Big Data Analytics for Fraud Detection in Social Media Data

Pravallika Mummadi

Northwest Missouri State University, Maryville MO 64468, USA
S555592@nwmissouri.edu and mummadipravallika@gmail.com

Keywords: Big Data · Data Analytics · Social Media · Data-driven Decision-Making · Data Science · Text Analytics · Anamoly · Machine Learning · Text Mining · SVM(Support Vector Machines) · Naive Bayes

Abstract. Social media platforms have developed into hubs for a variety of fraudulent operations in the digital age, which presents serious obstacles to consumer trust and data protection. In order to detect fraud in social media data, this research study presents a revolutionary methodology that combines Support Vector Machines (SVM) and Naive Bayes algorithms with the capability of big data analytics. By combining these two powerful machine learning techniques, our method delivers a multi-faceted analysis in contrast to conventional rule-based or single-algorithm approaches, offering a comprehensive and flexible answer to the intricate nature of social media fraud.

In order to provide scalability and real-time analysis, the technique is built to process enormous volumes of heterogeneous social media data, including text, photos, and videos. SVM is used because it is good at categorizing high-dimensional data, which is why structured social media content is a perfect fit for it. In addition, Naive Bayes improves overall predictive accuracy by effectively managing unstructured data via its probabilistic methodology. This combination dynamically adjusts to changing fraud strategies and dramatically lowers false positives, a typical problem in fraud detection systems.

Our strategy represents a major improvement over current approaches, which frequently lack the adaptability and all-encompassing analytic capabilities needed for the ever-changing social media landscape.Through a series of tests and comparisons with conventional fraud detection techniques, the study highlights the superiority of this methodology in terms of accuracy, adaptability, and real-time processing. This study advances the fields of social media analytics and cybersecurity by offering a reliable, scalable, and effective method of thwarting fraud in the dynamic social media environment

1 Introduction

Recent years have seen the rapid growth of social media bring in both opportunities and difficulties. Numerous people are drawn to various social media

sites, which in turn sparks a tremendous quantity of activity. The fast spread of social media sites like Twitter, Facebook, and Instagram has transformed how information is shared and consumed in the digital age. These platforms, which were initially intended to promote worldwide communication, have unintentionally turned into hubs for the proliferation of false information and phony accounts. The severity of this problem is made worse by the potential for misinformation to alter public opinion, affect election results, and even spark civil upheaval during globally significant events. The threat of false information is not new. The financial sector has always been plagued with fraud and Major scandals have rocked economies . New norms, laws, and regulations have been enacted in reaction to these incidents. Can these steps stop fraud involving financial statements in the future? This challenge is made harder by the digital environment. The rise of big data, analytics, and data science has transformed business and academic ecosystems, making them the trending topics of today . These developments offer beneficial options, but they also bring about additional challenges. For instance, platforms have incorporated identity verification mechanisms because of the growing reliance on social networks for information. The effectiveness of such actions in halting the spread of false information remains to be examined. Furthermore, novel methods for fraud detection are required due to the sizeable and dynamic nature of the data on OSNs. The size and complexity of modern data are too great for conventional procedures, such manual verification and simple content analysis. With its capacity to analyze large datasets and spot complex patterns, machine learning (ML)[9] and text analytics[2] stand out as potential game-changers in the fight against false information. Recent research has consistently discussed the application of text analytics and machine learning in the identification of fraud. Deep Learning algorithms [5] and Naive Bayes Classifiers[6] were used to identify fake news and demonstrated promising accuracy of detection. Analytical methods have discovered fake reviews on social media, where the content and related metadata, like user ratings, are both changed to serve fraudulent ends. The proliferation of fake news on social networks, while not a new issue, has seen its reach amplify thanks to websites like Facebook. This challenge is comparable to that problem. Particularly during crucial occasions like the 2016 U.S. presidential election, such "altered truths" have had a substantial social influence[7]. The detection of such fraudulent activity has been done via data mining, which looks for patterns in huge databases.

