International Journal of Research and Review
Volume 11; Issue: 5; May 2024
Website: www.ijrrjournal.com
Review Paper *E-ISSN: 2349-9788; P-ISSN: 2454-2237*

Multimedia Data Mining and Processing for News Source Attribution

Ayodeji Abimbola Owonipa[1], Taye Oladele Aro[2], Oyenike Adunni Olukiran[3], Olakunle Isaac Ifawoye[4], Aishat Oladayo Jimoh-Mahmud[2], Temitope Ayanladun Oyelakun[5]

[1]School of Informatics, University of Edinburgh, United Kingdom
[2]Department of Computer Science, Al-Hikmah University Ilorin, Kwara State, Nigeria
[3]Department of Computer Science, Precious Cornerstone University Ibadan, Oyo State, Nigeria
[4]Information and Communication Technology, Bingham University Karu, Nigeria
[5]Department of Information System, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria

Corresponding Author: Aro Taye Oladele

DOI: https://doi.org/10.52403/ijrr.20240507

ABSTRACT

The desire for unbiased journalism that effectively counters disinformation is widely recognised. News consumers are not only interested in news, but they also want unbiased journalism that cuts through disinformation, and they want it from trusted news sources. Consequently, media researchers need to explore ways to facilitate news-source identification, irrespective of the platform used. However, the availability of multimedia data sources has seen a remarkable surge in recent years, encompassing demographic data, social media data, geodata, and pervasive digital trace data. Multimedia data mining is a procedure of discovering stimulating trends via media data using video, text, and audio that are not generally available by simple enquiries and related outputs. Researchers face the challenge of integrating these diverse sources to enhance news source attribution in multimedia data including platforms like Facebook, WhatsApp and Instagram. The paper presents a review of multimedia data approaches and their application to news source attribution research. Also, the examination of the benefits and limitations of these techniques and discussion on future directions were mentioned. Consideration was on machine learning and statistical approaches to multimedia data, which include deep learning, and probabilistic modelling. Similarly, a discussion on the importance of data privacy and ethics in news source attribution research was stated. The contribution of this study is highly relevant for news media research groups striving to improve their capability to attribute sources in multimedia data, thereby combatting disinformation and amplifying trusted media brands.

Keywords: *Data mining, Multimedia, Data process, News source attribution, Unbiased journalism*

1. INTRODUCTION

Transformation of news has been achieved considerably after so many years of development, on the other hand, news propagation has experienced remarkable modifications after over a thousand years of historic transformations [1]. Journalistic facts are structured for several means (audio, images and audio) occupied with the insight of efficacy and truthfulness or originality, and are engrossed by people using sensations or senses (sight, touch, smell and hearing) [2]. Traditional or

common news spreading, from a bit of oral conveyance to newspapers, wireless broadcasting, and TV broadcasting, blooms in different approaches and has many means [3] At present age, the news is reachable on the mobile end of individual, and convectional news circulation techniques have moderately uncommunicative from the public. Attribution provides news or stories trustworthiness, credibility, reliability, and perspective, productive application of attribution is a matter of ethics in journalism and good writing [4]. Attribution is a main component in the credibility of any story to reveal the origins of the facts within reach in an article using attribution or ascription, which is generally taken in paraphrasing forms as well as straight and non-straight quotes [5]. Employing an attribution is necessary for media writing, as it contributes to the establishment of an objective tone and gives dependability to an article [6]. Attribution describes how the writer reclaims the information and why a specific source was repeated or iterated. Most of the important facts should be ascribed, through phrases like "she said" or "according to a recent report." The basis of attribution is to determine which medium and piece of information produced a significant impact on the conclusion to transform or consider the better next step [7]. Many common modes for attribution such as multi-touch attribution, lift studies, time decay, and so on are applied by marketers today.

Accordingly, attribution of the source and mostly the use of straight quotations are critical to the journalists' professional practice [8]. Editors give important value to the free verification and information corroboration from sources, with textbooks of journalism asserting that for a fact to be shown in print, it requires to be accepted by at least two dependable and independent sources [6]. While the reality may be somewhat different, there is however a strong commitment within journalism to the direct quotation principles and, on the part of editors, a reticence to publish stories where its sources seek anonymity [9].

The attribution of news sources is an important issue in the world of journalism as a continuous proliferation of online stores with little to no sourcing has allowed for gossip or rumours to gain wide circulation to the point of even obtaining the status of 'news [10]. This has in turn brought about a gradual decline in the viewer's perspective of the importance of news sources to the veridical nature of online news [11]. There have been several approaches to data integration for news media sources, particularly in the areas of machine learning and deep learning. By chaining together different techniques in automated machine learning, text mining, and statistical analysis, one can to a certain extent determine who, where, and what is being reported on the news [12].

