(IJISE) 2019, Vol. No. 9, Jan-Jun **e-ISSN: 2454-6402, p-ISSN: 2454-812X**

LEVERAGING THE MACHINE LEARNING ALGORITHMS AND TOOLS TO ENHANCE BUSINESS INTELLIGENCE LINKED TO SENTIMENT ANALYSIS

Shweta

Symbiosis College of Arts and Commerce, Pune

ABSTRACT

This paper proposes the utilization of sentiment analysis classification as a valuable approach for analyzing textual data sourced from various internet resources. Sentiment analysis, a data mining method employing machine learning techniques, offers insights into the opinions, reviews, feedback, and suggestions available online. Given the vast array of user opinions, it is imperative to uncover, analyze, and synthesize these viewpoints for informed decision-making. Sentiment analysis provides real-time, efficient feedback from consumers, significantly impacting the decision-making process in the business domain. Over the past decade, there has been a notable increase in research activity and emphasis on exploratory research methodologies. However, we observe certain gaps in Business Intelligence research methodologies, as well as identify areas that warrant further investigation.

INTRODUCTION

In today's world, people engage in various online forums and social media platforms like Twitter and Facebook for social interaction. This collective engagement through digital media occurs in real-time, offering many opportunities for business intelligence [14]. Individuals from diverse backgrounds, including countries, genders, classes, and races, share their experiences and opinions via the Internet. The rise of social media, the widespread use of mobile devices, and the advent of the Internet of Things contribute to the generation of vast amounts of real-time data. More than 1200 Exabytes of new data are generated each year from various sources [1]. Eighty per cent of this data is unstructured, making it challenging to store, process, and analyze using conventional tools [1, 2].

Analytics plays a crucial role in making sense of big data by transforming it into actionable insights [3]. Extracting value from big data involves two main steps: data management and analytics. Data management involves processes and technologies to acquire, store, and prepare data for analysis, while analytics refers to the technologies used to analyze and derive intelligence from big data [4]. Researchers have long been striving to enhance information systems that provide business intelligence. Business intelligence comprises both a product and a process. The product entails information that enables organizations to anticipate the behaviour of their competitors.

SENTIMENT ANALYSIS

A. What is Sentiment Analysis?

Sentiments encompass feelings that reflect attitudes, emotions, and opinions. Sentiment Analysis can be defined as the computational study of opinions, evaluations, attitudes, subjectivity, and views expressed in text. It involves assessing the emotional valuation of a situation, which can be either positive or negative based on physical or mental responses. The significance of sentiments in business is well-established, as customer responses serve as indirect motivators of purchasing behavior. Sentiment Analysis addresses these issues by systematically collecting and analyzing online sentiments from various sources and a large sample of customers in real-time. The phases of the sentiment analysis process are illustrated in Fig-1.

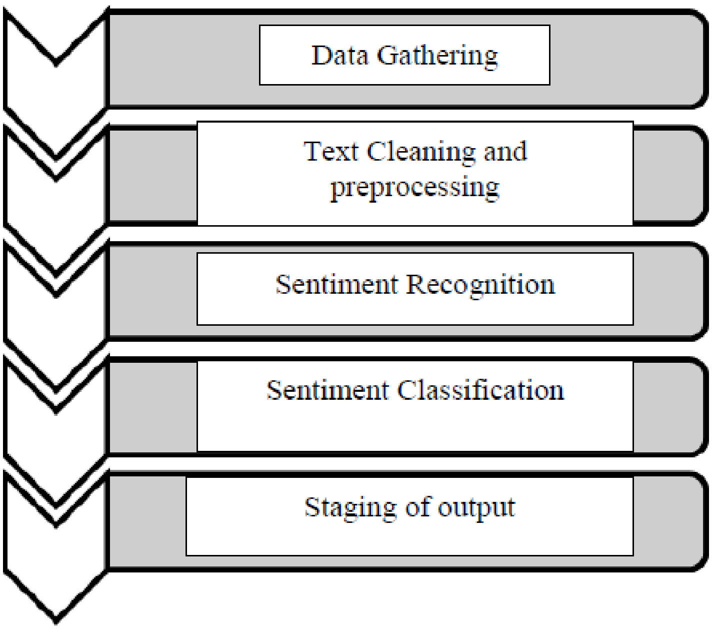

Fig-1 Sentiment Analysis Process

Sentiment analysis entails a systematic examination of online expressions, with a specific focus on gauging attitudes and opinions regarding a particular topic using machine learning techniques. The concept of sentiment analysis in data mining can be understood from two perspectives: functional and operational. Functionally, sentiment analysis is described as a process that categorizes a body of textual information to discern feelings, attitudes, and emotions toward a specific issue or object [6]. This definition elucidates how sentiment analysis operates and outlines the outcome of polar classification.

