing text data with dynamic topics in social media, the approach utilizes distant supervision, considering emoticons as natural sentiment labels in microblog texts. Unlike previous methods trained on isolated datasets, the lifelong learning approach sequentially learns on past tasks, retains acquired knowledge, and applies it to future learning. Evaluated on two large-scale distantly supervised social media datasets and nine benchmark datasets, the results demonstrate the feasibility and effectiveness of lifelong sentiment learning. The approach surpasses traditional single-task learning in both classification performance and computational efficiency, challenging the belief that more training data leads to better performance in large-scale social media sentiment analysis.  Algorithm:Lifelong Bagging: Averages or votes the predictions from classifiers trained on past partitions.  - Lifelong Stacking: Uses a meta-learning layer to re-learn the outputs of base classifiers from past partitions, enabling domain adaptation.  - Compatible with different single-task classification algorithms like logistic regression, naive Bayes, etc.  **Advantage**: - Continuous learning capability, retains and transfers knowledge from past tasks.  - Beats the traditional way of using all training data at once, in terms of performance and efficiency.  - Scalable to large-scale data by conducting sequential learning on partitions.  - Relies only on distant supervision from emoticons, no manual labelling needed.  - General framework adaptable to different single-task algorithms.  **Disadvantages:**  - Performance depends on the quality of distant supervision from emoticons, which can be noisy.  - Requires tuning of partition size and other hyper-parameters.  **Application: -** Large-scale social media sentiment analysis where data is continuously increasing and topics are dynamic.  - Domains where collecting enough manually labelled data is difficult. |
| 3. | Sentiment Analysis on social media | Federico Neri Carlo Aliprandi Federico Capeci Montserrat Cuadros Tomas | 2012 | Methodology:  - Uses a knowledge mining system with components like crawler, semantic engine, search engine, machine translation engine, geo-referencing engine, and classification engine.  - Crawler gathers data from web sources like social media (Facebook posts in this case).  - Semantic engine performs linguistic analysis, concept/entity extraction, sentiment analysis on the text data.  - Sentiment analysis considers polarity of words, syntax of sentences, idioms, negations etc.  **Algorithms:**  - Uses a bottom-up chart parser and slot grammar for linguistic analysis.  - Word sense disambiguation considers context and semantic relations.  - Sentiment analysis uses a combination of rules and lexicons to assign polarity scores.  - Classification uses supervised (Bayesian) and unsupervised (K-Means clustering) techniques.  **Advantages:**  - Can extract deep semantic information and sentiment from unstructured text data.  - Supports multiple languages like English, Italian, German, French etc.  - Allows querying data using natural language, semantic roles, combining concepts etc.  - Provides data visualization via clustering, geospatial mapping, concept graphs etc.  **Disadvantages:**  - Not clearly mentioned, but rule-based semantic/sentiment analysis can have limitations compared to modern neural models.  - Unsupervised techniques may not be as accurate as supervised for specific domains.  **Applications:**  - Demonstrated application is monitoring social media sentiment towards news channels/brands.  - Can be useful for marketing, reputation management, opinion mining etc.  - Used by government agencies for open-source intelligence gathering and analysis. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 4.  5.  6.  7. | Sentiment Analysis in Social Media and Its Application: Systematic  Literature Review  Sentiment Analysis of Users’ Reactions on Social Media during  the Pandemic  A comparative empirical study on social media sentiment analysis over various genres and languages  A Literature Survey on Sentiment Analysis  Techniques Involving Social Media and Online  Platforms | Zulfadzli Drus, Haliyana Khalid  Eldor Abdukhamidov  , Firuz Juraev  , Mohammed Abuhamad and Shaker El-Sappagh  and Tamer AbuHmed  Viktora Hangya – Richard Farkas  Raktim Kumar Dey, Debabrata Sardar, Indranil Sarkar, Rajesh Bose, Sandip Roy | 2019  2022  2024  2020 | **Methodology:**  - Lexicon-based (unsupervised) approach: Relies on sentiment lexicons/dictionaries to count positive and negative words. Common methods are SentiWordNet and TF-IDF.  - Machine learning (supervised) approach: Requires training data. Common algorithms are Naive Bayes, Support Vector Machines (SVM).  - Hybrid approach: Combining lexicon-based and machine learning methods.  **Advantages and Disadvantages:**  **Lexicon-based:**  **-** Simple counting of positive/negative words  - Flexible for different languages  - Fast analysis  - Struggles with sarcasm, negations, context-dependent meanings.  **Machine Learning:**  **-** Can learn complex linguistic patterns  - Requires large training data  **- T**ime-consuming model training, especially for complex models  - Performs poorly on noisy/irregular text like social media.  **Hybrid:**  **-** Combines strengths of both approaches  - Can improve handling of unstructured data  Generally better accuracy than single approach.  **Applications: - Business/Marketing:** Product/service feedback, brand monitoring, market analysis  - Politics: Election prediction, measuring public sentiment  - Healthcare: Detecting disease outbreaks, monitoring mental health  - Public Events: Analysing sentiment around disasters, sports, social issues  - Security: Early threat detection through sentiment monitoring.  **-** Twitter is by far the most commonly used, due to open API and real-time data  - Facebook data is messy with shortforms/errors, making analysis harder  - Other platforms like blogs, YouTube, Reddit used sometimes  The review highlights the trade-offs between the lexicon and machine learning approaches, and the potential benefits of combining them for better accuracy on the unstructured social media data. It also showcases the diverse applications of sentiment analysis for businesses, governments and public organizations.  **Methodology: -** Collected two large-scale COVID-19 related datasets from Twitter (131 million tweets from Jan-Jun 2020) and Instagram (3843 posts from Jan-Mar 2020)  - Performed sentiment analysis using Stanford CoreNLP, categorizing into 5 classes (very negative, negative, neutral, positive, very positive)  - Conducted topic modelling using Latent Dirichlet Allocation (LDA) on data during sentiment spikes to identify discussed topics  - Used Named Entity Recognition (NER) to identify mentions of countries in the data  - Performed word embedding using Word2Vec to analyze temporal patterns of words associated with top mentioned countries.  **Algorithm: -** Sentiment Analysis: Used Stanford CoreNLP based on deep learning compositional models over binarized sentence trees  - Topic Modelling: Utilized LDA implemented in Genism library along with Mallet  - Named Entity Recognition: Employed Stanford NER tagger  - Word Embedding: Implemented Word2Vec (CBOW and Skip-gram models) using Genism.  **Advantages: -** Large-scale datasets from popular social media platforms provide insights into public reactions during the pandemic  - Combining sentiment analysis, topic modelling, and temporal analysis gives a comprehensive understanding  - Identifying top mentioned countries and analysing their associated topics/reactions is insightful.  **Disadvantages:** - The datasets cannot be shared publicly due to privacy policies of social media platforms  - Instagram dataset covers a relatively shorter time period (Jan-Mar 2020) compared to Twitter.  **Applications:**  **-** Understanding public opinions, concerns, and reactions towards the COVID-19 pandemic through social media data  - Analysing how discussed topics and sentiments evolve over time and across geographic regions  - Identifying influential countries/localities and the focus of discussions around them during the pandemic.  **Methodology: -** Supervised machine learning approach using maximum entropy classifier for document-level and target-level sentiment analysis  - Document-level features: n-grams, word polarities from sentiment lexicons, character repetitions**,** negations, topic modeling (LDA)  - Target-level features: target name as feature, distance-weighted bag-of-words to capture target-relevant text, syntax-based features using dependency and constituency parsing to identify clauses related to target.  **Algorithm:**  - Preprocessing steps like lowercasing, stemming, replacing Twitter-specific elements, handling emoticons, numbers  - For word polarity, used SentiWordNet lexicon for English, custom lexicon for Hungarian  - Distance weighting of n-gram features based on distance from target mention  - Syntax-based: Dependency parsing to identify adjective modifiers of target, constituency parsing to extract relevant subtrees containing target.  **Advantages: -** Comparative analysis across different text genres (reviews, Twitter) and languages (English, Hungarian)  - Techniques to handle informal text like Tweets - preprocessing, distance weighting  - Usage of linguistic analysis like parsing to identify target-relevant text segments  - Combination of lexicon features and syntactic features.  **Disadvantages:** - Errors from lack of common-sense/background knowledge, unseen words, comparing to implicit references  - Parsers trained on well-formed text perform poorly on Tweets  - Morphological richness of languages like Hungarian causes data sparsity issues**.**  **Applications:** - Marketing - Analysing opinions about products/brands on social media  - Political campaigns - Monitoring sentiments towards candidates/issues  - Consumer analysis - Extracting positive/negative opinions about product aspects.  **Methodology: -** Sentiment analysis can be performed at document level, sentence level, or aspect level.  - It involves classifying text as positive, negative or neutral sentiment.  **Algorithm:** - Machine learning approaches:  - Supervised learning: Naive Bayes, Support Vector Machines (SVM), Decision Trees, Neural Networks, Maximum Entropy  - Unsupervised learning: Lexicon-based approaches  - Lexicon-based approaches:  - Dictionary-based  - Corpus-based  - Hybrid approaches combining machine learning and lexicon-based methods.  **Advantages:** - Enables understanding opinions and sentiments towards products, services, events, etc. from user-generated data  - Can help businesses make data-driven decisions based on public sentiment  - Useful for opinion mining in politics, marketing, product reviews, social media analysis, etc.  **Disadvantages:**  **-** Accuracy challenges like handling negations, sarcasm, context.  - Data sparsity problem.  - Polarity shift issue.  **Applications:**  - Politics - Gauging public sentiment towards political personalities, policies  - Business - Analysing opinions on products, services to make business decisions  - Social media monitoring  - Market research  - Movie/product reviews analysis. |