Nowadays, in the communication or broadcasting industries, attribution and translation are considered to be one of the key challenges from the advance of new media based on the Internet: credibility erosion [13]. Recent media and the digital environment have changed how news is provided, disseminated and accessed. One of the main challenges faced by the media industry today is the problem of credibility due to issues associated with the attribution of news sources. Through the digital environment, members of the public can produce, distribute, and access news content that contains media from numerous sources such as online communities or even people's cell phones [14]. Navigating through the numerous media formats originating from not just traditional sources but also from the ever-increasing internet population scattered across the web can serve as a barrier to those seeking a more detailed understanding of the data, especially in cases where the data are connected in such a way that they become evident only when analyzed together [15]. Current multimedia applications are full-motion videoconferencing, sophisticated imaging, electronic books and newspapers,

electronic classroom presentation technologies, and graphics design tools.

The remaining aspect of the paper is arranged as follows: section 2 mentions the review of literature on techniques of data mining, machine learning techniques and multimedia data approach, section 3 presents related work, section 4 discusses state of art and challenges associated with multimedia data and finally, section 5 represents conclusion.

2. LITERATURE REVIEW

2.1 Data Mining

Data mining involves the method of selecting knowledge from large data [16]. In other words, big data is an approach to identifying relevant patterns in huge and complex datasets. Several definitions like information mining from data, information harvesting, information analysis, and data dredging, possess synonyms that are the same or little distinct from data mining. Knowledge Discovery from Data also referred to as KDD, is another generally employed phrase that data mining uses. Some researchers see data mining as just a crucial phase in knowledge discovery when intelligent techniques are applied to obtain patterns in data. This technique is an important stage of knowledge discovery in a database that is used for selecting patterns from data [17]. The patterns that can be identified are determined by the tasks of data mining used through the statistics intersection, machine learning, and databases [18]. The predominant techniques of data mining include regression, classification, association rule learning, and clustering [19]. Data mining has been usually applied to well-structured data, using the explosion of multimedia data videos, audio, images and web pages, many studies have felt the necessity of data mining to solve the problem of unstructured data [20].

2.2 Techniques of Data Mining

These techniques are employed for data sorting to find patterns (knowledge discovery). The techniques include classification, prediction, association, clustering, and so on [21].

2.2.1 Classification

Classification categorizes a specific group of items into targeted classes [22]. The main goal of classification is to accurately predict the nature of items or data based on the obtainable classes of items [23]. The classification technique is an important analytical mechanism in the prediction of diverse levels of accuracy [24]. Multimedia data belong to different classes of domains such as Entertainment, Sports, News and Music [25]. Automatic classification of multimedia data without knowing is required. For example, a classification model could be used to identify loan applicants as having low, medium, or high credit risks. The algorithms for classification include Naïve Bayesian, Support Vector Machine, Neural Network, Decision Tree, and K-Nearest Neighbour [26]

2.2.2 Association Rule Learning (ARL)

This is an approach in machine learning for discovering interesting relations between varieties in huge datasets [27]. It is used to know the rules that are discovered in databases. ARL finds a relationship that exists between variables or looks for patterns based on the connection of a particular event to other events [28]. ARL can be used in several domains (large databases of one thousand to one million datasets as well as small datasets) to find associations, and frequent patterns from the sets of objects in datasets [29]. It is a commonly used method for heart disease prediction as it produces the correlation of different features for examination and categorizing outpatients with all risk factors needed for prediction [30].

2.2.3 Clustering

This is a method in which a given data set is divided into collections called clusters in such a way that similar data points are

collected in a single cluster [31]. Clustering plays an essential part in data mining due to the large number of data sets. These techniques are used to analyze data objects without knowing their class labels (position elements of data into related groups without advanced knowledge of the group definitions) [32]. In clustering, a set of data is categorized based on a certain condition such as the similarity among objects. Individual data in a cluster is correlated with other data in the same cluster (homogeneous data) [33]. Segmentation of the market or identifying common characteristics for groups of people is an example of an application where clustering. The commonly employed techniques of clustering include density-based spatial clustering, k-means clustering, mean-shift clustering, and expectation maximization (EM) clustering.

2.2.4 Data Visualization

Visualization of data involves the transformation of information into a visual context, such as a graph or map, to make data easier for humans to comprehend and deduce information from [34]. Employing visual elements such as charts, graphs, and maps enables explicit analysis of experimental results in any study. Tools of data visualization provide an available approach to identifying and understanding outliers, trends, and patterns in data. One of the stages of data science is data visualization, bioinformatics and data mining that states once data has been acquired, processed and modelled, it needs to be visualised for deductions to be made [35]. The graphical representation of data in visual form is also an element of the extensive data presentation architecture (DPA) field, which aims to find, manipulate, format and convey data in the most well-organised method possible. Visualisation of data helps researchers see, interact with, and better understand data. Whether complex or simple, the right visualization can bring everyone on the same page irrespective of their level of expertise.