From an operational standpoint within the realm of computational linguistics, sentiment analysis is characterized as automated subjectivity analysis akin to opinion mining and sentiment extraction, which revolves around extracting and classifying texts using machine language and computer programming [7]. Despite the differences between these perspectives, the overarching narrative remains consistent. In essence, sentiment analysis is a data mining technique that leverages natural language processing, computational linguistics, and text analytics to identify and extract relevant content from a corpus of textual data.

B. Business Process Decision

A business process comprises interrelated, organized activities or tasks aimed at delivering specific services or products to fulfill the needs of particular customers or customer segments [18]. In today's rapidly evolving business landscape, organizations must adapt to meet changing customer needs and environmental demands. Consequently, there is a need for dynamic business process modeling and predictive intelligence to accommodate these changes and make informed decisions [19, 20].

Researchers have long endeavored to enhance information systems that provide business intelligence. Business intelligence encompasses both a process and a product. The process involves methodologies that organizations employ to cultivate valuable information or insights, enabling them to thrive and succeed in the global economy. The product refers to information that empowers organizations to anticipate the behaviors of competitors, suppliers, customers, technologies, acquisitions, markets, products and services, and the general business environment with a certain level of confidence [5].

MACHINE LEARNING APPROACHES

The objective of machine learning is to develop algorithms that optimize the performance of models using historical data as training data. In traditional approaches, we provide data (input) and a program to the computer (machine), which processes it and yields a result (output). However, in the case of the machine learning approach, we provide both data (input) and the desired result (output), and the computer (machine) deduces a program based on the input-output relationship [15]. The machine learning paradigm is depicted in Fig-2.

Based on this device's program (algorithm), a model is constructed and applied to unseen data to generate or predict results. Machine learning algorithms can be broadly classified into three categories: Supervised, Unsupervised, and Semi-supervised.

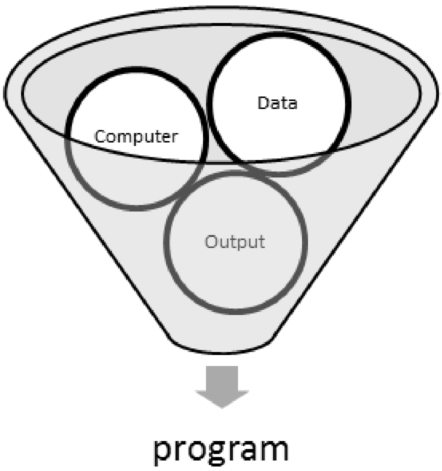

Fig-2 Machine Learning Paradigm

Supervised learning is inductive in nature, where the training data includes the desired output. It is commonly used in classification and regression tasks. In contrast, unsupervised learning involves training data that does not include the desired output and is often employed in clustering and pattern recognition [17].

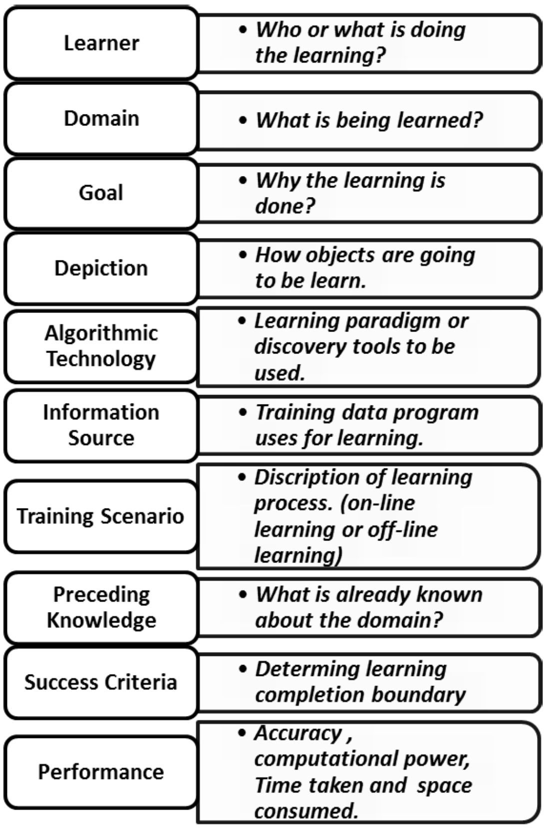

Fig-3 Computational Learning Model

A computational learning model should be transparent about several aspects, as depicted in Fig-3. The classification task itself comprises various subtasks, including:

