P-ISSN: 2618-0723
E-ISSN: 2618-0731

NAAS Rating: 5.04
www.extensionjournal.com

International Journal of Agriculture Extension and Social Development

Volume 7; Issue 9; September 2024; Page No. 82-87

Received: 22-06-2024
Accepted: 26-07-2024

Indexed Journal
Peer Reviewed Journal

Harnessing big data analytics for agricultural extension: Applications of automated data extraction and sentiment analysis

Keesam Manasa, Sidharth S and MD Saifuddin

Ph.D. Research scholars, Dairy Extension Division, National Dairy Research Institute, Karnal, Haryana, India

DOI: https://doi.org/10.33545/26180723.2024.v7.i9b.1020

Corresponding Author: Keesam Manasa

Abstract

The integration of big data analytics in agricultural extension offers a transformative approach to improving research, decision-making and outreach. This paper explores two key methodologies—automated data extraction and sentiment analysis—demonstrating their applications in agricultural extension. Using Krishikosh, an online repository of agricultural theses, automated data extraction identified research trends in agricultural extension, revealing key topics such as adoption and knowledge transfer. Sentiment analysis, applied to Twitter data on the topic of cow slaughter during Eid, provided insights into public opinion, highlighting a predominantly negative sentiment. By employing advanced tools such as R software, both techniques efficiently processed large datasets, enabling more precise, data-driven insights. Despite challenges related to data access, privacy, and technical expertise, the study emphasizes the potential of big data analytics to enhance agricultural productivity, sustainability, and responsiveness. This paper serves as a practical guide for researchers and extension professionals to harness big data for more informed decision-making and tailored outreach strategies.

Keywords: Big data analytics, agricultural extension, automated data extraction, sentiment analysis, social media analysis, agricultural research trends

Introduction

The digital revolution has fundamentally transformed the way information is generated, collected, and analyzed across numerous sectors, including agriculture. In recent years, the emergence of big data has created new opportunities to enhance the efficiency and effectiveness of agricultural practices, particularly in the area of agricultural extension. Agricultural extension services aim to disseminate knowledge, technologies, and innovations to farmers and stakeholders to improve productivity, sustainability, and livelihood outcomes. In this context, big data analytics can serve as a powerful tool, enabling stakeholders to process large volumes of data, extract valuable insights, and make informed decisions that can significantly benefit the agricultural community. Big data is characterized by its volume (the sheer quantity of data), velocity (the speed at which data is generated and processed), and variety (the different types of data, such as structured, semi-structured, and unstructured data). In agricultural extension, data can come from a multitude of sources, including weather reports, soil sensors, GPS devices, satellite imagery, social media, online research repositories, and government databases. As the agriculture sector increasingly integrates digital technologies, the capacity to manage and interpret big data becomes crucial for addressing complex challenges such as climate change, food security, and resource management.

The application of big data analytics in agricultural extension allows for a shift from traditional, intuition-based decision-making to a more data-driven approach. For instance, machine learning algorithms and artificial intelligence (AI) can identify patterns in crop performance, predict disease outbreaks, optimize resource use, and suggest tailored farming techniques. By using these advanced analytical tools, extension officers and researchers can offer more precise and actionable recommendations to farmers, leading to improved agricultural practices and outcomes. Furthermore, big data analytics not only enhances agricultural productivity but also supports sustainable practices. By analyzing large datasets that track weather patterns, soil quality, and water usage, scientists can advise farmers on how to use resources more efficiently, reduce waste, and minimize environmental impact. This is particularly important in the context of global efforts to combat climate change and achieve sustainable development goals (SDGs) in agriculture. Despite its potential, the integration of big data in agricultural extension also poses several challenges. Issues related to data access, privacy, ownership, and the digital divide between developed and developing regions can hinder the widespread adoption of big data technologies. Moreover, the complexity of handling large datasets requires technical expertise, which may be lacking among extension workers and farmers alike. Therefore, addressing these challenges through capacity building and the development of user-friendly data tools is essential for maximizing the benefits of big data analytics in agriculture.