1.1 Big Data Analytics

The industry of data science, business analytics, and information systems has seen a fundamental transformation because of the digital revolution. Analytical methods have advanced greatly since the 18th century, notably with the introduction of huge data and improvements in artificial intelligence. Rapid economic and social transaction digitalization has resulted in the collection of enormous amounts of organized and unstructured data. A startling 95 percent of big data is unstructured data, which includes written documents, movies, and sensor data. This increase in data has changed the nature of research as well as increased

the dimensionality of datasets. Historically, data gathering was done to test theories that people had come up with. However, in the modern day, which is characterized by vast amounts of data, computers are seen as more than just analytical tools; they are also seen as active players in research, capable of posing questions and coming to conclusions on their own . This evolution gives robots the ability to forecast and test theories that they may not have been aware of. Big data is distinguished by its quantity, speed, variety, authenticity, and worth. The market value of the business increased dramatically from 6.8 billion to 32 billion in just three years, indicating a potential market worth of 48.6 billion by 2019. This clearly demonstrates the industry's economic potential. Businesses now aggressively invest in data as a result of this rise, whereas some allocate over 10percent of their IT budgets exclusively for this use. Big data and analytics provide a strategic edge by assisting in decision-making, increasing company operations, and even enabling large-scale social phenomenon experiments. Big data, however, comes with a variety of difficulties. The data, which is frequently unstructured and comes from several networks, necessitates the creation of complex algorithms. Natural language processing has advanced with the rise of text and image processing tools like IBM's Watson, but there is still a long way to go until massive data on the internet can be fully utilized. The fact that many big data projects fall short of expectations also raises questions about the sustainability of the big data boom and forces a reconsideration of its true worth and return on investment. Researchers in Information Systems (IS) have a fantastic opportunity nowadays. IS researchers are well positioned to take advantage of the opportunities afforded by big data because of their long history at the nexus of technology and data. Their multidisciplinary approach demonstrates the adaptability and value of IS research in the big data era, tapping on disciplines like economics, sociology, and psychology. In a nutshell the rise of big data has fundamentally changed research by presenting unmatched chances for enquiring into and learning about. Although there are many obstacles to overcome, the potential for ground-breaking discoveries and new knowledge is enormous thanks to developments in algorithms, artificial intelligence, and interdisciplinary techniques like IS. .

1.2 Social Media Data, Disinformation and Fake News

A significant difficulty in the digital age is the proliferation of fake news, which was especially noticeable during the 2016 U.S. presidential election . The distribution of publications that are purposefully false has been accelerated by social media platforms, particularly Facebook . Users on these platforms tend to be less critical because they frequently approach content with a hedonistic perspective, which encourages the acceptance and dissemination of false information . Determining real news from fake news is made more difficult by the democratization of journalism on social media and users' passive selection of news sources [7] . The phenomenon of confirmation bias, where users are more likely to trust and spread information that supports their preexisting ideas, is a noteworthy effect of this environment [8]. There is a need for user education, the introduction of