- Data preprocessing, which involves cleaning the data.

- Feature selection and/or feature reduction, aiming to reduce dimensionality by decreasing the number of features.

- Representation, determining how objects will be learned.

- Classification, the actual categorization after completing the previous steps.

- Post-processing, which involves presenting the output.

The machine learning approach addresses classification problems through several steps, as illustrated in Fig-4:

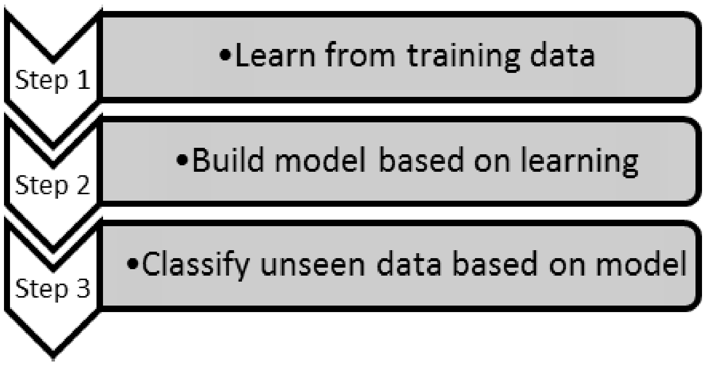

Fig-4 Machine Learning Classification steps

MACHINE LEARNING SOLUTIONS

Sentiment analysis of natural language texts represents a vast and burgeoning field that can be leveraged by businesses to make informed decisions regarding their services and products [21]. Sentiment analysis, or Opinion Mining, involves computationally analyzing opinions, sentiments, and subjectivity within text. It is a Natural Language Processing and Information Extraction task aimed at discerning the writer's sentiments expressed in positive or negative tones, queries, and appeals, by analyzing a large number of documents.

The initial stage in any data-driven approach to sentiment analysis involves transforming a portion of written data into a feature vector. In traditional Information Retrieval and Text Classification tasks, term frequency is considered essential. However, Pang and Lee [8] established that the presence of a term is more significant to sentiment analysis than its frequency. Hence, binary-valued feature vectors are employed, wherein entries indicate whether a term occurs (value 1) or not (value 0). Furthermore, it has been reported that unigrams outperform bigrams when classifying movie reviews by sentiment polarity, thereby shaping sentiment analysis research towards determining the semantic orientation of terms.

Determining the semantic orientation of words involves hypotheses such as those proposed by Hatzivassiloglou and McKeown [10], who suggest that adjectives separated by "and" have the same polarity, while those separated by "but" have opposite polarity. This information is utilized to group adjectives into two clusters, satisfying maximum constraints.

Sentiment classification is a burgeoning sub-discipline of text classification, focusing on the opinions expressed in documents rather than the topics they discuss. An essential aspect of opinion extraction from text is determining the orientation of "subjective" terms, i.e., discerning whether a term with opinionated content has a positive or negative connotation [11]. Esuli and Sebastiani proposed a novel method for defining the alignment of subjective terms, based on quantitative analysis of term glosses and the resulting term representations for semi-supervised term classification.

Sentiment classification encompasses several specific subtasks, including determining subjectivity, orientation, and the strength of orientation [11]. SENTIWORDNET, described by

Esuli and Sebastiani [12], is a lexical resource associating each WordNet synset with three numerical scores, representing objectivity, positivity, and negativity.

Traditionally, sentiment classification is approached as a binary classification task [8, 9]. Dave, Lawrence, and Pennock [9] utilize structured reviews for training and testing, employing appropriate features and scoring methods from information retrieval to determine whether reviews are positive or negative. These results perform comparably to traditional machine learning methods, which are then applied to identify and classify review sentences from the web, where classification is more challenging.