In this paper, we will explore how big data analytics can be applied within the field of agricultural extension. We will review various tools and techniques for automated data extraction, data analysis, and sentiment analysis, providing examples of how these methods can contribute to more effective and informed agricultural extension practices. Through the use of big data, agricultural extension can be transformed from a reactive, one-size-fits-all approach to a proactive, data-driven discipline that delivers tailored solutions to farmers in real time.

Materials and Methods

The materials used for applying big data analytics in agricultural extension research were selected based on their relevance to automated data extraction and sentiment analysis techniques. For data extraction, an online repository of agricultural research, such as Krishikosh, served as the primary source of data. This repository contains a vast collection of agricultural theses, providing a rich dataset for identifying key research trends in agricultural extension. Additionally, R software was utilized for extracting, cleaning, and analyzing the data, given its robust capabilities in text mining, data tidying, and visualization. In cases where users are unfamiliar with programming, alternative non-coding software tools can also be applied for similar purposes.

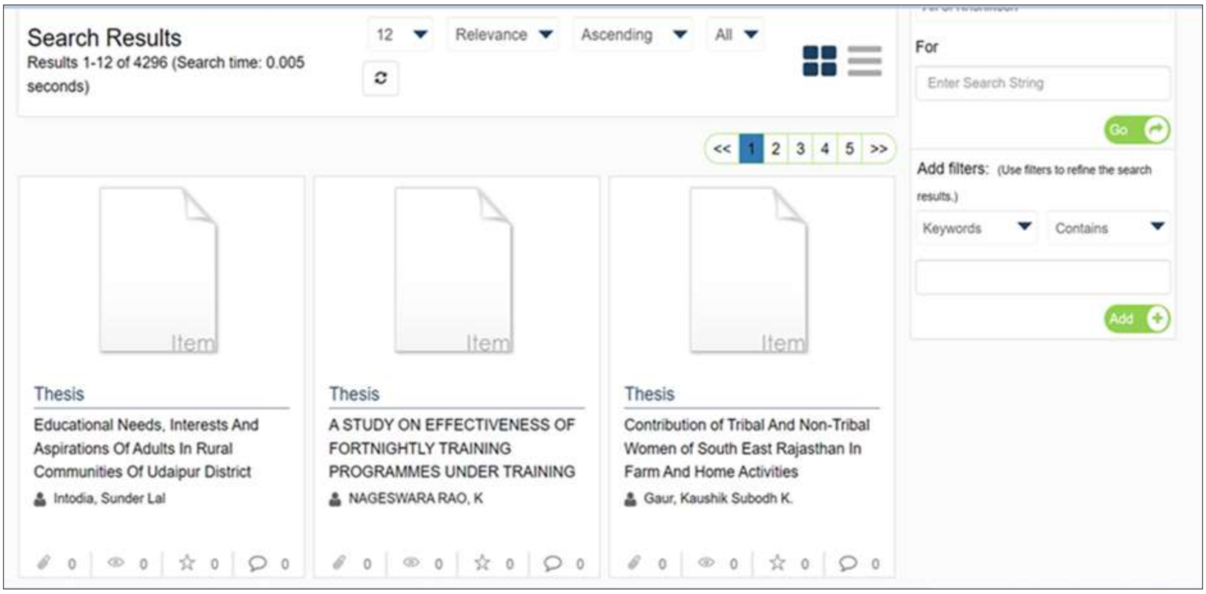
For sentiment analysis, data was sourced from online platforms such as Twitter, where public discourse around specific agricultural topics can be analyzed. Again, R software was employed to scrape and process the data, with packages like tidytext and ggplot2 aiding in sentiment classification and visualization. These tools enabled the systematic extraction of insights from large datasets and allowed for the categorization of sentiments (positive, negative, or neutral) related to agricultural issues. Overall, the combination of diverse data sources and advanced analytical software provided a comprehensive framework for conducting big data analytics in agricultural extension.