efficient news source rating systems, and adjustments to the design of social media platforms to solve these issues [7]. The reliability of news sources has been evaluated using a variety of approaches, including user and expert evaluations[8]. In the context of sharing and consuming information, the digital age has created a wide range of benefits and constraints. The rise of "fake news" and financial fraud, both of which have serious repercussions for people, businesses, and societies at large [9] are a major cause for concern. In the past, established financial institutions and traditional media channels were responsible for disseminating news and financial information. Although nearly anybody can now produce, disseminate, and consume information, including anything that might be incorrect, misleading, or deceptive , the rise of social media has democratized this process. Despite conventional techniques of preventing fraud in corporations have their roots in standard accounting principles, there is an increasing understanding of their shortcomings, particularly considering the manipulated nature of organized financial data [9]. It is well-established that internet deceptions provide problems, particularly in e-commerce and financial misinformation. Online deceptions, such as falsified product reviews, biased recommendation engines, and manipulated product information, can have a big impact on consumer decision-making and company reputations [10]. Additionally, it has been noted that the anonymity of the internet significantly facilitates the dissemination of false information, including financial misinformation, by lowering the perceived risks involved in publishing false information. There has been an increase in interest in extraction unstructured data, such as textual content from financial statements and user-generated content on social media platforms, for more efficient fraud detection and decision-making due to advancements in natural language processing and the rise of financial social media platforms[9] . Social media data, which is unstructured, qualitative, and subjective in nature, offers a plethora of information that, when properly tapped into, can offer businesses priceless insights, particularly in innovative product development . Researchers have suggested content analysis, statistical cluster analysis, and systematic coding methods to overcome the difficulties associated with obtaining value from social media data. These techniques, which have theoretical as well as managerial ramifications, seek to transform unstructured social media data into insights that can be used by decision-makers . In the final analysis, the landscape of information consumption and dissemination is quickly changing, with conventional approaches falling short of meeting the demands of the digital age. A promising path to more effective fraud detection, decision-making, and product development is the merging of structured financial data with unstructured social media data.

1.3 Text Analytic Framework

The rapid digital transformation has brought about significant advancements in various domains, from social media analytics to cybersecurity. Text analytics, in particular, has emerged as a powerful tool for extracting insights from vast amounts of unstructured data, such as social media comments. This process involves gathering, preprocessing, and extracting features from the text, followed

by advanced modeling and analytics [2]. The rise of the internet and social media has spurred interest in analyzing such data, with features extracted from text being categorized into lexical, stylistic, social, sentiment, distinctive terms, product, and semantic features [2]. However, while these technologies offer numerous benefits, including faster access to knowledge and reduced operational costs, they also present challenges in sense-making, a critical precursor to informed decision-making [3].

Historically, quality management research has primarily focused on Process Management and Business Results, with limited exploration of the connection between Measurement, Analysis, and Knowledge Management [2]. Traditional tools for Total Quality Management have been restricted to structured information, such as cause and effect diagrams and control charts. The analysis of unstructured content, especially from the internet, remains an underexplored area [2]. In the realm of cyber threat detection, there has been a notable shift towards proactive intelligence gathering, especially from sources like the Dark Web [1]. Platforms within the Dark Web, such as forums, are rich sources of information about hacker tools, strategies, and potential threats. However, the sheer volume of data and the unique language used by hackers pose challenges for conventional cybersecurity techniques [1]. The potential of artificial intelligence (AI) and advanced computational methods in addressing these challenges is evident, yet there remains a gap in its application, especially in synthesizing intelligence from the Dark Web [1]. Furthermore, while the potential of the Dark Web as a source of Cyber Threat Intelligence (CTI) is undeniable, it also raises ethical and legal considerations, especially concerning privacy and surveillance [1]. The act of accessing and using data from the Dark Web might be fraught with ethical dilemmas. In conclusion, while the digital landscape offers numerous opportunities for advancements in text analytics and cybersecurity, it also presents challenges that need to be addressed. The proactive harnessing of intelligence from sources like the Dark Web, complemented by the power of AI, presents a promising strategy. However, realizing this potential requires a multi-faceted approach, encompassing technological advancements, ethical considerations, and a deep understanding of the digital ecosystem.

1.4 Fraud Detection and Prevention

Financial fraud detection and anti-money laundering are dynamic and developing fields that are constantly exploring new approaches and using cutting-edge technologies. In recent times, it has been normal practice to rely on financial statement variables identified by experts, frequently in the form of financial ratios. However, a paradigm change in favor of the decomposition of classic fraud predictors into raw financial data has been seen. This change, when combined with sophisticated data-mining methods, presents the possibility of more accurate fraud detection. The type of data that can be used for fraud detection analysis has also attracted a lot of attention. While some research has concentrated on using internal data sources, the majority have seen the drawbacks of this strategy and have moved their attention to using external, or publicly

