International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 12 Issue 11, November-2023, Impact Factor: 7.957

Revolutionizing STEM Policy through Advanced Machine Learning Techniques

Anitha R[1], Meharuniza Nazeem[2], Dr. S Anil Kumar[3]

[1]Department of Futures Studies, University of Kerala, Trivandrum-695 581, India
[2]ICFOSS, Green Field Stadium, Karyavattom, Trivandrum, Trivandrum-695 581, India
[3]Department of Futures Studies, University of Kerala, Trivandrum-695 034, India

ABSTRACT

This paper provides a description of textual document analysis techniques from a social science perspective. The goal is to describe the current methodological standpoint of computer-assisted text analysis in the social sciences (not to provide an exhaustive list of all computer-assisted text analysis investigations) that are either directly or indirectly related to the social sciences, which use a quantitative and computer-assisted methodology as their text analytical tool, i.e., the processes performed in computer-assisted text analysis studies are described and discussed. Examining and describing a few of the more recent and cutting-edge approaches in the field and procedures applied Cataloging the different types of supplemental data and computational assistance that are still needed to increase the method's acceptability and usefulness for a wide range of text analysis aims the ranking of keywords is done by year-wise comparative analysis which can be used to determine which words are most commonly used in the policy document.

Keywords: NLTK, Word Cloud, and Network X

INTRODUCTION

Science and technology (S&T) are the foundation of the contemporary world, and in the knowledge-based 21st century, employment and education in STEM sectors are crucial for India. Higher education in STEM fields is absolutely necessary to prepare scientists and engineers who will create, adapt, and use new technologies. Despite having the world's greatest young population, India must also have the largest young workforce with the most advanced knowledge abilities. As a result, producing a sufficient pool of skilled STEM workers is a challenge for universities and colleges. STEM is therefore crucial to developing a workforce that can compete in the 21st century. She wants young individuals who have the necessary educational background and job-specific skills. Higher education and tertiary-level skills are required in India in order to raise both educational attainment levels and the population's proportion of scientists and engineers. The goal of the current study is to identify any distortions in the policy of STEM-related jobs and education. This is accomplished through the use of text analytics in STEM education policy materials.

Numerous linguistic, statistical, and machine-learning techniques are used in text analysis. Information is retrieved from unstructured data using text analytics, which also involves structuring the input text to identify patterns and trends and analyzing and interpreting the results. In our work, we compare text analysis methods using Python and R programming. To transform unstructured data into useful data for analysis, text analytics software solutions offer servers, analytic algorithm-based applications, data mining, and extraction tools. For analysis by other tools like business intelligence tools, big data analytics tools, or predictive analytics tools, the outputs—extracted entities, facts, and relationships—are typically stored in relational, XML, and other data warehousing applications. The process of turning unstructured text data into information that can be used for analysis, measurement of consumer feedback, product reviews, and sentiments, as well as for search functionality, entity modeling, and sentimental analysis to assist fact-based decision-making, lexical analysis, classification, clustering, pattern recognition, annotation, tagging, information extraction, link and connection analysis, visualization, and predictive analytics are all included. Text analytics and access to unstructured data have various applications. Big data analytics aims to comprehend and address actual issues. While a few studies have used new data sources to address significant research problems in the hospitality industry, big data analytic techniques have not been systematically applied in studies that focus on policy texts.

Analysis of the policy documents reveals that the earlier ones placed less emphasis on science and technology (S&T). The country is attempting to advance in its economic development through the enhancement of STEM education, on the other hand, which is given top priority in later documents.

STEM education is important in the knowledge-based economy. The term STEM, which has come to refer to any activity involving one or more of the STEM disciplines, has its roots in the 1990s at the National Science Foundation (NSE) of the US. STEM occupations are those that provide professional and technical support in the domains of computer science and mathematics, engineering, and life and physical science, according to the Economic and Statistics Administration (ESA). The involvement of students in STEM education is a topic of discussion in many nations. STEM workforce growth is cited as a major priority by leaders in business, government, and academia. This is a result of competing economic goals on a global scale, severe labor shortages in the STEM fields, and waning student enthusiasm in the field. Nevertheless, despite shared attitudes toward policy and identified issues, countries vary greatly in terms of the proportion of students that pick STEM fields of study.