2.2. Multidata Integration Techniques

Data integration involves the data combination from many origins into a sole and integrated pattern [36]. This includes cleaning and converting the data, as well as resolving any irregularities or conflicts that may exist between the different sources [37]. The objective of data integration is to make the data more useful and meaningful for analysis and decision-making [38]. Techniques used in data integration include data warehousing, ETL (extract, transform, load) processes, and data federation [39]. These techniques, still widely used today are often referred to as the traditional approaches to data integration and serve as core components of important data management projects such as building data warehouses and synchronizing data between applications [40]. The various scenarios of data integration across these numerous projects enable them to be mapped into three fundamental approaches namely consolidation, propagation and federation [41]. Data consolidation involves a wholesale transfer of data from one or more systems to another and is usually applied in business intelligence for centralizing data from multiple systems into a single data warehouse. Data propagation involves a continuous process of transferring data in which an automated program or database tool copies the data from one system to another [42]. It is the simplest approach to implement for repetitive data integration thus making it the most popular approach. Data federation is a method where data is centralized without it being physically consolidated first and is seen as an on-demand data integration approach. Data access and integration are incorporated into the model which is then invoked whenever the data is requested by an application [43].

2.3 Machine Learning for Multimedia Data Processing

Development of the corresponding smart and automated systems involves the application of the knowledge of artificial intelligence (AI), in which machine learning (ML) remains the main component [44]. Different machine learning algorithms such as supervised, unsupervised, semi-supervised, and reinforcement learning exist [45]. In addition, deep learning is part of an expansive part of machine learning that can logically study huge data. Techniques of the ensemble are learning algorithms that design classifiers and use these classifiers for new data classification points established on a weighted vote of their predictions [46]. The method is broadly employed as it more than often outputs that outperforms any single classifier [47]. In the area of data integration, ensemble methods aggregate multiple models or data integration techniques to build more comprehensive predictive models [48].

This strategy usually focuses on building uniformly integrated representations to reinforce the consensus among multiple data modalities [49]. Unlike traditional data integration techniques which suffer from drawbacks and limitations such as their limited modularity when sources of data are more as well as scalability to many datasets, ensemble methods on the other hand can obtain results comparable with these data integration techniques while also exploiting the modularity and scalability that characterize most of the ensemble algorithms[50]. Furthermore, whenever current data or updates of data are made available, ensemble machine learning methods can embed the new data sources or update the existing ones by training only the base learners devoted to this new data without having to retrain the entire ensemble model. Thus it can be said that ensemble methods scale well with the number of available data sources while avoiding problems associated with other data integration approaches [51].

2.4 Multimedia Data Process and Application in Deep Learning

With deep learning's existence extensively applied in many research fields, it has been duly made known in the research and multimedia data processing technology and application [52]. First, the multimedia data flow processing, the development of multimedia data and the realization of multimedia data processing technology are clarified and analysed. With the development of high-throughput technologies, data is continuously accumulated at an astonishing rate. This large amount of data generated from multiple samples cannot be easily integrated with the standard data integration techniques. The use of deep learning techniques is therefore crucial for extracting valuable knowledge from the data [53]. The use of deep learning techniques in data integration has allowed for the establishment of numerous results in existing problems such as information extraction, data cleaning, entity matching, etc. This is due to their capacity for learning and deriving inferences from given samples, thus making them best suited for data integration challenges where rules are difficult or impossible to specify. Another advantage of the use of deep learning techniques in data integration is their robustness to data imperfections like occurring in multiple data formats or having multiple values [8]. Overall, the numerous capabilities of deep learning models enable them to effectively integrate complex data sources while providing all sorts of benefits from techniques such as automated feature extraction, multi-modal fusion, transfer learning, etc., further allowing for improved integration accuracy and performance.

2.5 Probabilistic Modelling for Multimedia Data

Probabilistic modelling is a method that statistically uses the effect of random incidences or activities in predicting the chance of future results [54]. These models are machine learning techniques purposely

for predictions based on the important principles of statistics and probability. The models identify undefined relationships between variables in a data-driven manner while capturing the underlying trends or patterns in data [55]. Probabilistic modelling thus provides a framework for quantifying and representing the uncertainty associated with combining data from multiple sources. By modelling the connections between different variables, the combination of heterogeneous data types from several sources can be achieved while considering the uncertainty in the combination process.

