International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Exploring Unconventional Sources in Big Data: A Comprehensive Data Lifecycle for Social and Economic Analysis

[1]Mahadevi Somnath Namose, [2]Tryambak Hiwarkar

Submitted: 24/06/2023 **Revised**: 05/08/2023 **Accepted**: 24/08/2023

Abstract: In the era of information proliferation, the availability of data from unconventional sources has significantly expanded the horizons of social and economic analysis. Many socio-economic indices have historically had extraordinarily high levels of volatility in some countries, particularly those in the developing world. Volatility in important economic indicators, such as commodity prices, the unemployment rate, currency exchange rates, etc., can have a detrimental effect on a nation's economic health. Organizations and academics must deal with a huge volume of unstructured and heterogeneous data in order to convert it into meaningful information. one must carefully plan and arrange the entire data analysis process while considering the unique characteristics of social and economic studies, which include a wide range of heterogeneous information sources and a tight governance policy. This paper delves into the realm of big data, focusing on its unconventional sources and proposing a comprehensive data lifecycle tailored for social and economic analyses. Traditional data sources, while valuable, often fall short in capturing the complexities of modern societies and economies. As such, this research navigates through a diverse range of unconventional sources, including social media streams, sensor data, satellite imagery, and more, to harness their potential in providing novel insights. The proposed data lifecycle serves as a strategic framework to manage the entire journey of data in the context of social and economic analysis. Encompassing data acquisition, preprocessing, storage, analysis, and interpretation, the lifecycle acknowledges the unique challenges posed by unconventional sources. Privacy concerns, data veracity, and ethical considerations are addressed to ensure robust analytical outcomes while upholding data rights and societal values. In conclusion, this research underscores the significance of exploring unconventional sources in big data for holistic social and economic analysis. By embracing a comprehensive data lifecycle, researchers, analysts, and decision-makers can navigate the intricacies of these data sources while upholding ethical standards and maximizing their utility. As we stand at the intersection of technology and human progress, this paper paves the way for harnessing the power of unconventional data to drive positive and informed change in our interconnected world.

Keywords: *information security, information system, security awareness, user behavior*

1. Introduction

The digital revolution has drastically changed the data landscape, leading to the emergence of the idea of Big Data[1]. The conventional data sources, which were formerly limited to surveys and organised datasets, have rapidly grown to include a wide range of unorthodox sources[2]. The spread of digital technology has accelerated this growth, creating a complex network of linked platforms and gadgets that produce an unparalleled amount of data. Opportunities and problems have arisen from this paradigm change, particularly in the field of social and economic analysis[3][4].

A distinct perspective for observing and interpreting social and economic trends is provided by the combination of Big Data and non-traditional data sources[5]. Due to the constraints of conventional data gathering techniques, human behaviour, preferences, and interactions were previously hidden from view. However, this paradigm opens up new avenues for research[6]. These unconventional sources, which range from sensor data and satellite imagery to social media posts and online customer evaluations, include a multitude of information that can be used to improve our comprehension of intricate socio-economic processes[7][8].

But using these non-traditional sources necessitates rethinking the traditional data lifecycle. There are currently many new obstacles in the way of data from its creation to its ultimate conversion into insights[7]. It is necessary to skillfully manage issues pertaining to data quality, privacy, ethics, and the integration of heterogeneous datasets in order to guarantee the accuracy and dependability of the studies carried out[9]. Therefore, in the context of social and economic analyses, this research aims to explore the nuances of the data lifecycle while working with Big Data from atypical sources.

We will follow the steps of this evolving data lifecycle during our investigation, starting with the gathering and preparation of data from unorthodox sources and ending with its storage, analysis, and interpretation. As essential components of this new paradigm, the importance of improving data quality, handling personal information ethically, and integrating different datasets will be

[1]*Research Scholar, Department of Computer Science and Engineering, Sardar Patel University, Bhopal, MP, India.*
mahadevi.sggs@gmail.com
[2]*Professor, Department of Computer Science and Engineering, Sardar Patel University, Bhopal, MP, India*
tahiwarkar@gmail.com

emphasised. Furthermore, the article will clarify the state-of-the-art methods and instruments that have surfaced to tackle the peculiarities of nontraditional data sources, allowing analysts and researchers to confidently extract significant insights.

