TRENDS AND CHALLENGES IN ADOPTING NATURAL LANGUAGE PROCESSING IN EDUCATION

Asna Ashraf P and Naseeba V K, Students, PG Department of Computer Science,8075031795,8891917002,aznaazhar1@gmail.com,faizsheebu@gmail.com

ABSTRACT

Artificial intelligence (AI), the ability of a digital computer system or a system controlled- robot to perform multiple tasks commonly associated with intelligent beings. The term AI is frequently applied to the project of developing systems furnished with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from the past experience. Since the development of the digital computer systems in the 1940s, it has been demonstrated that computers can be programmed to carry out multiple complex tasks. Natural language procesing or NLP is a field of artificial intelligence in which computers can analyze, understand, and acquire meaning from human language in a useful way. NLP is used to analyzing text, allowing machines to understand how human speak. This human-computer interaction enables real-world applications like automatic text sum up, sentiment analysis, topic extraction, named entity recognition, parts-of-speech tagging, relationship extraction, stemming, etc... NLP is commonly used for text mining, machine translation, and automated question answering. This article illustrate on existing NLP challenges and applications that could be adapted to educational applications. In education, Trends and challenges in adopting NLP were reviewed and explored.

KEYWORDS: Artificial intelligence, Feature extraction, Natural language processing, Deep learning, Machine learning,

INTRODUCTION

Artificial Intelligence (AI) is a intresting topic withits human-like intelligence in building decision-making systems. With the capacity for prediction and classification, AI can revolutionize education. It is done by processing huge amounts of structured data sets such as SQL databases and unstructured datasets such as videos and audios. AI introduces machine learning methodologies to personalize the student learning experience via learning management sys-tems, deep learning, and transfer learning to use pre-trained concepts to deal with new similar problems, natural language processing (NLP) methods to listen tostudent feedback, process them and output predictive insights on their opinion towards learning infrastructure. AI can transform existing educational infrastructures namely online tutoring, learning management systems, curriculum, employment transitions, teacher training, assessments, and research training. The institutional project data are diverse and inclusive of student feedback in textual format classroom recordings in video and audio formats.

Chassignol et al. defined AI as an ''Artificial Intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment''. All Educational institutions have extensively adopted AI in different forms of service for students. One of the most widely used AI methodologies for student is NLP . The key role of NLP is in interpreting feedback or opinions of end-users. To understand end-users feedback, most institutions in the world invest their time and resources.

AI's impact on education and discovering the opportunities in the education domain, educational institutions have focused on building a cognitive intelligent system using AI. In this process, the foremost step is to listen to students' opinions and feedback on existing educational infrastructure, teaching practices, and learning environments. In academic institutions, it is traditional practice to request student feedback to gather students' perception of the teach-ing team and their learning experience in the course. Quantitative or qualitative formats are the type of students feedback. To rate the performance or textual comments to questions, we use numerical answers . Monitoring and tracking students' feedback manually is a time-consuming and resource demanding task. With annotation and summarization capabilities, NLP can contribute to this task. This article reviewed NLP methodologies that can contribute to the education domain, and the following questions were explored:

• What are the being used for NLP?

• In the education domain, What are the generic challenges of using NLP ?

• In student feedback analysis, What are the current trends of NLP ?

• How can NLP methodology in other disciplines be adopted to the education domain?

Machine learning and deep learning are part of AI methodologies. Machine learning is a set of algorithms can analyze data, learn and apply. Deep learning techniques holds multi-layer neural network with processing layers to train new concepts and link to previously known concepts. Deep learning enhances NLP with concepts like continuous-bag-of-words and skip-gram model. Convolutional neural networks (CNNs) , recurrent neural networks (RNNs), their special cases of long short-term memory (LSTM) and gated recurrent units (GRUs) are different forms of deep learning techniques used in text classification. In this article, analyzation of text data on existing works using the AI methodologies are explored. The methods can be adopted to students' feedback analysis, although few research works were not directly related to it.

The following are contributions of this research:

• Enhanced understanding of the impact of AI on education with open opportunities in the industry.

• Synthesis of existing NLP methodologies to student user feedback and annotate their views.

• Exploring trends and challenges in NLP that need to be addressed to be adopted to the education domain.