Results and Discussion

Automated Data Extraction

Step 1: Data Collection and Source Identification

The first step in applying big data analytics is identifying an appropriate data source. For agricultural extension, digital repositories such as *Krishikosh* offer a vast collection of research materials. In our example, *Krishikosh* housed over 4,296 theses on agricultural extension, providing a rich dataset to analyze research trends. Identifying such sources allows researchers to collect large volumes of data efficiently, which would otherwise be impossible to gather manualy



Source: https://krishikosh.egranth.ac.in

Fig 1: Search results for online thesises

Step 2: Extracting Keywords

Once the data source is determined, the next step is extracting relevant information. In this case, we focused on extracting the titles of theses related to agricultural extension from *Krishikosh*. R software was used for this task due to its robust data mining capabilities, which allow for automating the extraction process. By automating keyword extraction, large datasets can be processed quickly, enabling researchers to focus on more in-depth analysis rather than data collection.

Step 3: Tidying the Data

The extracted data often contains unnecessary elements such as prepositions and articles, which need to be removed to ensure clarity in the analysis. In this guide, we used text cleaning techniques to retain only meaningful words from the dataset. The cleaning process is essential for ensuring that the final analysis is based on relevant and useful information. R's tidytext and dplyr packages were used to accomplish this task, providing an efficient method for tidying up large datasets.