With nearly 17 million students enrolled in hundreds of institutions and colleges around the nation, India upholds the foundation of one of the world's largest higher education systems. Nearly all industrialized nations around the globe, as well as many of India's developing counterparts, have extensive STEM education programs. Although students in India are just as capable and inquisitive as those in any other country, few of them conduct real research. A strict educational system that prioritizes test scores over creativity and innovation forces our sizable youth population through it. Those who put substantial time and effort into research projects frequently discover that their work goes unnoticed, and this presents yet another issue. This occurs because our educational system does not promote invention and curiosity; in contrast to other nations where these young geniuses are recognized and invited to the most prestigious events, in India they are frequently overlooked. As a result, India has very low aspirations for innovation at the tertiary or school levels.

2. Challenges in STEM education for 'Skill India'

The "Skill India" program is anticipated to place a significant strain on the nation's educational institutions. The "Make in India" initiative of Indian Prime Minister Narendra Modi, which seeks to establish India as a significant global manufacturing hub, will place significant demands on the nation's educational institutions, particularly in the STEM (Science, Technology, Engineering, and Mathematics) disciplines, to create graduates with excellent credentials.

This shows that in order to prepare students for professions in the STEM disciplines, educational institutions need to pay close attention to the most recent developments in science and technology. In order to meet this demand for competent graduates, the "Skill India" initiative needs assistance from all parts of the nation, following the American example. The advertisement also raises new questions regarding the STEM labor force in India.

Quality STEM education

Infrastructure and the associated costs of building high-quality infrastructure are two major obstacles to STEM education. The corporate sector must work with the educational sector primarily for this reason. Over the years, the price of conducting high-quality research has undoubtedly skyrocketed. Furthermore, cutting-edge facilities are necessary for cutting-edge developments.

In the United States, a number of private foundations make major contributions to children’s education, beginning in elementary school. The largest privately sponsored education program in the United States is run by the Howard Hughes Medical Institute (HHMI), making it the most prestigious. From elementary schools to graduate institutions and beyond, the program assists students. Other examples of private initiatives to aid students in their scientific endeavors are the Burroughs Wellcome Fund (BWF) and Intel Education STEM resources.

The opportunities for training and participation in research should be available to all areas of society, including minorities, the underprivileged, and the economically disadvantaged classes, in order to make the Indian government's phrase "SabkaSaathSabkaVikas" (Development with everybody, for everyone) a reality. Because of the Higher Achievement program, which also arranges frequent visits to the laboratories of partner universities, underrepresented groups in the U.S. have the opportunity to learn about research.

In an effort to increase female students' engagement in the research workforce, promote innovation among them, and encourage their continued growth in these sectors, these programs also specifically target female students.

Additionally, having qualified teachers who are aware of the most recent developments in their fields is necessary for STEM education. For the next generation of pupils to be guided, India, a nation that has historically elevated gurus to high office, would need a well-trained army of teachers. Indian officials may find inspiration from the White House's plan to train 10,000 STEM teachers to become master instructors.

STATEMENT OF THE PROBLEM

STEM is crucial for creating a highly skilled workforce that is competitive in the 21st century; hence, it has long been difficult for Indian universities and colleges to produce a sufficient pool of qualified STEM graduates. India has the largest youthful population in the world, but she must also have the largest young labour force with the best degree of knowledge and skills. According to the most recent ADB Report (2014), India can grow into a significant global knowledge-based economy (KBE) in the years to come if it takes advantage of its human capital capabilities. She must raise educational levels and the population's proportion of scientists and engineers in order to accomplish this, and in order to do so, higher education and advanced tertiary skills are required. Given that India has the largest proportion of young people in the world, STEM has played a crucial role in the nation's economy in this knowledge-based 21st century. There are 226 million Indian youngsters between the ages of 10 and 19 who are prepared for higher education. According to the Economic Survey 2013–14, by 2020, 125 billion people will have an average age of 29 years, which would be the lowest age on record. Due to this demographic advantage, it is crucial to fund STEM education and increase its potential for contributing to human growth. However, there are still many gaps in the creation and application of policy. Due to these gaps and policy distortions, research is being proposed to determine ways to improve STEM education and employment

Objectives of the Study

1. To analyze the roadmap of policy related to STEM fields
2. To understand the relationship between keywords in STEM fields using text data visualization tools and techniques
3. To estimate the relationship between keywords in STEM fields using statistical tools and techniques.