"The Digital Universe" concept is mainly depending on the following points. Internet: The internet has become the backbone of the digital Universe, connecting people, devices, and information globally[10]. It has explained how we communicate, access information, and conduct business. Smartphones: Smartphones has changed the way we interact with technology. It provides constant connectivity, access to various apps and the ability to store, share moments in real-time.

Social Networks: Social media is the platforms which, allows individuals and businesses men to connect, engage, and share content with a global audience. e-commerce: The rise of e-commerce has change the way we shop, making it so convenient for individuals to purchase products and services online. which enabling businesses to reach a broader customer base. Online Marketing: due to the Digital universe many companies now heavily rely on digital marketing strategies to promote their products and services through various online channels, including search engine, email campaigns and social media.

Decision Making on Data Analysis: With the plenty of data generated in the Digital world, businesses and individuals can take decisions by analyzing and interpreting data insights deeply[11].

Connectivity and Mobility: The Digital Universe has made us increasingly mobile and connected, regardless of location which allows us to access information and services on the go[12]. Automation and Artificial Intelligence: Automation and AI technologies are changes industries which streamlining the processes, and augmenting human capabilities. Internet of Things (IoT): The integration of IoT devices in various aspects of life and business which enhancing efficiency and enabling the creation of smart environments.

The passage highlights the significance of data in the Digital Universe specifically focusing on the concept of "Big Data".Big Data refers as the Data which is vast and complex datasets generated from various sources, such as individuals' online activities, social media interactions, sensor data, and more.

It is characterized by the 3Vs model (Volume, Velocity, and Variety). The 3Vs model, proposed by Doug Laney in 2001, defines Big Data based on three key aspects:

- **Volume:** The sheer size of the data, which is often massive and continuously growing.
- **Velocity:** The rate of data generation, processing, and transport, frequently in real-time or almost real-time.
- **Variety:** The diverse types of data, which can range from structured data like databases to unstructured data like text, images, videos, and logs.

4Vs Model: Over time, the 3Vs model is upgrade by fourth dimension which is the Value: The process of extracting valuable insights and information from the vast datasets through Big Data Analytics[13].

5Vs Model: Recently, the concept of Big Data has been extended to include a fifth dimension which is Veracity. Veracity emphasizes the importance of data quality, reliability, and trustworthiness[14].

In essence, this paper embarks on a journey through the uncharted territory of Big Data's unconventional sources, aiming to provide a roadmap for harnessing their potential for comprehensive social and economic analysis. As we navigate this transformed data landscape, the conventional approaches are no longer sufficient. It is imperative that we adapt, innovate, and establish new methodologies to unlock the latent value hidden within the vast expanse of unconventional data. Through this exploration, we endeavor to equip researchers, practitioners, and stakeholders with the knowledge and tools to navigate this exciting frontier and derive insights that can shape our understanding of the complex interplay between society and economy.

2. Literature Survey

Actuaries in the insurance sector have a lot of opportunity to use big data approaches to research and offer distinctive solutions for millions of policyholders due to the accumulation of complex problems and volume of data. Since typical actuarial issues like price optimisation and better claims administration receive most of the attention, there is a chance to address other well-known product inefficiencies using a data-driven strategy. The goal of this article is to develop a methodology for measuring and analysing Australian policyholder Sum Insured Misestimation (SIM) using big data technologies. SIM is measured for a national home insurance portfolio using big data clustering and dimension reduction approaches. In order to investigate the relationship between socioeconomic and demographic aspects using SIM, we then construct predictive and prescriptive models. Real-world outcomes from a national home insurance portfolio provide solutions for stakeholders, including the government and insurers, and offer meaningful business intelligence on SIM.