available, data [4]. A common strategy has been to combine data from public financial accounts with statistical and machine learning-based classification techniques. However, the effectiveness of data based on financial statements for fraud detection has come under scrutiny, leading to the investigation of creative solutions. The concept of agency is a fundamental framework for comprehending the tensions that exist in contemporary businesses, emphasizing the possibility of fraud and the importance of audits in maintaining openness and accountability . The investigation of data-driven audit methodologies has become necessary due to the growing complexity of corporate processes and the enormous amounts of data created. Techniques for finding anomalies, or patterns that deviate from expectations, have been used in a variety of fields, including accounting and auditing. Data analytics integration in the auditing field has become a potent tool for improving audit effectiveness and efficiency as well as for detecting fraud. Multiple research efforts have been conducted in the specific context of AML, utilizing a wide range of methodologies including literature reviews, qualitative, quantitative, and mixed methods. Considering how corporate operations and fraud schemes are changing, the literature emphasizes the significance of developing audit and fraud detection tools. The agency theory and other theories that support classification-based models, ensemble models, and data-driven audit methodologies provide potential directions for improving the precision and effectiveness of audits.

1.5 A Comprehensive Approach to Fraud Detection in Social Media

The proposed strategy represents a major improvement over current techniques for identifying fraud in social media data by utilizing big data analytics and, in particular, Support Vector Machines (SVM) and Naive Bayes algorithms. Even if they are simple, traditional methods like rule-based systems are frequently rigid and cannot adjust to new and changing fraud patterns and variuos methods can be seen in table 1 . Their efficacy is usually limited to recognized fraud scenarios and they are impeded by their dependence on predetermined guidelines. Conversely, traditional machine learning methods, such Logistic Regression or Decision Trees, are more flexible than rule-based systems, but they have drawbacks including requiring labeled data and requiring human feature engineering. Despite being more adaptable than rule-based systems, these techniques are unable to fully handle the intricate and dynamic nature of social media fraud. While anomaly detection systems—which find anomalies or peculiar patterns—are capable of identifying fraud kinds that are not yet identified, they frequently have greater false positive rates. While these algorithms are good at spotting anomalies, they might not always be able to tell the difference between fraud and real abnormalities. By combining SVM and Naive Bayes with big data analytics, our method, on the other hand, offers a more sophisticated and practical answer. Structured vector machines (SVM) are widely used to handle high-dimensional data and are especially good at classifying complex datasets. For structured social media data, this makes SVM a good fit. Because of its probabilistic methodology, Naive Bayes is excellent in predicting outcomes

Table 1. Summary of Research Landscape in Detecting the fake data

Methods	SubMethods	Description	Source	Results	Research Gaps	Journal
Machine Learning and Deep Learning	Machine Learning	The Methods delves into the basics of ML algorithms, the theory behind it, and its applications. It also explores the concept of fake news	Sources are a pre-defined fake news dataset	Identifies that when coupled with TF-IDF for feature extraction, provides a promising approach to detect fake news with decent accuracy	The research points out the challenges in fake news detection and the need for larger, more diverse datasets.	[6,10]
	Deep Learning Algorithms	Liar Dataset, ISOT Fake News Dataset	The theory behind both traditional machine learning algorithms and deep learning neural networks. It also delves into the concept of fake news and its characteristics.	Deep learning models (especially LSTM) achieved better performance in fake news detection compared to traditional machine learning algorithms	need for more sophisticated models and larger datasets for improved performance	[8,5]
Text analytics	Text Analytics and ML algorithms	Hacker forum data, vulnerability assessment data	Development of proactive CTI using online hacker community data	Identification of system vulnerabilities and potential threats	The need for real-time detection mechanisms	[3,2]
	Text Mining	Financial news articles, propagation data	Detection of financial disinformation on social media platforms	aims to build a system rooted in the truth-default theory to detect deceptive communication	Effective detection of financial disinformation	[1]
Meta Analysis	Meta Framework	Sources used are public databases and wed pages data	The method aims to ensure	This method highlights the growth of big data research, and uncovers the most topics and trends in the domain	The study identifies gaps such as the lack of empirical research in the area and the need for more interdisciplinary research in big data	[4]
Big Data Analytics with ML techniques	BDA Framework	Analyzing financial markets for potential threats and opportunities	Financial datasets	Identifying market threats and opportunities using big data	Current auditing practices primarily focus on structured financial data, which might not capture the entire picture.	[9]