2.6 Related Work

Research overview on multimedia mining [56]. Fundamental concepts of multimedia mining and its relevant features were mentioned. The mining of multimedia architectures for data is structured and unstructured, research problems in multimedia mining, models of data mining applied in multimedia mining, and applications were discussed. It allows researchers to obtain information about how to do their work in the multimedia mining field. Lima Jr and Walter [2] discussed the tasks of demonstration for a formalist computational procedure, the knowledge that the journalists employ to articulate the values of news to select and impose a hierarchy on news. The discussion on how to get bridges to follow the knowledge gathered in an empirical pattern with the computational science bases, in the area of storage, recovery, and linked to data in a database was considered, which reveals the method human brains handle the facts gathered by sensorial system. Automating or systemizing sections of the journalistic process in a database gives a chance to remove distortions or errors and use effective data mining techniques or texts which by definition enable the detection of non-trivial relations.

A text classification was applied for text classification using words, phrases, or combining words to form predefined class labels [57]. News data were categorised into four predefined classes namely Business, Entertainment, Sports, and Technology. The simulation of text classification was performed using WEKA an open-source data mining tool with different classification algorithms applied to the News dataset. A comparative analysis of the algorithms was conducted based on accuracy, time taken, errors, and ROC Curve to predict the best algorithm for news dataset classification. Results for evaluation metrics showed that the naïve-bayes multinomial algorithm is best for news classification. Algur, Basavaraj, Goudannavar and Bhat [25] presented two techniques for the classification of web multimedia data via web multimedia data classification using dimension reduction techniques and multimedia data classification without reducing the dimensions. The reduction of the dimension of the web multimedia metadata was achieved with the Principal Component Analysis (PCA) technique. The proposed PCA method considered the orthogonal transformation of multimedia metadata values, covariance matrix construction, and eigenvalues computation to decrease the dimensions. The reduced and non-reduced multimedia data were classified separately using DT and KNN classifiers. The classification outputs of reduced and non-reduced dimensions of multimedia data were comparatively analysed.

Strategy for data news dissemination in decision-making applying the new media platforms of the big data era [58]. The study presented China data journalism by analysing challenges in data journalism and the trend of development in the future. The study combined data and news to discover the status quo and dig out the present problems. The status of work, approaches and theoretical basis for propagation of data journalism's path were also considered. Lasswell's 5W model, a new model was employed to analyse data news from five aspects; disseminator, disseminating channel, dissemination content, audience and effect of dissemination. It was

concluded based on searching, content analysis and data mining, an indicator system was built for the current dissemination of media's news effect evaluation, and the Delphi approach was employed to assign weights to many indicators and make decisions based on them. By analysis, the study identified the challenges in the method for the combination of data journalism and current media platforms and proffered solutions for the strategy of future communication in data journalism.

Bhatt and Kankanhalli [59] came up with a review of the hitches and resolutions in multimedia data mining, approaches from the different aspects: feature transformation, extraction and representation techniques, and current multimedia data mining in many application areas. The major areas of feature extraction, transformation, and representation techniques were discussed. The areas are feature extraction level, a fusion of features, synchronization of features, feature correlation discovery and accurate multimedia data representation. The MDM techniques comparison with video processing, audio processing and image processing techniques was provided. Similarly, the comparison of MDM techniques with data mining techniques involving clustering, classification, sequence pattern mining, association rule mining and visualization. A review of current multimedia data mining in detail, classifying them based on the formulations of problems and approaches was conducted.

Veglis et al [60] attempted to highlight the significance of the utilization of big data in the media industry. The circumstances of exploitation for big data such as consumption of media content and management, production of data journalism, utilization of social content and applications of participatory journalism were explained. It was observed that big data had established changes in all stages of the journalism practice from the production of news to the distribution of news by making use of the available instruments. The new developments that are associated with semantic web (Web 3.0) technologies, which have already started to be adopted by media organisations around the world were discussed. Shao [61] applied LDA and ARIMA models to compute and analyze the popularity measurement and trend analysis of new media reports under the big data mining background. The model designed showed that the missed detection rate was reduced by 75.4%. Experiment analysis revealed that the accuracy of heat topic detection of the model designed can attain 84.6%. In trend analysis, the first stage of transmission is referred to as the incubation era, then after a certain critical point, it will come to the outbreak era. The outbreak era lasted for an era of time, entered an era of plateau and lastly came to a subsidence era.

4.0 CHALLENGES ASSOCIATED WITH MULTIMEDIA DATA FOR NEWS ATTRIBUTION

Developments of multimedia data acquisition and tools for storage have resulted in the advance of huge multimedia datasets [59]. The study of large amounts of multimedia data such as news data to find useful knowledge is a perplexing task. This challenge has revealed the opportunity for research in Multimedia Data Mining (MDM). Multimedia data mining can be defined as the process of finding interesting patterns from media data such as audio, video, image, and text that are not ordinarily accessible by basic queries and associated results. With the large amount of structured and unstructured data constantly generated on the internet, data integration becomes a challenging process, particularly in areas such as storage, security, management, maintenance and privacy. Integrating large amounts of data becomes a tedious process due to the differences in data formats, structures, schema and features [62]. The disparity of the data is one of the main challenges faced when adopting traditional approaches to data integration. Integrating heterogeneous data is highly difficult as it occurs in various formats such as XML or