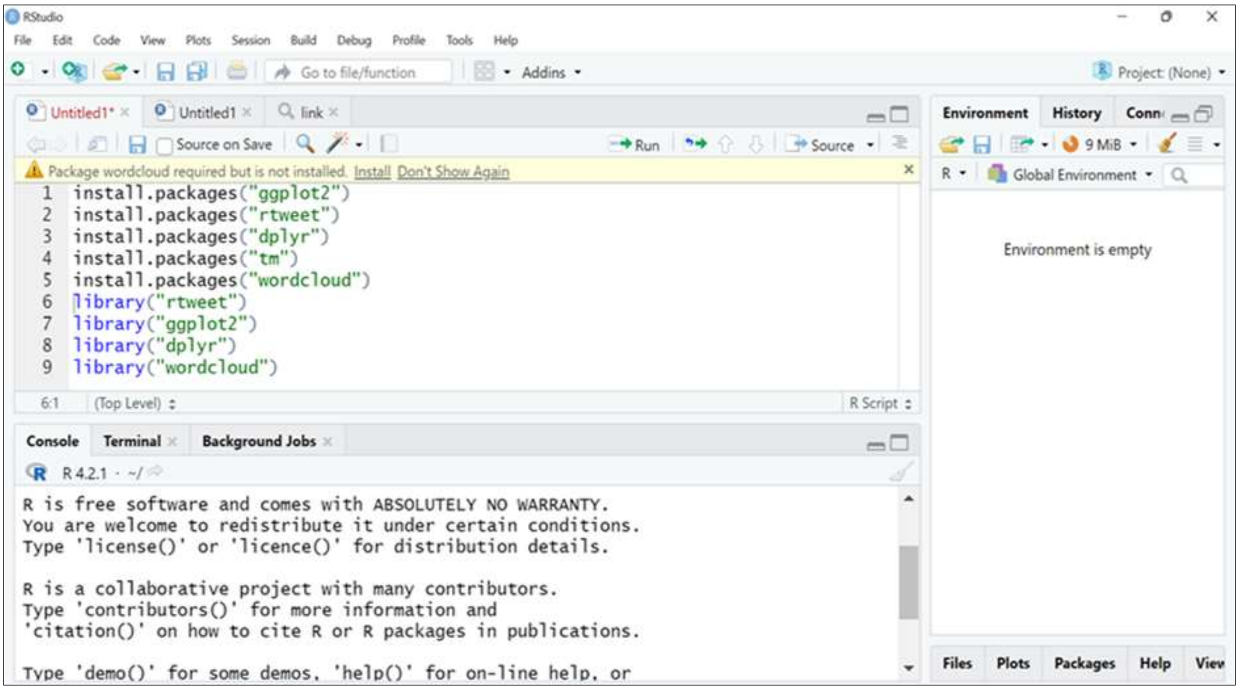


Fig 2: Data Extraction using R software

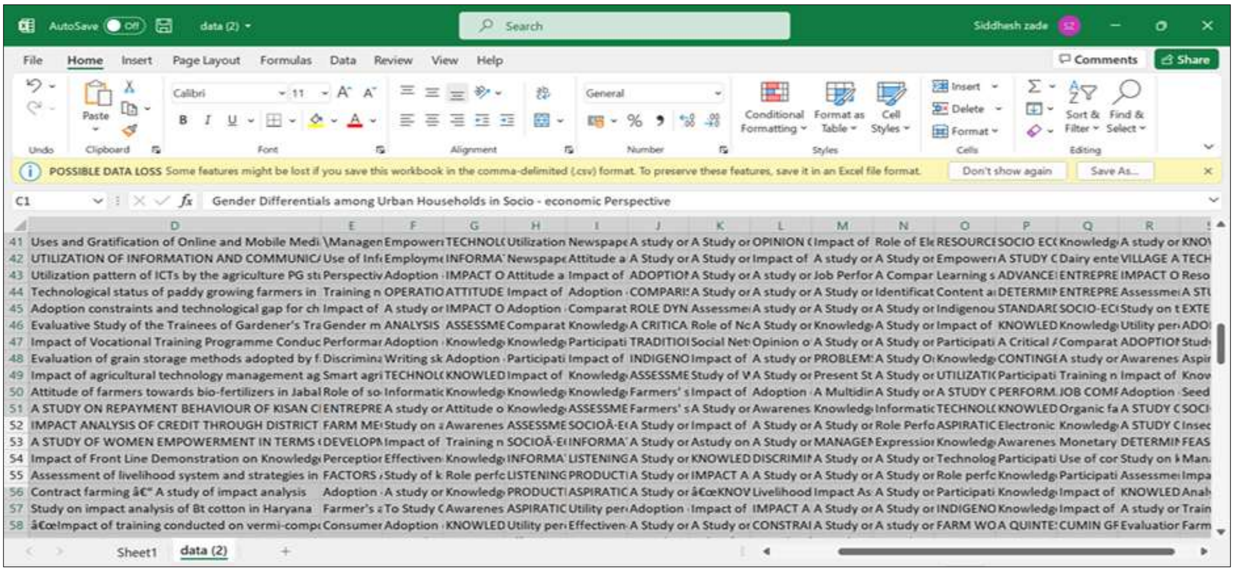


Fig 3: Tidying up of data

Step 4: Data Analysis and Visualization

Once the data is cleaned, the next step is analyzing and visualizing the information. A word cloud was generated to highlight the most common keywords in the thesis titles related to agricultural extension. Word clouds are an effective way to visualize the frequency of key terms in a dataset, where larger words represent more frequently occurring terms. In our example, words like “adoption” and “knowledge” were the most frequent, indicating that much of the research focused on these topics.

Additionally, bar graphs were used to present a clearer, more quantitative view of the frequency of research topics. Bar graphs provide a structured representation of the data, allowing researchers to easily identify dominant themes and areas that may require further investigation. In this case, the bar graph confirmed that studies on adoption and knowledge transfer were the most frequently researched topics within agricultural extension.

Through these steps, automated data extraction proves to be an efficient methodology for processing large amounts of academic and research data. By applying this approach, researchers can uncover significant patterns and trends in agricultural extension research, which would be difficult to detect manually.