based on the possibility of occurrences, contributing a further level of analysis that is particularly helpful in situations involving unstructured data. This combination, together with big data analytics' capabilities, makes it possible to

handle the enormous amounts of data that are typical of social media, giving these platforms the scalability and real-time processing that are necessary for their fast-paced environment. In addition to being flexible enough to accommodate novel and intricate fraud patterns, the approach uses both behavioral and textual data to provide a thorough analysis. This improves the accuracy of the detection process by significantly lowering false positives, a common problem in fraud detection. Furthermore, this method's adaptability allows it to be used effectively with a wide range of social media content, including text postings, photographs, and videos, and provides a comprehensive understanding of user behavior and interactions. Compared to approaches that mainly concentrate on one kind of data, like text, this represents a significant advancement.

In conclusion, the flexibility, thorough analysis capabilities, scalability, real-time processing, and lower false positive rates make the suggested methodology exceptional. These characteristics make it especially appropriate for the ever-changing and intricate social media landscape, when more conventional approaches might not be adequately

1.6 PROPOSED METHODOLOGY

With an emphasis on Support Vector Machines (SVM) and Naive Bayes algorithms, a complete strategy is taken to construct a model for detecting fraud in social media data using big data analytics. This technique integrates several stages from data collection to model deployment which can be seen in the block diagram 1.

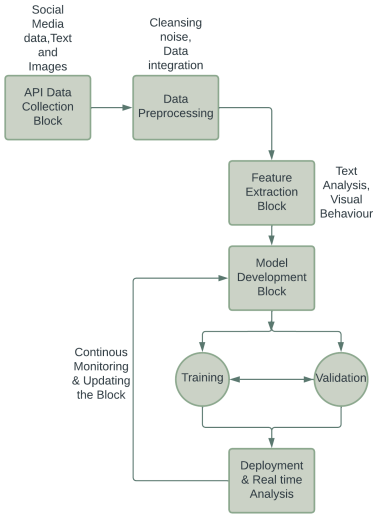

Fig. 1. Block Diagram for proposed methodology

A wide range of data is first gathered from social media networks to start the process. Text posts, photos, videos, and analytics like likes, shares, and comments are all included in this data. Web scraping tools and platform-specific APIs are used to collect this data. Web scraping is used to supplement the organized and trustworthy data that APIs offer with extra information that APIs do not provide.

The data is preprocessed after it is gathered. Ensuring the quality and consistency of the data requires careful attention to this stage. Preprocessing is the process of cleansing the data by eliminating superfluous information, fixing mistakes, and managing missing values. It also comprises data integration, which combines data from multiple sources into a single dataset, and normalization, which standardizes data from various platforms into a common format.

The crucial next step in converting unprocessed data into a format appropriate for machine learning models is feature extraction. Natural Language Processing (NLP) techniques including tokenization, stemming, and lemmatization are used for textual data. The emotional tone of the posts is deciphered using sentiment analysis, and popular themes are found through topic modeling, especially when algorithms such as Latent Dirichlet Allocation (LDA) are applied. Images and movies are analyzed for visual content using image processing methods and deep learning models such as Convolutional Neural Networks (CNNs). Additionally, network analysis is used to examine social relationships and interactions, and user interaction patterns and engagement metrics are examined to spot anomalies. This is known as behavioral analytics.