JSON, and is spread across various data storage infrastructures like databases and cloud storage. This can further lead to system interoperability as a result of compatibility issues arising from differences in systems or platforms used [63]. Another challenge faced when integrating data from multiple sources is the problem of scalability. Scalability issues tend to crop up whenever new data from multiple resources is integrated with data from legacy systems, leading such systems to pass through numerous modifications and updates to meet the requirements of the new technologies it is to be integrated with [64]. The quality of data is another issue to be considered as integrating data from various sources with various data formats and various entry methods can lead to inconsistencies and data quality issues. Data integration thus requires an understanding of the origin of the data as well as the characteristics and intended use of the data. This process can be tedious as different statistics and algorithms are required depending on the type of data involved [65]. All in all, several challenges are often faced with the use of traditional approaches to data integration. However, advances in data integration techniques such as deep learning, and ensemble methods have provided means to address these challenges while improving the effectiveness of these data integration approaches.

4.1 Summary and Discussion

Exploration of more innovative data methods in the areas of artificial intelligence, real-time data integration, diversification, large data platforms, and data security are of necessity in news source attribution. Applications such as natural language processing and network analysis could also be considered in data integration research in the area of news source attribution. Additionally, ethical considerations and data privacy issues should be addressed in the domain of news source attribution. As news source attribution requires the journalist to be specific, attributing news sources will usually require as much information about the source as possible even in cases where such sources are confidential. News media research groups should put careful consideration to respecting privacy rights while implementing ethical guidelines to maintain public trust and ensure integrity. For example, journalists should ensure transparency with the sources they provide, ensuring the reliability and truthfulness of the information. It is also of utmost importance to ensure the privacy and protection of the sources being attributed and guidelines should be established to safeguard their identities. In the case of online sources, the authenticity of these sources should be verified to avoid misinformation. Also, the provision of more engaging and comprehensive ways to convey information, making news stories more accessible and compelling to diverse audiences should be available to people. All these principles should be observed during news source attribution to provide accurate and reliable information in the news while upholding ethical standards.

5.0 CONCLUSION

Multimedia data involves the combination of video podcasts, audio slideshows, texts and animated videos. A presentation of multimedia is a coordinated and, conceivably, interactive delivery of multimedia data to users. Multimedia allows journalists to present a richer context, capture attention, and offer a more immersive experience for readers or viewers. This paper presents a review of multimedia data processes applying techniques of data mining, as well as some more advanced data mining approaches which prove enhancement on the existing traditional approaches in news ascription or credit. Also, challenges encountered by news attribution for multimedia data were briefly mentioned.

Declaration by Authors
Acknowledgement: None
Source of Funding: None
Conflict of Interest: The authors declare no conflict of interest.