Fig 4: Word cloud of most common topics researched

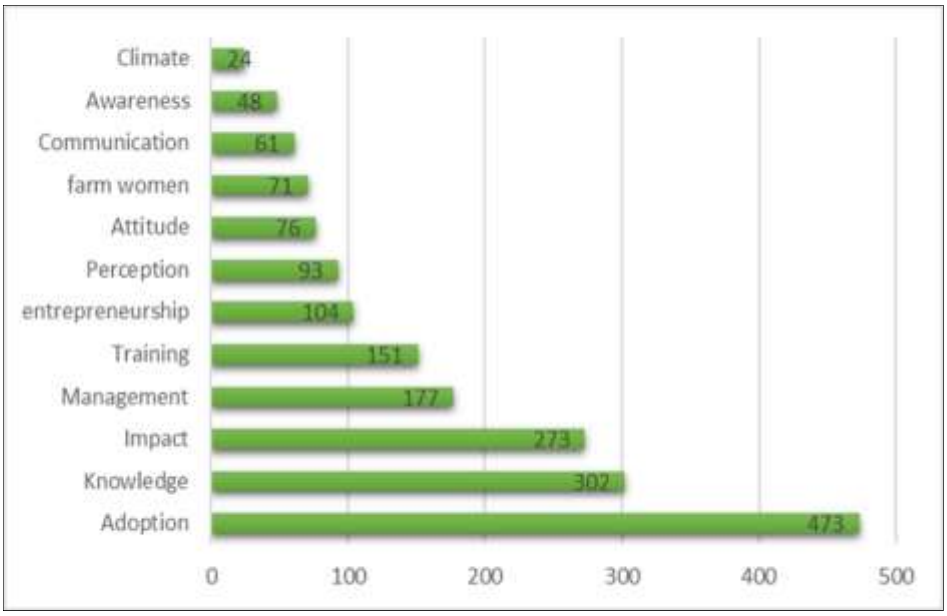


Fig 5: Bar graph of most common topics researched

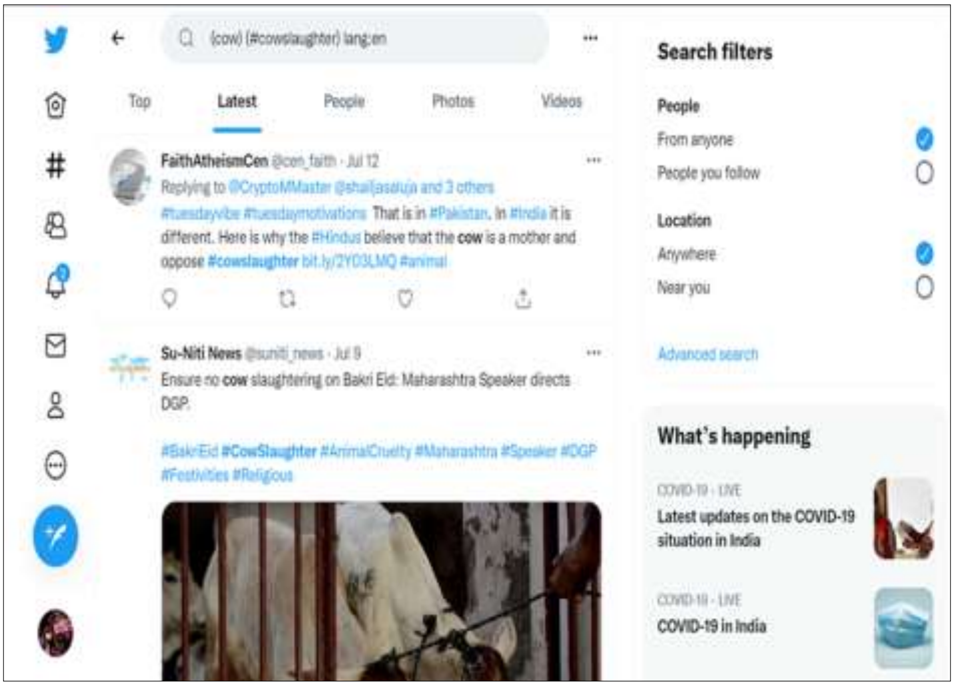


Fig 6: Tweets about cow slaughter. (Source: https://twitter.com)

Sentiment Analysis

Step 1: Data Extraction from social media: Sentiment analysis, another key method in big data analytics, begins with extracting relevant data from social media platforms, such as Twitter. In our example, we collected tweets related to *cow slaughter* during the Eid festival, a topic that was trending at the time. Social media data offers a unique opportunity to analyze public opinion in real time, providing valuable insights into how people feel about specific agricultural issues. Using R software, we scraped tweets that contained the keyword "cow slaughter." This automated process enabled the collection of large volumes of data quickly and efficiently. Similar approaches can be applied to other agricultural topics, allowing researchers to track public sentiment on a wide range of issues.

Step 2: Data Cleaning: As with the automated data extraction process, the collected tweets need to be cleaned before analysis. Social media data often includes irrelevant elements such as hashtags, links, and common words that do not contribute to the sentiment analysis. Removing these elements ensures that the analysis focuses solely on meaningful content. In this example, R's tidytext package was used to filter out unnecessary parts, leaving a dataset of relevant words for further analysis.