EXISTING METHODOLOGY

Feature extraction and feature selections are mandatory data preprocessing steps to transform text data into quantitative vector formats.In this section, existing methods in feature extraction, feature selection, and topic modelling will be discussed.

A) FEATURE EXTRACTION

The students feedback data are collected and transform it for machine learning modelling. This is done by applying Feature extraction techniques. For example, in NLP, there are feature extraction techniques like Bag of Words (BoW), Term Frequency (TF)-Inverse Document Frequency (IDF), and Word Embedding.

Bag of Words (BoW) is a common feature extraction method that involves a vocabulary of known words and a measure of the presence of known words. With known words in a document the BoW is only concerned. It will not bother about the structure or order of words in a document that ignores the context of the words. TF-IDF estimates the importance of each word or term in a document based on their weights. IDF gives how common or rare a word is in a corpus. The closer the value is to zero, the more common a word is in a corpus. TF-IDF means multiplication of TF and IDF. Word Embedding is a learned representation of text with similar meaning. It reduces dimensionality and enhances the generalization process.

The most common word embedding techniques are Word2Vec, GloVe, Doc2Vec and Bidirectional encoder representations from transformers (BERT) . The neural networks model using Word2vec algorithm. To learn word associations from a large corpus of text this will help. The synonymous words or even suggest additional words for a partial sentence can be detect the trained algorithm. Word2Vec generates the number of dimensions for each word in a corpus and then searches at the context level of the occurrence of the words in a sentence. In a vector space, all the words with similar contexts are grouped. A GloVe approach combines the matrix factorization technique and latent semantic analysis (LSA) with a context-based learning in Word2Vec. Doc2Vec is a tool to create vector or numeric representations of documents. BERT is a pre-trained deep bidirectional representations from unlabeled text. Both left and right context in all layers jointly conditioning to do this. BERT can perform word or sentence embedding to extract vectors from text data.

B. FEATURE SELECTION

Feature selection is a process of reducing data dimensionality in terms of features. This would maintain or enhance the performance of a machine learning algorithm. The reduction criteria would simplify a model's complexity and consistently maintain accuracy. Considering n features in a dataset, the number of the feature subsets would be 2n. An increase in the features count would make the modelling infeasible. The stability or robustness of feature subsets was evaluated by grouping similar features or considering all feature subsets, removing the non-contributing features, and the size of the feature subsets. The feature subset evaluation methods are broadly categorized as filters, wrappers, or embedded methods.

1) FILTER METHODS

Filter methods rank the key features and select high representative features by setting a threshold. As shown in Figure 1, filter methods rank the features and select them before actual modelling. In addition, this technique filters the low importance features before training a model.

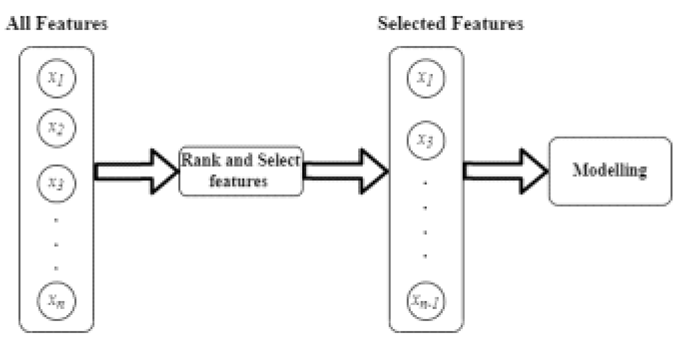

FIGURE 1. Filter methods.

The feature importance technique assesses two measures in ranking the features. The first measure is to check the predictive power of each feature toward the target variable(s). These are called correlation criteria or dependence measures. Mutual information, χ2 statistic, Markov blank, and minimal-redundancy-maximal-relevancy techniques extract a feature's correlationwith a target variable. The second measure in the feature importance technique is redundancy, which assesses the features with redundant information. This detects the redundant features by evaluating relevant measures among the independent variables. An article by Wang presented a redundant feature analysis. Its process is to find the most relevant features in predicting the target variables and use the relevant features to estimate the redundancy in other features.

2) WRAPPER METHODS

Wrapper methods defined as a subset of features using a predefined classifier and then the performance of the subset of features is evaluated using predefined classifiers. In wrapper methods, a machine learning algorithm is used to upload the feature selection performance.