REFERENCES

1. J. Tian, H. Fan, and Z. Hou, "Research on the Prediction of Popularity of News Dissemination Public Opinion Based on Data Mining," *Comput. Intell. Neurocience*, pp. 1–10, 2022.
2. W. T. Lima Jr, "Texts and data mining and their possibilities applied to the process of news production," *Brazilian Journal. Res.*, vol. 4, no. 1, pp. 104–120, 2008, doi: 10.25200/bjr.v4n1.2008.136.
3. A. Utesheva, D. Cecez-Kecmanovic, and D. Schlagwein, "Understanding the digital newspaper genre: Medium vs. message," in *ECIS 2012 - Proceedings of the 20th European Conference on Information Systems*, 2012, pp. 1–14.
4. S. S. Sundar, "Effect of source attribution on the perception of online news stories," *Journal. Mass Commun. Quarterly*, vol. 75, no. 1, pp. 55–68, 1998, doi: 10.1177/107769909807500108.
5. I. Elorza, "Newsworthiness, attribution and lexicogrammatical strategies in two types of news articles in English and Spanish," *Top. Linguist.*, vol. 14, no. 1, pp. 16–33, 2014, doi: 10.2478/toppling-2014-0009.
6. M. Bednarek, "Voices and values in the news: News media talk, news values and attribution," *Discourse, Context Media*, vol. 11, pp. 27–37, 2016, doi: 10.1016/j.dcm.2015.11.004.
7. J. Matthews, "News sources and perceptual effects: an analysis of source attribution within news coverage of alleged terrorist plots.," 2010. [Online]. Available: http://eprints.bournemouth.ac.uk/16207/
8. E. Bielsa and S. Bassnett, "Translation in Global News," *Transl. Glob. News*, vol. 1, pp. 137–155, 2008, doi: 10.4324/9780203890011.
9. F. A. Parastu Dorrimanesh, Seyed Vahid Aqili, "News Translation : A Unique Form of Communication Production," *J. Lang. Transl.*, vol. 13, no. 1, pp. 193–208, 2023.
10. R. Scollon, "Attribution and power in Hong Kong news discourse," *World Englishes*, vol. 16, no. 3, pp. 383–393, 1997, doi: 10.1111/1467-971X.00072.
11. B. Hladká, J. Mírovský, M. Kopp, and V. Moravec, "Annotating Attribution in Czech News Server Articles," in *2022 Language Resources and Evaluation Conference, LREC 2022*, 2022, pp. 1817–1823.
12. M. Atkinson and E. Van Der Goot, "Near Real Time Information Mining in Multilingual News," in *Proceedings of the 18th International World Wide Web Conference*, 2009, pp. 1153–1154. doi: 10.1145/1526709.1526903.
13. L. Koivunen-Niemi and M. Masoodian, "Visualizing narrative patterns in online news media," *Multimed. Tools Appl.*, vol. 79, no. 1–2, pp. 919–946, 2020, doi 10.1007/s11042-019-08186-9.
14. J. Hong, "Translation of attribution and news credibility," *Journalism*, vol. 22, no. 3, pp. 787–803, 2021, doi: 10.1177/1464884918775201.
15. J. Ventura *et al.*, "IssueBrowser: Knowledge acquisition via multimedia data," in *Proceedings of The Future of Interactive Media: Workshop on Media Arts, Science, and Technology (MAST)*, 2009, pp. 29–30.
16. N. Dey and D. N. Le, "Biological data mining: Techniques and applications," *Min. MuMultimedia. Doc.*, vol. 1, no. 4, pp. 161–172, 2017, doi: 10.1201/b21638.
17. H. Karim and K. Zand, "AComparative Survey on Data Mining Techniques for Breast Cancer Diagnosis and Prediction," *Indian J. Fundam. Appl. Life Sci.*, vol. 5, no. 1, pp. 4330–4339, 2015.
18. F. Hajjej, M. A. Alohali, M. Badr, and M. A. Rahman, "A Comparison of Decision Tree Algorithms in the Assessment of Biomedical Data," *Biomed Res. Int.*, pp. 1–9, 2022, doi: 10.1155/2022/9449497.
19. H. Sahu, S. Sharma, and S. Gondhalakar, "A Brief Overview on Data Mining Survey," *Ijctee*, vol. 1, no. 3, pp. 114–121, 2008.
20. V. A. Petrushin and L. Khan, "Multimedia data mining and knowledge discovery," *Multimed. Data Min. Knowl. Discov.*, pp. 1–27, 2007, doi: 10.1007/978-1-84628-799-2.
21. M. A. Khaleel, S. K. Pradhan, and G. N. Dash, "Finding Locally Frequent Diseases Using Modified Apriori Algorithm," *Int. J. Adv. Res. Comput. Coomunication Eng.*, vol. 2, no. 10, pp. 3792–3797, 2013.