Various supervised or data-driven techniques for sentiment analysis include Naïve Bayes, Maximum Entropy, and Support Vector Machines. Pang and Lee [8] compared their performance on different features, such as considering only unigrams, bigrams, or combinations of both, incorporating parts of speech and position information, or focusing solely on adjectives. The results of this comparison are illustrated in Fig-5.

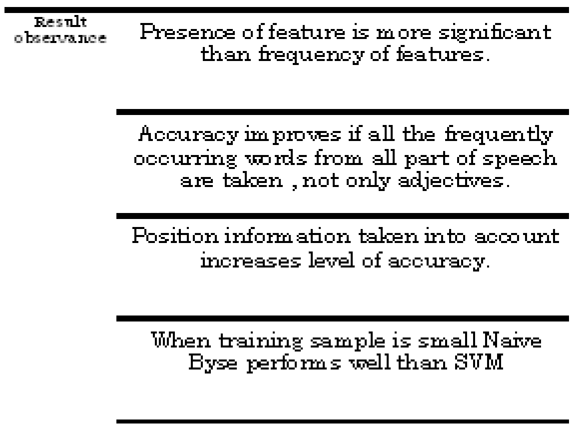

Fig-5 Comparison of performance of Naïve Bayes, Maximum Entropyand Support Vector Machines in Sentiment analysis on different features

CONCLUSION

In summary, this paper has thoroughly examined various machine learning techniques, including Naïve Bayes and Support Vector Machines, and has provided an up-to-date overview of the latest developments in sentiment analysis. It has demonstrated how sentiment analysis can be effectively utilized to inform business process decisions and has addressed the unique challenges presented by big data. Most importantly, the paper has served as a comprehensive introduction to sentiment analysis and its diverse applications in the business environment, leaving the reader with a solid understanding of this important field.

Sentiment analysis presents a significant challenge but potentially profoundly influences decision-making across domains. Despite the plethora of machine learning approaches available for application in various business domains, a generalized solution is still needed to determine the most suitable technique. Organizations often need a standardized methodology for business prediction to avoid trial and error with different approaches. Thus, there is a pressing need to delve deeper into developing standardized machine-learning approaches tailored to various business domains.

REFERENCES

[1] L. Issacs, “Rolling the Dice with Predictive Coding: Leveraging AnalyticsTechnology for Information Governance”, Information Management(47:1), 2013, pp. 22- 26.

[2] P. Malik, “Governing Big Data: Principles and Practices”, IBM Journal ofResearch and Development (57 :¾), 2013, pp. 1-13.

[3] R.M. Hogarth, E. Soyer, “Using Simulated Experience to Make Sense ofBig Data”, MIT Sloan Management Review (spring), 2015, pp. 5-10.

[4] A.Gandomi, and M. Haider, “Beyond the Hype: Big Data Concepts,Methods, and Analytics”, International Journal of InformationManagement (35), 2015, pp. 137-144.

[5] Vedder, R. G., Vanecek, M. T., Guynes, C. S., & Cappel, J. J. (1999).CEO and CIO Perspectives on Competitive Intelligence.Communications of the ACM, 42(8), 108–116.

[6] Liu, Bing. (2010). Sentiment analysis and subjectivity. Handbook ofNatural Language Processing, 2nd ed. Chapman and Hall: Florida.

[7] Kumar, Akshi, and Sebastian, Teja, Mary. (2012). “Sentiment analysis. Aperspective on its past present and future.” International Journal ofIntelligent Systems and Applications, 4 (10): 1-14.

[8] B. Pang, L. Lee, and S. Vaithyanathan,“Thumbs up?:Sentimentclassification using machine learning techniques,”inProc.ACL-02Conf.Empirical Methods Natural Lang. Process., 2002, pp. 79–86.

[9] K. Dave, S. Lawrence, and D. M. Pennock, “Mining the peanut gallery:opinion extraction and semantic classification of product reviews,” inProc. 12th Int. Conf. World Wide Web, New York: ACM, 2003, pp.519–528.