Hypotheses of the study

1. In India, universities play a major role in upholding STEM fields
2. A high incidence of STEM keywords is observed in later periods compared to earlier periods due its increasing importance in India’s policy.

METHODOLOGY

The research begins with data gathered from the UGC website and uses LDA modeling to categorize the annual report data gathered from reports. Following that, keywords were analyzed using SPSS to discover how each year's keywords were clustered.

Step 1: Collecting documents for the analysis

Initially, data collected from the UGC website from the years 1968 to 2018 is there. The qualitative data of social sciences collected from the site is policy documents for stating the annual reports on newly launched policies and the changes that are brought to the existing policies.

Step 2: Converting a PDF file into a text file

The data collected was in pdf format, which we further converted into a.txt file so as to access the data.

Step 3: The text files are after the Python code.

The text data is processed after the word cloud to extract keywords and then plot those keywords into a graphical representation using the Network tool.

Step 5: Creating Topic modeling for clustering

The output obtained from the analysis consists of a topic-wise categorized list with the weight of each word in the topic. This needs to be executed for all the documents from 1968 to 2018.

Step 5: Insert the data keywords into the SPSS

In this step, SPSS clusters the keywords in each year and then compares the two related data using the **Wilcoxon Signed Ranks Test**. In this step, the previous year was then compared.

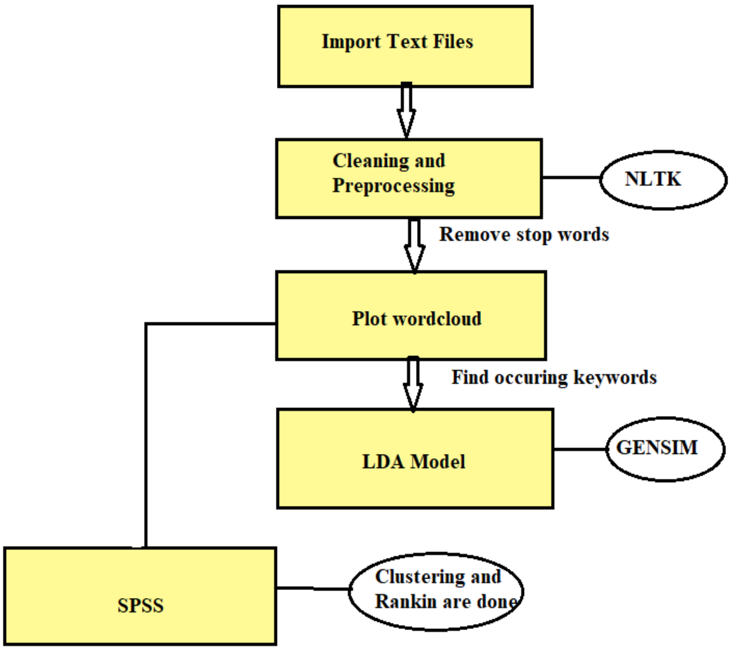

Figure 1: Diagram for Methodology

REVIEW OF LITERATURE

This paper illustrates Science, technology, engineering, and mathematics are subjects that will play a significant role both today and in the future in order to survive in the world of science and technology [1]. To adapt to advances brought about by the rapid advancement of information and technology, people must be qualified. With the use of technology in education, this would be accomplished. In this regard, the number of studies being done to advance STEM education (science, technology, engineering, and mathematics) is also rising daily. A total of 40 academic research, including articles and papers published in national and international journals, are discussed in this study. Information on the STEM studies is provided. Within this context, studies on the issue conducted in accordance with the findings of the content analysis are looked into. Evaluations are made of research methodologies, research approaches, learning environments, learning outcomes, data collection tools, research samples, STEM study subjects, and the body of literature. Descriptive statistical techniques are used to analyze the study data. Tables with frequencies and percentages show the results. The study's findings highlight the preference for qualitative research in studies of STEM education, the use of experimental research and surveys as research techniques, and the prevalence of primary school children in these studies. It is believed that this study will be useful to researchers who are enthusiastic about researching the topic.