After the features are recovered, the emphasis switches to developing the model—more especially, utilizing the SVM and Naive Bayes techniques. SVM is used because it can represent intricate relationships in text and structured data and works well in high-dimensional environments. Because of its ease of use and effectiveness when working with sizable datasets, Naive Bayes is chosen for probabilistic predictions based on attributes that are taken from social media data.

These models' validation and training are essential. Finding the ideal hyperplane to divide the data's classes into distinct groups is the aim of SVM training. The procedure for Naive Bayes entails figuring out the odds that various traits will belong to particular classes. Next, the models are assessed using measures such as F1-score, recall, accuracy, and precision.

Lastly, the models are used to detect fraud in real time on social media sites. Creating APIs to incorporate the models into the platforms is part of this. Additionally, a mechanism for ongoing observation and updating is put into place. This method makes sure that the models are updated frequently to reflect changing social media trends and deceptive techniques, and they are always learning from fresh data and user comments.

To summarise, this methodology encompasses multiple phases of data management, ranging from gathering to preprocessing, feature extraction, creating models, training, authentication, and implementation. Combining SVM with Naive Bayes algorithms offers a well-rounded method that makes effective use of

each algorithm's advantages in managing intricate, high-dimensional social media data and producing probabilistic predictions. This all-encompassing strategy is necessary to identify and evaluate fraudulent activity in an efficient manner.

2 Case Study: Detecting Fake News Using SVM and Naïve Bayes Algorithms

2.1 Data Collection

The spread of false information in the digital age has become a serious problem that influences public perception and confidence. In order to tackle this problem, a study was carried out to determine how machine learning algorithms differentiate between authentic and fraudulent news.

The study's dataset was painstakingly gathered from a variety of social media sites. The main goal was to compile a wide variety of content, such as posts, comments, photos, and videos, that represents the complexity of social media interactions. The machine learning algorithms (Support Vector Machine and Naïve Bayes) that properly recognize and categorize authentic and fraudulent news depended on this data for training and testing.

2.2 Algorithms Used

The Support Vector Machine (SVM) and Naïve Bayes Classifiers algorithms were the main subjects of the study.

Support Vector Machine (SVM): In order to classify data points, the cut plane in an N-dimensional space had to be found. When utilizing a linear kernel function to handle multidimensional data points, the technique performed especially well. Naïve Bayes Classifiers: Based on the assumption of conditional independence of attributes, this algorithm utilized Bayes' rule. In comparison to other classification algorithms, it was renowned for its computational efficiency and quicker processing time. The dataset from the news stories was first entered into a DataFrame, which dispersed the data into a two-dimensional structure to facilitate handling. After that, stop words were extracted from the articles' text features so that the more important information may be highlighted. In order to identify stop words and provide predictions based on word usage, TF-IDF Vectorizer was utilized. A confusion matrix was used to compare the effectiveness of the SVM and Naïve Bayes algorithms for classification.

Metrics for Performance Evaluation The study employed a number of criteria, such as accuracy, precision, recall, and F1-score, to assess how well the algorithms performed on the dataset. The confusion matrix, which divided the predictions into true positives, false negatives, false positives, and true negatives, served as the foundation for these measurements.

The SVM and NB algorithm showed a level of accuracy in predicting real and fake news, with the following performance metrics can be seen in 2 :

The study showed how well machine learning algorithms—more especially, SVM and Naïve Bayes—work for identifying false news. Despite SVM's almost

flawless accuracy, it was pointed out that applying a single strategy might not always work in practical situations. But this method offered insightful information about how machine learning might be used to detect false information.

3 Conclusion

The research presented here explores the rapidly developing topic of big data analytics with a particular emphasis on how it might be used to identify and examine fraudulent activity on social media platforms. Our unique approach, which combines Support Vector Machines (SVM) and Naive Bayes algorithms, is notable for its precision, versatility, and capacity to handle massive amounts of heterogeneous data in real time. Our approach addresses the dynamic and complex nature of social media fraud, offering a significant improvement over traditional rule-based and single-algorithm methods.