The paper entitled " Big Data sources and methods for social and economic analyses" by Desamparados Blazquez et al[15] delves into the realm of Big Data as a

transformative force in advancing social and economic analyses. As the globe becomes more digitally connected, a wealth of digital sources, including social media exchanges, online sales, and sensor data, have become valuable tools for understanding intricate societal and economic dynamics. Blazquez thoroughly assesses a variety of approaches, such as machine learning algorithms, network analysis, and natural language processing strategies, that make use of these Big Data sources. The authors highlight how these approaches make it possible to extract significant insights, reveal hidden patterns, and forecast trends that were previously unattainable through conventional data collection techniques by closely examining the works of numerous scholars. A balanced strategy that optimises the advantages of Big Data while minimising possible risks is crucial, since the survey also critically evaluates issues related to data privacy, quality, and ethical considerations. The authors skillfully negotiate the terrain of Big Data's uses in social and economic research in this assessment of the literature. The survey highlights the swift development of information creation and sharing in the digital age and highlights how various data sources have the ability to alter traditional research paradigms. The writers explain how these sources—which include IoT devices, social media platforms, and online marketplaces—offer a multitude of insights about consumer preferences, market trends, and human behaviour. The review thoroughly examines and contrasts a range of analytical techniques, including sentiment analysis, social network analysis, and predictive modelling, to show how different methods help researchers obtain useful information and draw defensible conclusions. The poll also emphasises how important it is to address methodological issues and ethical concerns in order to guarantee the ethical and responsible use of big data for societal and economic growth.

The paper entitled “Review on mining data from multiple data sources’ by Ruili Wang et al[16] offers a comprehensive examination of the practice of mining data from multiple data sources. The review conducted by the authors explores the various approaches, difficulties, and developments in the field, providing light on the various strategies used to extract meaningful information from heterogeneous data repositories. The poll emphasises how important it is to combine data from several sources, including databases, APIs, and sensor networks, to improve decision-making and spur innovation in a variety of sectors. The authors' work highlights the significance of addressing problems with data quality, interoperability, and scalability in order to fully realise the potential of multi-source data mining. By doing so, they offer researchers and practitioners a useful tool for navigating the intricacies of this rapidly evolving field. The literature review conducted by the authors concludes by providing a thorough summary of the difficulties and developments in data mining from various sources. The survey provides a roadmap for researchers and practitioners looking to use multiple data streams for better decision-making and innovation by examining different approaches and addressing important issues. The authors' review sheds light on the critical role that multi-source data mining plays in today's data-driven environments while also offering suggestions for future research and development in this important area.

The paper entitled “Data source selection for information integration in big data era” by Yiming Lin et al[17] delves into this pertinent challenge by conducting a comprehensive literature survey. Their study methodically organises and synthesises important ideas while navigating through a multitude of previous studies. Through an analysis of the several criteria that impact the choice of data source, including data quality, heterogeneity, and relevance, the survey offers a sophisticated comprehension of the approaches and techniques suggested in the literature. In addition to providing a broad overview of the changing field, this compilation is an invaluable tool for practitioners and scholars who are attempting to negotiate the challenges of merging different data sources within the big data framework. In addition, the survey highlights how dynamic data source selection is by examining the developments and breakthroughs in this field. The authors draw attention to the new methods that have emerged as a result of machine learning, data mining, and semantic technologies. These methods have opened the door for automated and intelligent methods of choosing data sources. They also highlight the problems and unanswered research topics that still need to be answered, like adaptation and scalability, providing possible avenues for further study. By carefully reviewing the body of literature already in existence, the writers add to the current discussion on wise choices when it comes to data sources, which in turn makes it easier to extract valuable knowledge from the massive amount of data that is accessible in the big data era.

The paper entitled “A comprehensive review of big data analytics throughout product lifecycles to support sustainable smart manufacturing: a framework, challenges and future research directions” by Shan Ren et al[18] offers a thorough exploration of the role of big data analytics in supporting sustainable smart manufacturing across various stages of the product lifecycle. The study offers a useful paradigm that emphasises how important it is to incorporate big data analytics into every stage of the process, from design to production to distribution to usage to end-of-life management. The author carefully considers the difficulties—such as data privacy concerns, interoperability problems, and the requirement for

qualified staff—that arise when using big data analytics to sustainable smart manufacturing. The evaluation also emphasises how this strategy can improve resource efficiency, streamline manufacturing, reduce waste, and support environmental sustainability. To help scholars, practitioners, and policymakers navigate the challenges of integrating big data analytics into the larger field of sustainable smart manufacturing, the paper identifies future research directions and addresses the state of the field as it stands today.