This study uses content analysis to assess studies on STEM education in the math and science disciplines conducted between 2010 and 2015 [2]. Five criteria—publication year, disciplines (mathematics, science, and science and mathematics), methodology (qualitative, quantitative, and mixed), study regions, and participants—were utilized to evaluate the papers that were reached. Particularly, full-text articles obtained via the Internet were indexed in SSCI (Social Science Citation Index). The research comprised a total of 51 papers. Search criteria included STEM, STEM education, scientific education, and mathematics education. According to the study's findings, 23 papers and the majority of studies published in 2014 dealt with teaching-learning research.

A comparison of the interplay between inclusive policies, governmental initiatives, knowledge gaps from the literature, and possible long-term problems Australia and India might encounter in STEM education at the school level[3]. The question we posed was, "What are the key findings and best practices in STEM education that can grow the Strategic Partnership in a way that is beneficial to both parties?" The governments of both countries have made it plain that they are committed to utilizing the youth in their populations to train them as skilled workers in order to meet the need of industry in the future.

DATA SOURCE, TOOLS AND METHODOLOGY

Data source

The year-wise policy document is in PDF format; it should be extracted as a text file after that file.

Text analytics using Python

Text analysis is the process of separating texts into machine-readable facts. Text analysis is used to convert unstructured text into structured data. The procedure can be compared to cutting and dicing mountains of unstructured, heterogeneous documents into manageable, understandable data chunks. Text mining, text analytics, and information extraction are terms that are similar to text analysis.

Text analytics is another name for text mining. Finding patterns in large amounts of textual data is called text mining. While NLP processes the underlying metadata, text mining processes the text itself. Text mining is the process of determining word frequency counts, sentence length, and the presence or absence of particular terms. Text mining has various parts, one of which is natural language processing. NLP aids in identifying sentiment, locating sentence entities, and determining the blog/article category. Preprocessed data for text analytics is known as text mining. Information is categorized using statistical and machine learning algorithms in text analytics.

The term "natural language processing" (NLP) refers to a broad range of approaches for categorizing and measuring textual material by applying computational analytical methodologies. Cluster analysis is one of the NLP techniques that can assist researchers in examining their textual data.

NLTK

A strong Python module called NLTK offers a variety of effective natural language algorithms. It has a strong community, is open source, free, and extensively documented. The most popular algorithms, including part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition, are all included in NLTK. NLTK assists the computer with text analysis, preprocessing, and comprehension. Steps are includes as follows.

Tokenization

The initial stage in text analytics is tokenization. Tokenization is the process of dividing a text paragraph into smaller units, such as words or sentences. A token is a single object that serves as the foundation of a sentence or paragraph.

Sentence Tokenization

Sentence tokenizers break text paragraphs into sentences.

Word Tokenization

A word tokenizer breaks a text paragraph into words.

Stopwords

Stopwords are considered noise in the text. Text may contain stop words such as is, am, are, this, a, an, the, etc.
In NLTK, to remove stopwords, you need to create a list of stopwords and filter out your list of tokens from these words.

Stemming

Stemming is a linguistic normalizing process that strips words of their derived affixes or reduces them to their word roots. For instance, the words connection, connected, and linking can be reduced to the word "connect".

Lemmatization

When words are lemmatized, they are reduced to their basic words, which are lemmas in the proper sense. With the aid of vocabulary and morphological analysis, it converts root words. Typically, lemmatization is more complex than stemming. Stemmer analyzes a single word without taking context into account. As an illustration, the lemma of the word "better" is "good."