By combining the benefits of SVM and Naive Bayes algorithms, our approach is resistant to the constantly changing strategies used by social media scammers. Because of its probabilistic approach, the SVM algorithm's skill with high-dimensional data complements the Naive Bayes algorithm's efficacy with unstructured data. This synergy not only improves fraud detection accuracy but also dramatically lowers false positives, a prevalent issue in this field. Our study emphasizes how important big data analytics is to improving social media platform cybersecurity safeguards. Our established methodology offers a complete framework that enables researchers and businesses to better analyze and combat social media fraud.

4 Future Work

There is a wealth of prospects for future study in the field of big data analytics for social media insights, especially in the area of fraud detection. Predictive models could become even more precise and effective by incorporating cutting-edge machine learning techniques, like deep learning, into the current processes. Extensive testing on a range of social media platforms and real-world implementation are necessary to confirm the efficacy of the suggested approaches in a variety of dynamic settings. It is imperative to investigate the degree to which these approaches can be adapted to other platforms, considering the distinct user behaviors and attributes that are specific to individual social media platforms. Data privacy and ethical issues are becoming more and more crucial to handle

Table 2. SVM and Naïve Bayes Performance Results

Algorithm	Accuracy	Precision	F1-score
SVM	96	97	96
NB	89	95	90

as big data analytics develops. Future research should concentrate on creating techniques that strike a balance between the protection of user privacy rights.

Future study must also focus on adapting to the ever-changing strategies used by con artists. Keeping fraud detection systems successful will require developing algorithms that can adapt dynamically to new fraudulent techniques. Multidisciplinary insights from the fields of criminology, sociology, and psychology may be combined to create a more comprehensive understanding of the mechanics and motivations underlying social media fraud. Considering social media's global reach, localizing fraud detection techniques should be a focus of future research. It is imperative to include language, cultural, and regional disparities in social media usage patterns to guarantee the efficacy of fraud detection in diverse international settings. In general, there is much potential for the area to grow significantly, improving both the security of social media platforms and our comprehension of the complex dynamics at play in digital communication environments.

References

1. Abbasi, A., . C. H. Cybergate: A design framework and system for text analysis of computer-mediated communication. *MIS Quarterly 32*, 4 (2008), 811–837.
2. Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., and Jiao, J. An integrated text analytic framework for product defect discovery. *Production and Operations Management 23*, 11 (2014), 1771–1788.
3. Ahmad Abbasi, Yili Zhou, S. D., and Zhang, P. Text analytics to support sense-making in social media: A language-action perspective. *MIS Quarterly 42*, 2 (2018), 427–464.
4. Ahmed Abbasi, Conan Albrecht, A. V., and Jam. Metafraud: A meta-learning framework for detecting financial fraud. *MIS Quarterly 36*, 4 (2012), 1293–1327.
5. Al-Khalifa, M., and Al-Kabi, M. Fake news detection using machine learning and deep learning algorithms. *International Journal of Engineering and Advanced Technology (IJEAT) 9*, 1 (2020), 677–684.
6. Granik, M., and Mesyura, V. Fake news detection using naive bayes classifier, 2017.
7. Kondamudi, M. R., S. S. R. C. L. . Y. N. A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches. *Journal of King Saud University – Computer and Information Sciences 35* (2023).
8. SAEEDREZA SHEHNEPOOR, ROBERTO TOGNERI, W. L., and BENNAMOUN, M. Social fraud detection review: Methods, challenges and analysis. *The University of Western Australia* (2021).
9. Wei Dong, S. L. Z. Z. Leveraging financial social media data for corporate fraud detection. *Journal of Management Information Systems 35*, 2 (2018), 461–487.
10. Xiaohui Zhang, Qianzhou Du, Z. Z. A theory-driven machine learning system for financial disinformation detection. *Production and Operations Management 31*, 8 (2022), 3160 – 3179.