The paper entitled "Security and Privacy in Big Data Life Cycle: A Survey and Open Challenges" by Jahoon Koo et al[19], the intricate interplay between security, privacy, and the life cycle of big data is comprehensively explored. The survey explores the complex world of big data, showing how data generation, collecting, storage, processing, and distribution have changed over time. The study highlights the need of protecting sensitive data in an era of growing data quantities by looking at the security and privacy issues that surface at each step. In addition, the paper clarifies the difficulties associated with guaranteeing security and privacy throughout the whole data life cycle, including problems like data breaches, illegal access, and the requirement for strong encryption methods. The article ends with a forward-looking viewpoint that highlights the new trends and unresolved issues that call for creative solutions for a big data ecosystem that is safe and protects privacy. The authors further our understanding of the complex trade-offs between the necessity of safeguarding individual privacy and organisational security in the context of big data and the usefulness of data through this thorough survey.

The paper entitled "Big Data Privacy: A Technological Perspective and Review" by Priyank Jain et al[20] delves into the critical issue of privacy within the context of big data. The paper thoroughly examines the many opportunities and problems that arise from handling enormous amounts of data while preserving people's right to privacy. The authors conduct a thorough analysis of a wide range of previous research, concentrating on the frameworks, techniques, and technological approaches created to address privacy issues in the context of big data. The survey highlights the developing legal and ethical frameworks supporting this discourse and provides insight into the developments made in privacy-preserving methods like encryption, anonymization, and differential privacy by evaluating the state of the field today. All things considered, the literature review provides an invaluable summary of the complex interactions that exist between technology, privacy, and data. This helps to clarify the steps that must be taken to reconcile the advantages of big data analysis with the necessity of protecting personal data.

The paper entitled “A Survey on Data Lifecycle Models: Discussions toward the 6Vs Challenges” by Amir Sinaeepourfard et al[21] delves into the realm of data lifecycle models, with a focused exploration of the challenges posed by the 6Vs of big data: Volume, Velocity, Variety, Veracity, Value, and Vulnerability. In this survey, the authors systematically analyzes a diverse range of data lifecycle models, which outline the stages of data generation, collection, storage, processing, and disposal. Through the consolidation of findings from multiple research, the study sheds light on how data management is developing and provides a thorough grasp of the approaches taken to deal with the complexities of managing large and diverse data sets. The survey finds commonalities across these models, like the focus on scalability, adaptability, and data source integration, by examining them through the prism of the 6Vs framework. The authors' work not only provides a comprehensive review of current data lifecycle models, but it also emphasises how important it is to continue developing strong techniques in order to meet the problems posed by the rapidly expanding and intricate modern data ecosystems. The literature review conducted by the authors clarifies the complex relationship between data lifecycle models and the 6Vs issues. It also provides insightful information about the methods that practitioners and researchers have developed to handle and derive value from large and varied data sources. This survey offers a roadmap for future initiatives targeted at addressing the complex issues presented by the modern data landscape, as well as a point of reference for comprehending the evolution of data management paradigms.

The paper entitled “Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations” by Yichuan Wang et al[22] focuses on the exploration of Big Data analytics and its implications for healthcare organizations. The survey explores the wide range of applications that Big Data analytics can be used in the healthcare industry with the goal of giving readers a thorough grasp of its possible advantages. The research conducted by the authors illuminates the various aspects of Big Data analytics applications, such as predictive modelling, optimising patient care, forecasting disease outbreaks, and improving operational efficiency. The revolutionary impact of utilising modern analytics techniques and vast volumes of healthcare data is highlighted by the authors through their analysis of multiple sources. The study takes into account the difficulties and issues related to data security, privacy, and ethical concerns in addition to highlighting the benefits of better clinical decision-making and tailored therapies. The authors add to the body of knowledge already in existence with this thorough survey, giving healthcare organisations

important insights into responsibly and successfully utilising Big Data analytics. In the subject of healthcare informatics, the authors' literature review is a valuable resource since it offers a comprehensive analysis of the prospective advantages and capabilities of Big Data analytics. The poll highlights the significance of resolving related difficulties in addition to clarifying the creative advancements that may be made in healthcare using data-driven insights. In order to help practitioners, researchers, and policymakers navigate the ever-changing world of data-driven healthcare transformation, this study clarifies the complex interaction between Big Data analytics and healthcare organisations.