Visualization using Word Cloud

It is a novel form of visualizing text data that is generally employed to display keyword metadata on websites or to visualize text-free data. Tags are often single words, and the font size or color of each one indicates its significance. This style is helpful for finding a term alphabetically to establish its relative popularity as well as for rapidly identifying the most significant terms.

Frequency

In the first step type, size represents the number of times that tag has been applied to a single item. In the second, more commonly used type, size represents the number of items to which a tag has been applied, as a preparation of each tag’s popularity.

Significance

When compared to a background corpus, size can be used to indicate the importance of words and co-occurrences instead of frequency.

Categorization

In this third category, content pieces are categorized using tags. Greater tags in a tag cloud indicate that there are more content items in that category. Word clouds for extracting text into PDF files are shown in the accompanying figure.

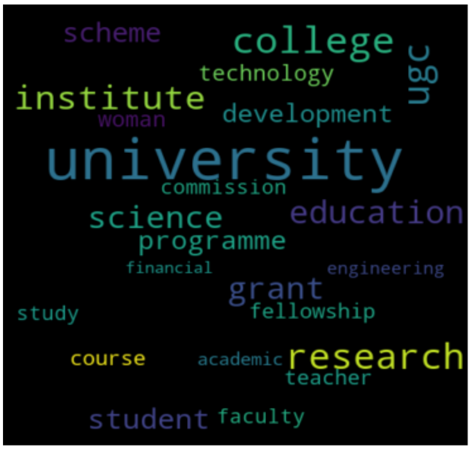

Figure 2: Diagram for Word Cloud

Network X

NetworkX is a Python module for building, modifying, and researching the composition, dynamics, and purposes of complicated networks. NetworkX offers a standard programming interface and graph implementation, an interface to existing numerical algorithms, a rapid development environment for collaborative, multidisciplinary projects, and tools for studying the structure and dynamics of social, biological, and infrastructure networks. It also makes it simple to work with large nonstandard data sets.

OBJECTIVES OF THE STUDY

While analyzing the objective of the roadmap of policy related to STEM fields, it is generally found that the government of India has given proper concern to STEM education in the later periods, which in turn is reflected in the implementation of the policies in the field

While trying to understand the relationship between keywords in STEM fields using text data visualization tools and techniques, the attempt was successful to some extent. It could bring out the prevalence of the keywords in the documents used for analysis.

In order to realize the objective of estimating the relationship between keywords in STEM fields using statistical tools and techniques, the general conclusion was made in such a manner that there is a progressive importance of STEM fields in the later periods of implementation.

HYPOTHESIS

The first hypothesis stated that “In India, the universities play a major role in upholding STEM fields” has proved positively using the words cloud and network that the universities, especially the public universities, play a critical role in delivering STEM education throughout the country.

The second hypothesis, “High incidence of STEM keywords found in later periods compared to earlier periods due its increasing importance in India’s policy” has clearly been proved on the grounds of various statistical analyses which in turn points to the fact that the policies on higher education have a keen role in changing the STEM fields in the country.

CONCLUSION

The study thus made a humble attempt to explore the different dimensions of policy documents related to STEM fields using text analytics and network analysis tools. The study is an eye opener for policymakers and those working in the field of higher education, as it gives a path to how to approach policy documents using the tools from computer-assisted machine learning techniques.

REFERENCES

[1]. STEM Education Research: Content Analysis Devkan Kaleci *, Özge Korkmaz Department of Computer and Instruction Technology, Faculty of Education, İnönü University, Turkey Universal Journal of Educational Research 6(11): 2404-2412, 2018 DOI: 10.13189/ujer.2018.061102

[2]. A Content Analysis Study About Stem Education Sevda Göktepe Yıldız, Yıldız Teknik University goktepe@yildiz.edu.tr, Ahmet Şükrü Özdemir Marmara University, ahmet.ozdemir@marmara.edu.tr

[3]. Perspectives of 'STEM education and policies' for the development of a skilled workforce in Australia and India Jyoti Sharma Department of Science & Technology, Ministry of Science & Technology, Government of India, Delhi, India Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia