Ethical Behaviour with Avoiding Text Mining in Digital Education

Tamil Selvan
Alagappa University, India
https://orcid.org/0000-0001-8772-8181

G. Kalaiyarasan
Alagappa University, India
https://orcid.org/0000-0002-7155-3062

Abstract

Behavioural psychology is to connect our minds and activities. Our intentions and actions must be such that they do not affect others. Education is changing to digital education so we force to act with good behaviour. Nowadays, text mining is to be challenged for educationists. This paper addresses the ethical behaviour of avoiding text mining in digital education in the twenty-first century such as have explained how to act with ethical behaviour in others' text data in the current situation. This article advises following ethical behaviour in an educational text message, information, materials, etc. Further, this article explains what text mining is and how to use it with ethical behaviour. We will hope that text mining protection will get to everyone in future. Beginning with an overview of text data mining, this paper concludes with how to play with good behaviour.

Keywords: Ethical Behaviour, Text Mining, Data Mining, Digital Education.

Introduction

Information is collected from several sources such as newspapers, articles, books, digital libraries, e-mail messages, WhatsApp, text messages, web pages, etc. text databases are growing rapidly. As a researcher should follow good behavioural activities and should behave honestly because our society wants us to be discovering new things from us. So, we must follow respectable behavioural methods. We should follow better behaviours as educationists, without text mining write better articles as good researchers. The objective of text data mining is the creative display of others' data and to show the work of others' as self-work. According to text data standards, the kinds of data belong to the multimedia data, which include word forms, pdf files, excel page-aligned sequences of forms,and electronic or digital links. Text data is sequences of a stylus, data glove sensors, graphical, temporal, relational, and categorical. Text data mining is exploited from structured information and used in another way. Hence, can be analyzed words from documents, in addition, to be changed to our style. Therefore, we shouldn't spoil our noble behavioural image.

Text Mining

Text Mining is also known as Text Data Mining. Nowadays, every educational institution is being advised, to give learning process and educational information, through social media. For example, is made WhatsApp group education, SMS, e-mail, etc. most of the students are following that education with or without interest. But they are advised to follow that way because of the pandemic situations. Some of the learning material, education content, or

information is being sent to others. At the same time, is prepared the content, study material,and keynotes by a knowledgeable person. The data is used unethically by some friends or institutions. As many study centres are using other study centre materials in money minded. This is made negative thought midst in writers. So, they cannot get at a good position in a particular department. It is the foremost general medium for the proper exchange of information but now text mining is used to disrespect. Text is one of the issues in deriving implicit and explicit concepts (Bais, 2017). Natural Language Processing techniques are used to find the semantic relations between concepts. A large amount of text data is accounted for by the knowledge discovery (Oliveira-Neto et al., 2013), (Samha et al., 2015). In recent times, language analysis would be done by the computer is better than the human being. The manual techniques were expensive and time-consuming methods. There are various technologies are deployed in text mining. Text mining technologies are information extraction, summarisation, topic tracking, classification, and clustering (Usai et al., 2018). The term text mining is commonly used to denote any system that analyses large quantities of natural language text and linguistic usage patterns in an attempt to extract probably useful (although only probably correct) information (Natarajan et al., 2006)(L. Kumar & Bhatia, 2013). Reflecting a cultural therapeutic ethos, such labels blur popular and professional understandings of our emotional and psychological selves (Ecclestone, 2009) (Füredi et al., 2003). Studies in schools and colleges show how this leads teachers and support staff to categorise certain children and young people as, variously, 'vulnerable', 'at risk', having 'fragile learning identities', 'no self-esteem' and being 'dysfunctional' and 'disaffected and disengaged' (Ecclestone, 2009). Like that, we have to highlight our behaviour within this world and show it. Data mining is a process previously unknown by self-knowledge. With the extensive use of information technology and the recent developments in digital education systems, the amount of multimedia data available to users has increased exponentially. Text data is one of the multimedia data collections. Multimedia contains several kinds of data such as images, meta-data, visual audio and video. It is widely used in many major potential applications like coaching centres, study circles, communication, and educational programs. Text mining is one of the big problems of the data-mining to text creative community. When compared to the mining of other types of data, text data mining is still in its infancy. There are many challenges in digital education existing with text data mining. Well, behavioural activities will be got a good name among educationists.