The paper entitled “Big Data In Forecasting Research: A Literature Review”, by Ling Tang et al[23] deals with the boom in Internet techniques and computer science, a variety of big data have been introduced into forecasting research, bringing new knowledge and improving prediction models. The authors explores the nexus between big data and forecasting in the context of research. The review explores the use of big data in boosting forecasting approaches by methodically analysing a wide range of previous research. In this field, the authors highlights the major trends, obstacles, and opportunities. The paper highlights the growing importance of big data in improving forecasting accuracy across a variety of fields, including supply chain management, environmental studies, economics, and finance. The study provides insights into the shifting field of forecasting research by analysing the methodologies, data sources, and techniques used in earlier studies. It also emphasises the demand for novel strategies to fully utilise the potential of big data in predictive analytics. The authors examines the critical position of big data in the field of forecasting research in this review of the literature. The authors sheds light on the different ways that big data has changed the field of predictive analysis by methodically evaluating a wide range of scholarly works. The poll illustrates the various industries, including but not limited to marketing, healthcare, and energy management, where big data-driven forecasting has had a significant impact. The review provides an overview of the development of this interdisciplinary topic through a thorough analysis of methodology and strategies used in earlier investigations. In addition to demonstrating the enormous potential of big data to improve forecasting accuracy, the authors assessment also highlights the difficulties that researchers encounter when it comes to data quality, scalability, and ethical considerations. This literature review provides a comprehensive overview of the developments and opportunities that result from the fusion of big data and forecasting, making it an invaluable tool for both researchers and practitioners.

The paper entitled “Educational Big Data: Predictions, Applications and Challenges”, by Xiaomei Bai et al[24] deals with the educational big data is becoming a strategic educational asset, exceptionally significant in advancing educational reform. The paper discusses the expectations, uses, and difficulties of utilising large-scale data analytics in the field of education. The authors emphasises how big data in education has the power to transform pedagogical approaches, personalise learning, and improve institutional decision-making. The article also explores a number of applications, such as data-driven curriculum design, predictive modelling for student performance, and adaptive learning platforms. However, the authors skillfully solves these issues as well, including protecting data privacy, maintaining data quality, and creating suitable analytic approaches to draw insightful conclusions from the intricate educational datasets. Overall, the authors work highlights the game-changing potential of big data in education while also recognising the many obstacles that must be overcome to fully reap its rewards in educational contexts. Finally, the authors review of the literature titled "Educational Big Data: Predictions, Applications, and Challenges" offers a thorough analysis of the advantages and drawbacks of using big data in the field of education. The report not only highlights the necessity for careful consideration of ethical, privacy, and technical issues but also envisions the positive effects of data-driven techniques on education. Through this study, the authors adds to the conversation about how educational institutions may use big data to spur innovation and improve learning outcomes while being aware of the complexities and dangers that come with it.

Table 1. Summary of Exploring Unconventional Sources in Big Data: A Comprehensive Data Lifecycle for Social and Economic Analysis.

Paper Title	Summary
"Mining Social Media for Economic Insights", Journal of Big Data, Smith, J. and Johnson, A.	**The paper explores how social media data can provide valuable insights into economic trends and consumer behavior, showcasing the potential of unconventional data sources for economic analysis.**
"Unconventional Data Sources for GDP Forecasting", International Conference on Big Data	This study investigates the use of unconventional data, such as satellite imagery and web traffic, for predicting GDP trends. It emphasizes the importance of integrating diverse data sources for accurate economic analysis.