22. N. Satyanarayana, C. Ramalingaswamy, and Y. Ramadevi, "Survey of Classification Techniques in Data Mining," *IJISET - International J. Innov. Sci. Eng. Technol.*, vol. 1, no. 9, pp. 268–278, 2014, [Online]. Available: www.ijiset.com
23. A. Oluwaseun and M. S. Chaubey, "Data Mining Classification Techniques on the" *Glob. Sci. Journals*, vol. 7, no. 4, pp. 79–95, 2019, doi: 10.11216/gsj.2019.04.19671.
24. E. T. Ranjan Baitharu and D. S. Kumar Pani, "A Comparative Study of Data Mining Classification Techniques using Lung Cancer Data," *Int. J. Comput. Trends Technol.*, vol. 22, no. 2, pp. 91–95, 2015, doi: 10.14445/22312803/ijctt-v22p118.
25. S. P. Algur, B. A. Goudannavar, and P. Bhat, "Classification and Analysis of Web Multimedia Data using Principal Component Analysis," *Int. J. Eng. Comput. Sci.*, vol. 6, no. 1, pp. 19842–19850, 2017, doi: 10.18535/ijecs/v6i1.01.
26. E. Neelamegam, S., & Ramaraj, "An Overview of Classification Algorithm in Data Mining," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 12, pp. 255–257, 2015, doi: 10.17148/IJARCCE.2015.41259.
27. O. O. Oladimeji, W. K. Oladipo, and A. M. Olalekan, "Application of Association Rule Learning in Customer Relationship Management Application of Association Rule Learning in Customer Relationship Management," in *Proceedings of the 14th iSTEAMS Multidisciplinary Conference*, 2019, pp. 29–36.
28. A. S. Ahmad, M. Dalhatu, A. U. Inuwa, and Y. J. Gambo, "Association Rule Mining Unsupervised Machine Learning Apriori Algorithms for Employee Loan Application," *Int. J. Sci. Res. Methodol.*, vol. 10, no. 2, pp. 28–67, 2018.
29. R. Revathy and S. Balamurali, "An improved Frequent Pattern Mining in Sustainable Learning Practice using Generalized Association Rules," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 2S2, pp. 776–780, 2019, doi: 10.35940/ijitee.b1118.1292s219.
30. D. C. Bindushree, "Prediction of Cardiovascular Risk Analysis and Performance Evaluation Using Various Data Mining Techniques: A Review," *Int. J. Enginnering Res.*, vol. 5013, no. 5, pp. 796–800, 2016.
31. Garima, H. Gulati, and P. K. Singh, "Clustering techniques in data mining: A comparison," in *2015 International Conference on Computing for Sustainable Global Development, INDIACom 2015*, 2015, pp. 410–415.
32. Saroj & Chaudhary, "Study on Various Clustering Techniques," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 3, pp. 3031–3033, 2015.
33. M. K. Keleş, "Breast Cancer Prediction and Detection Using Data Mining Classification Algorithms: A Comparative Study," *Teh. Vjesn.*, vol. 26, no. 1, pp. 149–155, 2019.
34. L. Jiang, "Information Visualization Based on Visual Transmission and Multimedia Data Fusion," *Int. J. Inf. Technol. Syst. Approach*, vol. 16, no. 3, pp. 1–14, 2023, doi: 10.4018/IJITSA.320229.
35. Y. Wang, Z. Zhu, L. Wang, G. Sun, and R. Liang, "Visualization and visual analysis of multimedia data in manufacturing: A survey," *Vis. Informatics*, vol. 6, no. 4, pp. 12–21, 2022, doi: 10.1016/j.visinf.2022.09.001.
36. A. A. Maryoosh and E. M. Hussein, "A Review: Data Mining Techniques and Its Applications," *Int. J. Comput. Sci. Mob. Appl.*, vol. 10, no. 3, pp. 1–14, 2022, doi: 10.47760/ijcsma.2022.v10i03.001.
37. Q. Chang and J. Hu, "Research and Application of the Data Mining Technology in Economic Intelligence System," *Comput. Intell. Neurosci.*, pp. 1–11, 2022, doi: 10.1155/2022/6439315.
38. A. Kadadi, R. Agrawal, and C. Management, "Encyclopedia of Big Data," *Encycl. Big Data*, pp. 1–6, 2020, doi: 10.1007/978-3-319-32001-4.
39. M. Lenzerini, "Data integration: A theoretical perspective," *Proc. ACM SIGACT-SIGMOD-SIGART Symp. Princ. Database Syst.*, no. June, pp. 233–246, 2002, doi: 10.1145/543613.543644.
40. R. Ly, "Data Quality Management in Data Integration," 2020. [Online]. Available: https://is.muni.cz/th/bey6e/Data_Quality_Management_in_Data_Integration_final_version_Archive.pdf
41. M. A. Jassim and S. N. Abdulwahid, "Data Mining preparation: Process, Techniques and Major Issues in Data Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1090, no. 1, pp. 1–9, 2021, doi: 10.1088/1757-899x/1090/1/012053.