[10] V.Hatzivassiloglou and K. R. McKeown, “Predicting the semanticorientation of adjectives,” in Proc. 8th Conf. Eur. Chap. Assoc.Comput.Linguist., Morristown, NJ: Assoc. Comput. Linguist, 1997,pp.174–181

[11] A. Esuli and F. Sebastiani, “Determining the semantic orientation ofterms through gloss classification,” in Proc. 14th ACM Int. Conf. Inf.Knowl.Manage., 2005, pp. 617–624.

[12] A. Esuli and F. Sebastiani, “SENTIWORDNET: A publicly availablelexical resource for opinion mining,” in Proc. 5th Conf. Lang. Res.Eval., 2006, pp. 417–422

[13] International Journal of Ad hoc, Sensor & Ubiquitous Computing(IJASUC) Vol.4, No.1, February 2013, “Opinion Mining and SentimentAnalysis –An Assessment of Peoples’ Belief: A Survey”S Padmaja andProf. S Sameen Fatima.

[14] Hsinchun Chen, Roger H. L. Chiang(2012), BUSINESSINTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIGIMPACT, MIS Quarterly Vol. 36 No. 4, pp. 1165-1188/December 2012

[15] Ferguson, Rebecca (2012). Learning analytics: drivers, developments andchallenges. International Journal of Technology Enhanced Learning,4(5/6) pp. 304–317

[16] Vasant Dhar(2012), Data Science and Prediction, May2012,http://hdl.handle.net/2451/31553

[17] Chid Apte(2010), The Role of Machine Learning in BusinessOptimization, in Proceedings of the 27th International Conference onMachine Learning, Haifa, Israel, 2010.

[18] Chiang, R. H. L., Goes, P., and Stohr, E. A. 2012. Business intelligenceand analytics education, and program development: A uniqueopportunity for the information systems discipline. ACM Trans.Manage. Inf. Syst.3, 3, Article 12 (October 2012), 13 pages.

[19] Ranjit Bose (2008), Advanced Analytics, opportunities and challenges,industrial management & data systems ISSN:0263-5577

[20] Sharma, R., Mithas, S. and Kankanhalli, A. (2014). Transformingdecision-making processes: a research agenda for understanding theimpact of business analytics on organisations. European Journal ofInformation Systems, 23 (4), 433-441.

[21] Zack Jourdan, R. Kelly Rainer, and Thomas E. Marshall, BusinessIntelligence: An Analysis of the Literature, Information SystemsManagement, 25: 121–131Copyright b© Taylor & Francis Group, LLCISSN: 1058-0530 print/1934-8703

[22] Sonosy, O. A., Rady, S., Badr, N. L., & Hashem, M. (2016). A study ofspatial machine learning for business behavior prediction in locationbased social networks. 2016 11th International Conference on ComputerEngineering & Systems (ICCES). doi:10.1109/icces.2016.7822012.

[23] Singh B., Kushwaha N., & Vyas O. P. (2016). An interpretation ofsentiment analysis for enrichment of Business Intelligence. 2016 IEEERegion 10 Conference (TENCON). doi:10.1109/tencon.2016.7847950

[24] Chen J., Huang C., & Cheng C. (2016). The monitoring system ofBusiness support system with emergency prediction based on machinelearning approach. 2016 18th Asia-Pacific Network Operations andManagement Symposium(APNOMS).doi:10.1109/apnoms.2016.7737239

[25] Ghiassi, M., Zimbra, D., & Lee, S. (2016). Targeted Twitter Sentiment Analysis for Brands Using Supervised Feature Engineering and the Dynamic Architecture for Artificial Neural Networks. Journal of Management Information Systems, 33(4), 1034-1058.doi:10.1080/07421222.2016.1267526

[26] Cambria, E. (2016). Affective computing and sentiment Analysis. IEEE Intelligent Systems, 31(2), 102-107. doi:10.1109/mis.2016.31

[27] Kumar, A., & Joshi, A. (2017). Ontology driven Sentiment Analysis onSocial Web for Government Intelligence. Proceedings of the SpecialCollection on eGovernment Innovations in India - ICEGOV 17.doi:10.1145/3055219.3055229

[28] Swain, A. K., & Cao, R. Q. (2017). Using sentiment analysis to improvesupply chain intelligence. Information Systems Frontiers. doi:10.1007/s10796-017-9762-2

[29] V. Singh, R. Adhikari, and D. Mahata, "A clustering and opinion miningapproach to socio-political analysis of the blogosphere," inComputational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on, Dec 2010, pp. 1–4.