Ethical Behavior

If our behavioural activities are fine, the behaviour of those who are also with us will be fine. Bible says, For with the measure you useit will be measured back to you, Luke: 6;38. Nowadays, every education newspaper says that educationists are facing text mining activity by unethical behaviour person. Plagiarism, copyright, publication ethics, ethics of writing, ethics of downloading, reuse ethics and open access ethics are followed by an ethical behaviour person. We can improve our good ethically activities when we can feel others' difficulties. At Virginia Tech, officials stated that cheating involving electronic media rose dramatically within one academic year, from 80 cases in 1995–1996 to 280 incidents in 1997(Zack, 1998). As reported by USA Today on May 21, 2001, at UC Berkeley, academic dishonesty cases doubled between 1995 and 1999 alone (Groark, 2001). A large 2000/2001 survey conducted by McCabe indicates that cheating is rampant in high schools as well. More than half of the high-school students have plagiarized writing assignments in some form specifically with the help of the Internet (CAI, 2002–2003).The ethical review might be accused us. It is the best way to ensure that people are acting with moral behaviour (McMillan, 2005).

Unethical Behavior
Information Mining

Text mining is the work of information mining. Text mining is information mining from unstructured information to structured information (Bais, 2017). The information mining work is to recognize phrases and finds the relation between them. Information extraction is a key that extracts meaningful information from a large amount of

text. Experts specify the attributes and relations according to the topic. Information extraction is used to extract specific attributes and concepts from the document and establish their concept relationship. The extracted concept is stored in a database for further processing. The information mining process is used to check, and evaluate the relevance of results on the extracted text. The modularity of documents may be adjusted so that each subsection or paragraph comprises a unit in its own right, in an attempt to focus results on individual nuggets of information rather than lengthy documents (Abrol & Malhotra, 2013) (Agrawal & Batra, 2013).

Cluster Mining

Clustering focus on similarity measures of different objects and places. A cluster is a set of collection work. For example, around our living place, there are many educational institutions and study centres. Educationists are given many study materials standardized for study to their students or learners. But unauthorized institutions have collected the mining of text from more standardized institutions' material and remade their text material ontheir own also is followed this method, and get more money. In-text mining, the clustering key can be applied to group similar words, sentences, and concepts. Cluster is processed by assigning documents to pre-defined categories and grouping documents into the natural cluster. A Cluster is a group of related documents, and Clustering is the operation of grouping documents based on some similarity measure, automatically without having to pre-specify categories(Gupta & Lehal, 2009). The most common Clustering algorithms are used hierarchical, binary relational, and fuzzy. Hierarchical clustering creates a tree with all documents in the root node and a single document in each leaf node. The intervening nodes have several documents and become more and more specialized as they get closer to the leaf nodes. It is very useful when we are exploring a new document collection and want to get an overview of the collection. The most important factor in a Clustering algorithm is the similarity measure. All Clustering algorithms are based on similarity measures(Gupta & Lehal, 2009).

Text Cleanup

Text Cleanup means removing unnecessary or unwanted information, such as removing content or text from web pages. Normally, Text is converted from binary formats. It deals with tables, figures and formulas(L. Kumar & Bhatia, 2013).

Classification

Classification is a key for text data mining analysis, can classify every specified set of documents. It is followed to achieve a new goal. Classification is the process of constructing data into categories for its better effective and efficient use. Well-planned data creates predefined classes. Nowadays, critical minded people are changing classification concepts without losing exactness (Chena & Chena, 2015). Text classification has broadly been used in different communities. Such as data mining, database, and machine learning are used in a massive number of applications in various domains.

Text Summarisation

Many text mining applications need to summarise the text file to get a concise overview ofa large file or a collection of files on a topic (FREEMAN & SISTRUNK, 1978),(E B & P Haroon, 2016).There are two categories of summarisation techniques in general extractive summarisation where a summary comprises information units extracted from the original text and contrary abstractive summarisation where a summary may contain "synthesized" information that may not occur in the original document (Allahyari et al., 2017).A text summariser produces a compressed representation of its input, which specifies Human Consumption. It also contains individual documents or groups of documents. Text Compression is a related area but the output of text summarisation is specific to be human-readable. It only supports decompression, that is, automatic reconstruction of the original text. Summarisation differs from many other forms of text mining in that there are people, who are skilled in the art of producing summaries and carrying out the task as part of their competitive life. Many institutions are gathered educational text material with legal or illegal from online or others' preparation and to use to develop their institutions.

Educational Data Mining

Educational Data Mining is concerned with developing methods for exploring the unique types of data that comes from educational environments (Bakhshinategh et al., 2018). It can also be defined as the application of data mining techniques to this specific type of dataset that comes from educational environments to address important educational questions (Romero & Ventura, 2013). Nowadays, there is a wide array of well-known general-purpose tools and frameworks that can be used to conduct educational data mining(Slater et al., 2017) such as Rapidminer, Weka, SPSS, Knime, Orange, Spark Lib, and so on. However, these tools aren't easy for educators to use due to they are required to select the specific method to use and to provide the appropriate parameters in advance to obtain good results. So, the educators must possess a certain amount of expertise to find the right settings(Romero & Ventura, 2013). However, they only work with specific data to solve specific educational problems. Most educational data mining normally uses their data for solving specific educational problems. But it is a complex and very time-consuming task to gather and pre process educational data (Slater et al., 2017). So, another option is to use some of the public datasets that are currently available for free download on the Internet.

Educational Text Mining Software

Text analysis software also called text analytics or text mining software helps users gain insights from both structured and unstructured text data using natural language processing. Such insights include sentiment analysis, key phrases, language, themes and patterns, and entities, among others. Text mining tools can process text parts of speech, sentence structure, and semantic word meaning. Additionally, some device scan identify representational relationships between different words and sentences. The tool's analysis reflects the diversity and complexity of the language that it seeks to measure and assess. Different groups of linguistics researchers often have different approaches to describing and analyzing text, and the wide range of tools available for text mining (M. Kumar et al., 2007) (Das et al., 2012). Text mining software leverages natural language processing and machine learning to complete other insights and provides visual representations of the data for more straightforward interpretation.

Open Access and Text Data Mining

Traditionally, the Copyright Transfer Agreement is followed to publish text material and research papers by the experts. Copyright transfer as the default has far-challenging consequences on the ability of both the experts and users to re-use that published research. Many authors did not know of the awareness and impact of these transfers on copyright. So, Academics frequently give the copyright to the publishers in exchange for the perceived prestige of publishing in one of their venues (Muller-Langer & Watt, 2010).In some cases, institutes adopt rights-retention O A policies that grant authors non-exclusive rights to their institutes before signing copyright agreements with publishers, which enables them to make articles O A without requiring permission from publishers (cyber. law.harvard. edu/hoap/Good practices for university open-access policies). Fundamentally, copyright is a pre-digital tool wielded by traditional publishers to maintain revenues rather than fostering creativity, innovation, or protecting authors (Willinsky, 2002). In this current climate of digital education, copying is essential to perform necessary activities ranging from viewing the article to reusing figures from downloading materials. This interaction of open access and copyright is complex and earns extended gaining knowledge (Eger et al., 2015).We will follow how OA views copyright and relate this to its effects on text data mining. Text data mining is not only a knowledge-generation tool. Text data mining allows for automated screening for errors and automated literature searches that renew scientific discovery. Text data mining has become possible to easily compare one's results with those of the published literature, identify convergence of suggestions and enable knowledge discovery (Natarajan et al., 2006). Text data mining has discovered frequent tentative hypotheses that can be used for new research (Abrol & Malhotra, 2013). It has already been used to make major advances in fields such as biomedicine (Gonzalez et al., 2016). It also allows for computer applications that can download all academic literature given certain search terms, simplifying

and shortening the tedious literature search. Well, behavioural activities create new common-sense assumptions about the state of the self.

Challenging

Twenty-first-century education is very challenging for educationists,learners, and parents. Recently education is followed by digital or online mode because,during this pandemic period,students did not move to schools, colleges, institutions, learning centres, and universities. So, all educational lessons, content, information, and materials are uploaded on a website with an internet connection. In addition, more and more materials are added and uploaded by various educationist people in helping the mind. But nowadays educational materials, information or educational related data are being collected from many online sources by some educational organizations. After that mined material is used to remake or reproduce like own. Most educational organizations are being used to earn money. The same challenge has been for online course programmers and authorised competitive educational centres. Theft cannot be eradicated if the thief does not change their activities. Therefore, we must be vigilant and careful. We should not be deceived in our text data. So, unethical behaviour activities should not be used by us.

Protection

In this pandemic period, text mining protection is important for educationists, educational organizations, research centres, competitive coaching centres etc. Text mining is misbehaviour, according to mining the creative knowledge of a genius person and changing it into one's work. Bible says,give instruction a wise man, and he will be still wiser; teach a righteous man, and he will increase in learning, Proverbs: 9;9. Preserving is a rare act rather than creating. So, every educational organization should use specialized software, protection keywords, and password by them to avoid text mining. Otherwise, individualised can make user IDs, the password for files, and folders. We should keep trusted people to handle the educational properties. Only needy information, files, and materials should be uploaded or discussed online. To protect from text mining,

every educational institution, coaching centre, preparing material group or research centres must be given protection keys to educational information, text messages, documents, files, materials, etc.

Conclusion

This paper addresses the ethical behaviour of avoiding text mining in digital education inthe twenty-first century as it has explained how to act with ethical behaviour in others' text data in the current situation. This article covered behavioural activities in text mining. Nowadays, text mining is to be challenged for educationists. Many authors' quotes have been given to protect from text mining. This article advises following ethical behaviour in an educational text message, information, materials etc. Further, this article explains what text mining is and how to use it with ethical behaviour. We will hope that text mining protection will have been got to everyone in future. Well, behavioural activities will be got a respectable name among educationists.

References

Abrol, V., and J. Malhotra. "Data Dashboard-Integrating Data Mining with Data Deduplication." *International Journal of Computer Applications*, vol. 71, no. 22, 2013, pp. 28-33.

Agrawal, R., and M. Batra. "A Detailed Study on Text Mining Techniques." *International Journal of Soft Computing and Engineering*, vol. 2, no. 6, 2013.

Ajmal, E.B., and Rosna P. Haroon. "Maximal Marginal Relevance based Malayalam Text Summarization with Successive Thresholds." *International Journal on Cybernetics & Informatics*, vol. 5, no. 2, 2016, pp. 349-56.

Allahyari, M., et al. "Text Summarization Techniques: A Brief Survey." *International Journal of Advanced Computer Science and Applications*, vol. 8, 2017, pp. 397-405.

Bakhshinategh, B., et al. "Educational Data Mining Applications and Tasks: A Survey of the Last 10 Years." *Education and Information Technologies*, vol. 23, 2018, pp. 537-53.

Chena, Z.X., and C.C. Chena. "The Intention of Using Smart Device Messages as Parent-Teacher Communication from the View of

Parents", *Procedia Manufacturing*, vol. 3, 2015.

Das, Dipanjan. et al. "An Exact Dual Decomposition Algorithm for Shallow Semantic Parsing with Constraints." *Conference on Lexical and Computational Semantics*, 2012, pp. 209-17.