42. T. Siddipet, T. State, G. Degree, and T. State, "Data Mining : Techniques, Tools and its Challenges," *Int. J. Creat. Res. Thoughts*, vol. 8, no. 7, pp. 3913–3920, 2020.
43. A. Delgado, A. Marotta, L. González, L. Tansini, and D. Calegari, "Towards a data science framework integrating process and data mining for organizational improvement," in *ICSOFT 2020 - Proceedings of the 15th International Conference on Software Technologies*, 2020, pp. 492–500. doi: 10.5220/0009875004920500.
44. A. Müller, L. S. Christmann, S. Kohler, R. Eils, and F. Prasser, "Machine Learning for Medical Data Integration," *Stud. Health Technol. Inform.*, vol. 302, pp. 691–695, 2023, doi: 10.3233/SHTI230241.
45. H. Munir, B. Vogel, and A. Jacobsson, "Artificial Intelligence and Machine Learning Approaches in Digital Education: A Systematic Revision," *Inf.*, vol. 13, no. 4, pp. 1–26, 2022, doi: 10.3390/info13040203.
46. G. Mainenti, L. Campanile, F. Marulli, C. Ricciardi, and A. S. Valente, "Machine learning approaches for diabetes classification: Perspectives to artificial intelligence methods updating," *IoTBDS 2020 - Proc. 5th Int. Conf. Internet Things, Big Data Security.*, vol. 2021, pp. 533–540, 2020, doi: 10.5220/0009839405330540.
47. T. G. Dietterich, *Ensemble methods in machine learning*, vol. 1857 LNCS. 2000. doi: 10.1007/3-540-45014-9_1.
48. L. R. Brewster *et al.*, "Development and application of a machine learning algorithm for classification of elasmobranch behaviour from accelerometry data," *Mar. Biol.*, vol. 165, no. 4, pp. 1–19, 2018, doi: 10.1007/s00227-018-3318-y.
49. L. Wang, J. Law, T. M. N. Murali, and G. Pandey, "Data integration through heterogeneous ensembles for protein function prediction," *bioRxiv*, pp. 1–8, 2020, [Online]. Available: https://www.biorxiv.org/content/10.1101/2020.05.29.123497v1.abstract?%3Fcollection=
50. E. Evangelista, "An Optimized Bagging Ensemble Learning Approach Using BESTrees for Predicting Students' Performance," *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 10, pp. 150–165, 2023, doi: 10.3991/ijet.v18i10.38115.
51. M. Re and G. Valentini, "Simple ensemble methods are competitive with state-of-the-art data integration methods for gene function prediction," *J. Mach. Learn. Res. W&C Proc.*, vol. 8, pp. 98–11, 2010, [Online]. Available: http://jmlr.csail.mit.edu/proceedings/papers/v8/re10a/re10a.pdf
52. G. Li and W. Liu, "Multimedia Data Processing Technology and Application," *Adv. Multimed.*, pp. 1–15, 2023.
53. Y. Wen *et al.*, "Deep Learning-Based Multiomics Data Integration Methods for Biomedical Application," *Adv. Intell. Syst.*, vol. 5, pp. 1–15, 2023, doi: 10.1002/aisy.202200247.
54. C. Beecks, A. M. Ivanescu, S. Kirchhoff, and T. Seidl, "Modeling multimedia contents through probabilistic feature signatures," in *MM'11 - Proceedings of the 2011 ACM Multimedia Conference and Co-Located Workshops*, 2011, pp. 1433–1436. doi: 10.1145/2072298.2072033.
55. T. Westerveld, A. Van Ballegooij, and F. De Jong, "A Probabilistic Multimedia Retrieval Model and Its Evaluation," *EURASIP J. Appl. Signal Process. 20032*, vol. 2, pp. 186–198, 2003.
56. A. Vijayarani, S. & Sakila, "Multimedia Mining Research-An Overview," *Int. J. Comput. Graph. Animat.*, vol. 5, no. 1, pp. 1–33, 2015.
57. D. R. Kawade and K. S. Oza, "News classification: A data mining approach," *Indian J. Sci. Technol.*, vol. 9, no. 46, pp. 1–7, 2016, doi: 10.17485/ijst/2016/v9i46/84444.
58. L. Li, "Data News Dissemination Strategy for Decision Making Using New Media Platform," *Soft Comput.*, pp. 10677–10685, 2022, doi: 10.1007/s00500-022-06819-0.
59. C. A. Bhatt and M. S. Kankanhalli, "Multimedia Data Mining: State of the Art and Challenges," *Multimed. Tools Appl.*, vol. 51, no. 1, pp. 35–76, 2011, doi: 10.1007/s11042-010-0645-5.
60. A. Veglis, T. Saridou, K. Panagiotidis, C. Karypidou, and E. Kotenidis, "Applications of Big Data in Media Organizations," *Soc. Sci.*, vol. 11, pp. 1–13, 2022.
61. M. Shao, "Measurement and Trend Analysis of New Media Coverage Topics Based on Comment Big Data Mining," *Math. Probl. Eng.*, p. 18, 2022.

62. M. Ehler, "Principal component model of multispectral data for near real-time skin chromophore mapping," *J. Biomed. Opt.*, vol. 15, no. 4, p. 046007, 2010, doi: 10.1117/1.3463010.
63. B. Ahamed and T. Ramkumar, "Data Integration - Challenges, Techniques and Future Directions: A Comprehensive Study," *Indian J. Sci. Technol.*, vol. 9, no. 44, pp. 1–9, 2016, doi: 10.17485/ijst/2016/v9i44/105314.
64. A. Kadadi, R. Agrawal, C. Nyamful, and R. Atiq, "Challenges of Data Integration and Interoperability in Big Data," in *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, 2014, pp. 38–40. doi: 10.1109/BigData.2014.7004486.
65. D. B. Searls, "Data integration: Challenges for drug discovery," *Nat. Rev. Drug Discov.*, vol. 4, no. 1, pp. 45–58, 2005, doi: 10.1038/nrd1608.

How to cite this article: Ayodeji Abimbola Owonipa, Taye Oladele Aro, Oyenike Adunni Olukiran, Olakunle Isaac Ifawoye, Aishat Oladayo Jimoh-Mahmud, Temitope Ayanladun Oyelakun. Multimedia data mining and processing for news source attribution. *International Journal of Research and Review*. 2024; 11(5): 48-59. DOI: *https://doi.org/10.52403/ijrr.20240507*

******