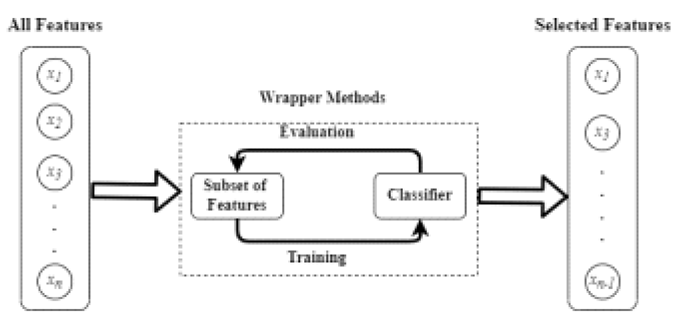

FIGURE 2. Wrapper methods.

As shown in Figure 2, a subset of features is selected and trained by a classifier with the selected features. Then, the performance of the classifier is examined. Sequential forward selection (SFS) is an example of a wrapper method with sequential feature selection methods. It is a greedy search algorithm that extracts an optimal subset of features iteratively based on the classifierperformance. Features are selected one-by-one from the pool of all features iteratively.

3) EMBEDDED METHODS

Embedded methods normally a combination of a filter method and a wrapper method. They defeat the challenges of low accuracy in filter methods and slow computation speed in wrapper methods. Embedded methods analyze the optimum features, contributing to the classifier's accuracy.

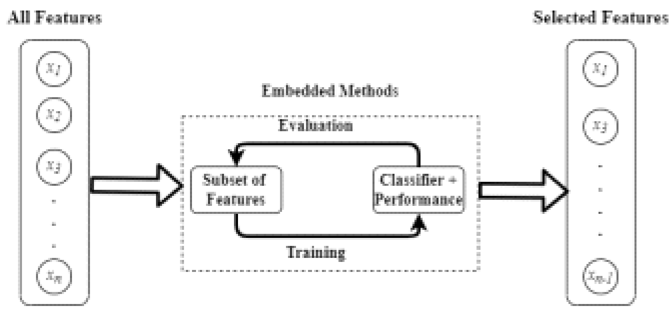

FIGURE 3. Embedded methods.

As shown in Figure 3, embedded methods evaluate the performance of each subset of features. One of the most common embedded methods is regularization, which is to reduce the degree of overfitting or variance of a model by adding a penalty against its complexity for L1 regularization methods.

C. TOPIC MODELING

Topic modelling automatically analyzes a entity of documents with text data techniques using machine learning techniques and determines cluster words. The technique does not need any training to cluster the words from the entity. This is an unsupervised machine learning technique.Topic modelling divides a corpus of documents into groups to extract a list of topics covered, and several sets of documents are grouped by the topics they covered. The topic modelling techniques are broadly classified into probabilistic and non-probabilistic models.

1) NON-PROBABILISTIC MODELS

Non-probabilistic models are matrix factorization algebraic approaches. These models are use with latent semantic analysis (LSA) and Non-negative matrix factorization(NMF). Both LSA and NMF mechanisms works on BoW approaches. As discussed in Section II-A, BoW converts a corpus into a term-document matrix to extract the frequency of the terms and ignores the order of the terms. LSA is an algebraic method that produce a matrix with words presented in a corpus. It concludes that words that are similar in meaning will occur very close in the text. The technique is based on single value decomposition (SVD) which optimizes the number of words while preserving a similar structure. The similarity of the texts will be computed using vector representation and organized into semantic clusters.NMF transforms high dimensional data into low dimensional data with no negative components and clusters simultaneously. This is also known as positive matrix factorization (PMF). It is an unsupervised machine learning technique that can extract relevant information without previous insights into the original data.

2) PROBABILISTIC MODELS

Probabilistic models are fully unsupervised approaches that are tweaked to guide in latent dirichlet allocation (LDA) modelling and semi-supervised learning in a probabilistic latent semantic analysis . Probabilistic latent semantic analysis (PLSA) is to detect semantic co-occurrence of words or terms in a entity. This is built based on the first statistical model, a model that revealed the semantic cooccurrence in a document term matrix of the corpus. Due to its unsupervised nature, PLSA is capable of determining the number of topics, the probability of a topic and the probability of a document containing the topic. It groups unknown topics of every existing document. LDA is a commonly used technique in topic modelling which is built based on De Finetti's theorem, which states that positively correlated exchangeable observations are conditionally independent relative to some latent variable. It can capture inter and intra document statistical structures on assumptions that a corpus has a predefined number of topics and each document in the corpus has a different proportion of the topics. It is a hidden variable model which uncovers hidden patterns in gathered data in a corpus.