Step 3: Generating a Word Cloud: Once the data is cleaned, it can be visualized using a word cloud. Similar to the earlier example, a word cloud was generated to display the most frequently mentioned terms in the tweets about *cow slaughter*. In this case, words related to the debate over *cow slaughter* were highlighted, offering insight into the dominant themes in the conversation. Word clouds allow for a quick and intuitive understanding of the key terms people are using in social media discussions.

Fig 7: Word cloud of sentiments of twitter data

Step 4: Sentiment Classification

The final step in sentiment analysis is classifying the words into categories based on the sentiment they express-positive, neutral, or negative. In our example, sentiment analysis revealed that the majority of words related to *cow slaughter* expressed neutral or negative sentiments. Using R's sentiment analysis tools, we assigned a sentiment score to each word and calculated the overall sentiment of the dataset. The results indicated that public sentiment toward *cow slaughter* during the Eid festival was predominantly negative. Sentiment classification is a powerful tool for understanding how the public feels about specific agricultural issues. By tracking public opinion through sentiment analysis, agricultural extension services can tailor their messaging and outreach efforts to address public concerns more effectively.

Fig 8: Sentiment Analysis of Twitter data

Conclusion

The integration of big data analytics into agricultural extension offers a transformative approach to enhancing the effectiveness and precision of research and outreach efforts. By applying automated data extraction and sentiment analysis methodologies, researchers and extension professionals can unlock valuable insights from vast datasets, ranging from academic repositories to real-time social media discourse. The use of advanced tools such as R software enables efficient data collection, cleaning, and visualization, allowing for more informed decision-making and targeted interventions. Automated data extraction, as demonstrated with the *Krishikosh* repository, provides an efficient means to analyze research trends within agricultural extension, highlighting key areas of focus and opportunities for further investigation. By leveraging techniques such as keyword extraction and data visualization, researchers can identify patterns and gaps in the literature that would otherwise be difficult to detect manually.

Similarly, sentiment analysis of social media platforms like Twitter offers a powerful way to gauge public opinion on critical agricultural issues. In this study, the analysis of tweets related to *cow slaughter* provided a clear understanding of the negative and neutral sentiments surrounding this topic during the Eid festival. This type of analysis helps extension services tailor their communication strategies, ensuring that they address public concerns more effectively and foster a stronger connection with the agricultural community. While the potential of big data analytics in agricultural extension is significant, challenges such as data access, privacy concerns, and the need for technical expertise must be addressed to maximize its impact. Capacity building and the development of user-friendly tools are essential to ensure that both researchers and extension workers can fully leverage the benefits of big data.

In conclusion, the adoption of big data analytics represents a critical shift from traditional, intuition-based practices to a more proactive, data-driven approach in agricultural extension. By harnessing the power of big data, extension services can provide more accurate, timely, and customized solutions to farmers, ultimately improving agricultural productivity, sustainability, and resilience in the face of global challenges such as climate change and food security. The methodologies discussed in this paper serve as a guide for how researchers and practitioners can effectively apply these tools, paving the way for a more modern and responsive agricultural extension system.

References

1. Eastwood C, Klerkx L, Ayre M, Dela Rue B. Managing socio-ethical challenges in the development of smart farming: From a fragmented to an inclusive socio-ethical approach to responsible innovation in agriculture. Geoforum. 2019;99:95-103.
2. Jayashankar P, Nilakanta S, Johnston WJ, Gill P, Burres R. IoT adoption in agriculture: The role of trust, perceived value and risk. J Bus Ind Mark. 2018;33(6):804-821.
3. Kamilaris A, Kartakoullis A, Prenafeta-Boldú FX. A review on the practice of big data analysis in agriculture. Comput Electron Agric. 2017;143:23-37.
4. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. Comput Electron Agric. 2018;147:70-90.
5. Lambrecht I, Vanlauwe B, Maertens M. Agricultural extension in Eastern Democratic Republic of Congo: Does gender matter? Eur Rev Agric Econ. 2016;43(5):841-874.
6. Sonka S. Big data and the ag sector: More than lots of numbers. Int Food Agribusiness Manag Rev. 2016;19(A):11-20.
7. Wolfert S, Ge L, Verdouw C, Bogaardt MJ. Big data in smart farming – A review. Agric Syst. 2017;153:69-80.