Ecclestone, Kathryn, and Dennis Hayes. "Affect, Knowledge, Communication, Creativity and Emotion." *Beyond Current Horizons*, 2008.

Eger, Thomas, et al. "The Determinants of Open Access Publishing: Survey Evidence from Germany." *European Journal of Law and Economics*, vol. 39, 2015, pp. 475-503.

Freeman, D.W., and W.A. Sistrunk. "Effects of Post-Harvest Storage on the Quality of Canned Snap Beans." *Journal of Food Science*, vol. 43, no. 1, 1978, pp. 211-14.

Füredi, J., et al. "The Role of Symptoms in the Recognition of Mental Health Disorders in Primary Care." *Psychosomatics*, vol. 44, no. 5, 2003, pp. 402-06.

Gonzalez, G.H., et al. "Recent Advances and Emerging Applications in Text and Data Mining for Biomedical Discovery." *Briefings in Bioinformatics*, vol. 17, 2016, pp. 33-42.

Groark, Kevin P. "Taxonomic Identity of "Hallucinogenic" Harvester ANT *(Pogonomyrmex Californicus)* Confirmed." *Journal of Ethnobiology*, vol. 21, no. 2, 2001, pp. 133-44.

Gupta, V., and G.S. Lehal. "A Survey of Text Mining Techniques and Applications." *Journal of Emerging Technologies in Web Intelligence*, vol. 1, no. 1, 2009, pp. 60-76.

Kumar, Lokesh, and Parul Kalra Bhati. "Text Mining: Concepts." *Journal of Global Research in Computer Sciences*, vol. 4, no. 3, 2013 pp. 36-39.

Kumar, Mohit, et al. "Summarizing Non-Textual Events with a 'Briefing' Focus." *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, 2007, pp. 210-25.

McMillan, Julie. *Course Change and Attrition from Higher Education*. Longitudinal Surveys of Australian Youth Research Report 39, 2005.

Muller-Langer, Frank, and Richard Watt. "Copyright and Open Access for Academic Works." *Review of Economic Research on Copyright Issues*, vol. 7, no. 1, 2010, pp. 45-65.

Natarajan, Jeyakumar, et al. "Text Mining of Full-Text Journal Articles Combined with Gene Expression Analysis Reveals a Relationship between Sphingosine-1-Phosphate and Invasiveness of a Glioblastoma Cell Line." *BMC Bioinformatics*, vol. 7, 2006.

Oliveira-Neto, Francisco Moraes, et al. "An Online Self-Learning Algorithm for License Plate Matching." *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, 2013, pp. 1806-16.

Romero, Cristobal, and Sebastian Ventura. "Data Mining in Education." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, 2013, pp. 12-27.

Samha, Amani K., et al. "Aspect-Based Opinion Mining from Product Reviews using Conditional Random Fields." *Conferences in Research and Practice in Information Technology Series*, 2015, pp. 119-28.

Slater, Stefan, et al. "Tools for Educational Data Mining: A Review." *Journal of Educational and Behavioral Statistics*, vol. 42, no. 1, 2017, pp. 85-106.

Usai, Antonio, et al. "Knowledge Discovery Out of Text Data: A Systematic Review Via Text Mining." *Journal of Knowledge Management*, vol. 22, no. 7, 2018, pp. 1471-88.

Willinsky, John. "Democracy and Education: The Missing Link May Be Ours." *Harvard Educational Review*, vol. 72, no. 3, 2002, pp. 367-93.

Zack, M.H. "What Knowledge-Problems Can Information Technology Help to Solve?." *Americas Conference on Information Systems*, 1998, pp. 664-46.

Author Details
P. Tamil Selvan, *Alagappa University, India*, **Email ID**: *tamilselvankirubai@gmail.com*
Prof. G. Kalaiyarasan, *Alagappa University, India*, **Email ID**: *gkauce@yahoo.com*