An LDA technique based on a topic modelling methodology was selected in mobile learning research to find the topic trends. Out of 50 topics extracted from the LDA, 25 topics were selected and grouped into three dimensions of technology, learning and learners in that mobile learning. Similarly, as part of designing a course structure for virtual reality with augmented reality and mixed or extended reality, the LDA technique was employed in the research study. The study was to understand the motive of learners (students) in joining the course using topic modelling.

It revealed that learners had little experience in designing virtual applications. Also, the learners had little experience in a programming language. Designing a massive open online course without understanding learners' engagement would lead to a high dropout rate.

D. TEXT EVALUATION

In this subsection, NLP applications like text summarization, document categorization, text annotation, and knowledge graphs are discussed.

1) TEXT SUMMARIZATION

There has been an exponential growth in collection of student feedback for evaluation in educational institutions. Consolidating the content and extracting useful resources is a tedious task and would consume massive efforts. The summary of the content would be easier for readers to digest and comprehend. Text summarization technique provides a summary of a student feedback or corpus of the feedback text without losing critical information. Text summarization can be categorized into extractive, abstractive, and hybrid approaches.

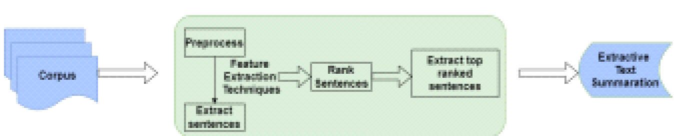

FIGURE 4. **Extractive text summarization.**

Extractive Text Summarization is a traditional text summarization method. It extracts significant sentences as it is from the document and adds to the summary. As shown in Figure 4, the technique selects a subset of the sentences in an original text using feature extraction techniques like BoW, N-gram, graphs and so on. The extracted sentences are ranked based on their importance. It creates an intermediate representation that highlights the most important information included in the original text.

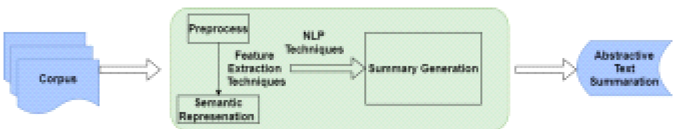

FIGURE 5. Abstractive text summarization.

Abstractive Text Summarization extracts sentences from documents in an intermediate representation and generates a summary of the sentences instead of the original sentences as shown in Figure 5. The technique paraphrases the sentences using NLP techniques and generates a summary that is suitable to human interpretation.

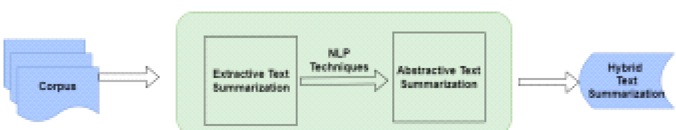

FIGURE 6. Hybrid text summarization.

Hybrid Text Summarization is an ensemble of extractive and abstractive text summarization as shown in Figure 6. In this mechanism, the top-ranked sentences extracted fromextractive text summarization are paraphrased using NLP techniques and summarizes the content abstractedly.

2) DOCUMENT CATEGORIZATION

Document categorization is one form of annotation to annotate a document in a text corpus. It analyzes the content, intent and sentiment within a document and classifies them into predefined labels.

Document categorization or text classification analogizes end-to-end entity linking where an entity linking labels individual words or phrases, document categorization annotates an entire text or body of a document with a single label. Sentiment annotation and linguistic annotation are part of document categorization to extract latent semantic and linguistic elements in a document.

3) ENTITY EXTRACTION

To identify named entities, parts of speech and key phrases within a text, an entity annotation technique can be used. Annotators read the text thoroughly to locate the target entities based on predefined labels. The located entities in entity annotation can be connected to larger repositories of data using entity linking. In end-to-end entity linking, preprocess a piece of text for named entity extraction. In entity disambiguation, extracted named entities will be linked to knowledge databases.

4) KNOWLEDGE GRAPHS

Knowledge graphs can represent information extracted using NLP in an abstract form and integrate the information extracted from multiple data sources. Domain knowledge from knowledge graphs are input into a machine learning model to produce better predictions. A knowledge graph can be served as a data structure which can store information. A combination of human input, automated and semi automated extracted data can be applied to a knowledge graph.

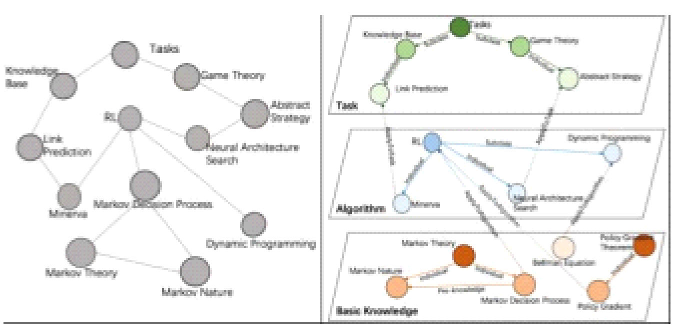

FIGURE 7. Sample knowledge graph [118].

5) SENTIMENT ANNOTATION

One of the most trending annotations in NLP is sentiment annotation which is to label emotion, opinion and sentiment inherent within a text. The label could be a positive sentiment, neutral sentiment, or negative sentiment. It deals with emotional intelligence quotient in sentiment analysis or opinion mining. In natural language, understanding the context is critical. Without comprehensive understanding, it is difficult to predict the true emotion behind a text message or email. It is much more difficult for machines to mine customer intention in reviews or feedback, especially with sarcasm and humour. Sentiment annotated data are used to train machine learning models and assist them to do sentiment analysis or opinion mining.

CHALLENGES

In this section, we are discussing challenges for implementing NLP techniques in the education domain

A. DOMAIN-SPECIFIC LANGUAGE

This is considered one of the challenges in imple-menting NLP in educational domain. Considering many student feedback being generated from different surveys,questionnaires, and other educational feedback acquiring por-tals on a course teaching or a learning management system. Without understanding or getting trained on the specific domains,submitted a domain-specific NLP for students, faculty mem-bers, universities in computer science or information tech-nology in higher education sector. The authors extracted some tech-related skills using named entity recognition (NER) and built a personalized multi-level course recommendation system.The will authors calculate the grades based on the relevance to the topics created by a teacher or auto-generate the text from the subject area.Extracting entities or concepts from a huge database available using a data scraping technique and processing them with considerable manual annotation would assist in building body for an application domain.

B.SARCASM

Decoding sarcasm is critical in NLP tasks like opinion analysis. This helps to student opinions and perceptions on course structure and educational infrastructure. To detect sarcasm, Rule-based, statistical, and deep learning are three different approaches that reported by authors. In a rule-based approach, sarcasm can be identified based on key indicators of sarcasm captured as evidence ,In a statistical approach to detect sarcasm, punctuations, sentiment-lexicon-based features, unigrams, word embedding similarity, frequency of the rarest words, sentiment flips and so on ,In deep learning algorithms, RNN models and LSTM methods can be used individually as well as in combination with CNN for automatic sarcasm detection .The survey article will be provided a comprehensive understanding of sarcasm detection.

C. AMBIGUITY

Ambiguity in natural languages is common as it based on context and user perception in reading a text. With challenges in decoding a context, ambiguity in machine learning language processing is more complicated. Ambiguity could be due to the structure, syntactic, or lexical nature of a sentence or text. In structural ambiguity, a sentence includes more than one syntactic structure. In syntactic ambiguity, a grammatical construction error occurs in a sub-part of a sentence that causes grammatical ambiguity in a whole sentence. Lexical ambiguity is due to a word having two different meanings . Addressing the ambiguity challenges is crucial in analyzing feedback.

D. EMOTICONS AND SPECIAL CHARACTERS

Emoticons and special characters play an important role in opinion mining especially students' feedback containing the special symbols to express their emotions. NLP has a challenging phase in processing the emoticons and characterize them with appropriate emotion tags. In 2020, the authors are studied and analyzed cross-cultural reactions to the novel coron-avirus and detected sentiment polarity and emotion from their tweets and validated them with emoticons. Six emotions of joy, surprise, sadness, anger, fear, and disgust were validated using different types of emoticons with their unicodes.Cappallo et al.proposed a large dataset and explained three challenges in emoticons processing. They are emoji processing, emoji anticipation, and query-by-emoji.

E. ASPECT-BASED SENTIMENT ANALYSIS

Most of the research works to process student comments or their feedback to arrange the positive or negative sentiment using lexicon-based or machine learning methods at document level. Nazir et al. conducted a survey. It is based on issues and challenges. They are related to extraction of different aspects. The study was divided into three topics aspect extraction, aspect sentiment analysis, and sentiment evolution. Each topic was breakdown into sub-categories explicit aspect extraction, implicit aspect extraction, aspect level sentiment analysis, entity level sentiment analysis, multi-word sentiment analysis, recognition of factors in sentiment evolution, and predicting sentiment evolution over social data.

F. DATA IMBALANCE

Data imbalance is one of the most common challenges in AI in which number of samples in one class exceeds the amount in other classes. A subset of AI, the challenge is inherited by considering NLP. It is difficult to acquire of massive labelled data as it requires manual annotation from domain experts, especially in education domain,. Although the acquired labelled data fed to deep learning algorithms. Due to data distribution discrepancy, the classification performance is based . A potential tool to overcome this challenge could transfer learning, where a deep learning model trained on a large corpus of student feedback to perform similar tasks on another data source. Other techniques could be sampling techniques to under-sample majority classes or over sample minority classes, which might demand text augmentation tasks.

CONCLUSION

The study was aimed is to explore existing NLP methodologies, that can be implemented or adopted in education domain. This assist to understand AI impact on education. With open opportunities, synthesize the methods to process student feedback, and annotate their views. The literature review has been performed using Google Scholar. Covering bibliographic databases like Wiley, Scopus, Springer, ACM Digital Library, IEEE Xplore, Pub-Med, Science Direct, and Multidisciplinary Digital Publishing Institute (MDPI) and so on.

The impact of AI on education was discussed, in this review article. And also the scope of introducing AI into educational institutions is detailed in this study. It is based on the opportunities. Existing NLP methodologies were explored, in this article. Explanation about Feature extraction, feature selection and topic modelling methodologies are mentioned in this study. It is explained with brief definitions. Moreover, text evaluation techniques text summerization, annotation, and knowledge graphs were reviewed. Each of these applications was defined. And also existing approaches were discussed. Review of Challenges in adopting NLP methodologies to the education domain is taken. This study is confined to AI implementation methodologies. These methodologies are with less focus on pedagogy concepts. This is limitation of this research. Data scarcity and class imbalance are data specific challenges. This were not discussed. The future direction of this research would be to explore data challenges. While extracting feedback or opinions without affecting privacy. This would affect the model learning for deep learning algorithms. They are data hungry. The strategies to interpret deep learning models (black box) were not explored.

REFERENCES:

[1] M. Chassignol, A. Khoroshavin , A. Klimova , and A. Bilyatdinova, ‘‘Artificial intelligence trends in education: A narrative overview,’’ Proc. Comput.Sci., vol. 136, pp. 16–24, Jan. 2018, doi: 10.1016/j.procs.2018.08.233.

[2]M. Wang, X. Tao, and F. Han, ‘‘A new method for redundancy analysis in feature selection,’’ in Proc. 3rd Int. Conf. Algorithms, Comput. Artif. Intell., Dec. 2020, pp. 1–5, doi: 10.1145/3446132.3446153.

[3] A. Nazir, Y. Rao, L. Wu, and L. Sun, ‘‘Issues and challenges of aspect-based sentiment analysis: A comprehensive survey,’’ IEEE Trans. Affect. Comput., early access, Jan. 30, 2020, doi: 10.1109/TAFFC.2020.2970399.

[4] Thanveer Shaik, Xiaohui Tao X. Li, Yan Li, Christopher Dann, Jacquie McDonald Petrea Redmond, Linda Galligan, ‘‘A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis,’’ IEEE Access, vol. 10, pp. 56720 - 56739, 25 May 2022.