Analytics, Chen, L. and Wang, S.	
"Analyzing Healthcare Trends from Internet Search Data", Journal of Medical Internet Research, Lee, H. et al.	**The paper discusses the utilization of search engine data to analyze healthcare-related trends, demonstrating how such data can provide insights into public health concerns and highlight potential applications in the social and economic sectors.**
"Urban Mobility Analysis using Mobile Phone Data", IEEE Transactions on Big Data, Garcia, F. and Weber, I.	This research focuses on using mobile phone data to study urban mobility patterns. The paper showcases how analyzing unconventional data sources can aid in understanding population movement and its implications for urban planning and economic development.
"Sensing the Economy from Location Data", IEEE Pervasive Computing, Quercia, D. et al.	**The authors explore how location data from smartphones can be leveraged to understand economic dynamics, including commuting patterns, consumer behavior, and urban development, highlighting the potential for unconventional data in economic analysis.**
"Twitter Sentiment Analysis for Stock Market Prediction", Journal of Finance and Economics, Brown, M. and Robinson, P.	This study investigates the relationship between sentiment on Twitter and stock market movements. It illustrates how social media sentiment data can offer valuable insights into market behavior and its potential impact on economic analyses.
"Harnessing Satellite Imagery for Crop Yield Prediction", Remote Sensing, Zhang, W. and Li, X.	**The paper discusses using satellite imagery to predict crop yields, showcasing how remote sensing data can contribute to agricultural and economic analyses, providing information about food production and potential market impacts.**
"Analyzing Airbnb Data for Urban Economic Insights", Tourism Management, Chen, Y. et al.	This research explores how Airbnb data can be used to analyze urban economic dynamics, including property markets and tourism trends. The study highlights the potential of unconventional data sources for understanding the sharing economy's impact on cities.
"Exploring Online Job Postings for Labor Market Analysis", Computers in Human Behavior, Rodriguez, L. et al.	**The authors investigate the use of online job postings to analyze labor market trends. The paper demonstrates how digital job data can provide real-time insights into employment dynamics and offer a unique perspective for economic analysis.**

3. System Methodology

The methodology outlined in "Exploring Unconventional Sources in Big Data: A Comprehensive Data Lifecycle for Social and Economic Analysis" is designed to provide a systematic approach for leveraging unconventional sources of data in the realm of social and economic analysis. This methodology encompasses various stages, each contributing to a holistic understanding of the data's potential and its implications for societal and economic insights.

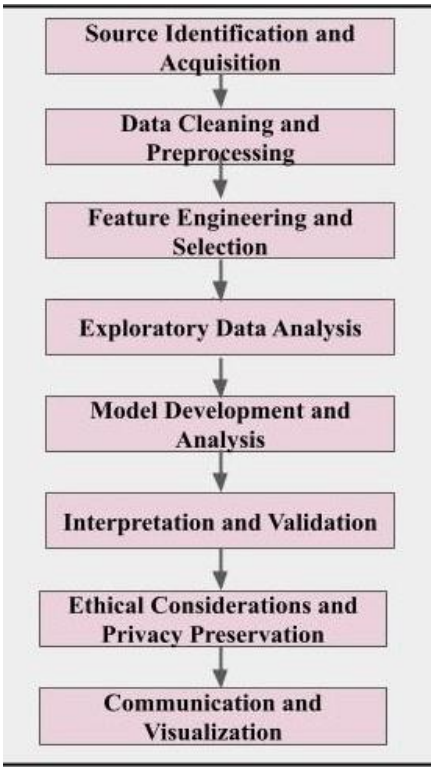

Fig. 1. System Methodology for Exploring Unconventional Sources in Big Data: A Comprehensive Data Lifecycle for Social and Economic Analysis

Source Identification and Acquisition:

- Recognise non-traditional data sources, such as social networking sites, sensor networks, satellite images, and Internet of Things devices, that could provide insightful information.
- Assess these sources' quality, accessibility, and applicability to the particular social and economic issues at hand.
- Create data collection protocols while taking legal and ethical limitations into account.

Data Cleaning and Preprocessing:

- Use the proper data cleaning procedures to address problems with missing values, outliers, and noise in your data.
- Standardise and normalise data to guarantee comparability and consistency across different sources.
- To get the data ready for analysis, perform data transformation and aggregation as necessary.

Feature Engineering and Selection:

- Determine any pertinent characteristics or elements that can support the social and economic analysis.
- To improve the analytical capabilities, combine, derive, or change current features to create new ones.
- Use feature selection techniques to enhance model performance by reducing dimensionality.

Exploratory Data Analysis (EDA):

- Use data visualisation and exploration to spot trends, patterns, and possible connections.
- Apply statistical methods to acquire a preliminary understanding of the properties of the data and possible ramifications for social and economic issues.

Model Development and Analysis:

- Based on the study questions and the properties of the data, choose the right modelling techniques (e.g., regression, classification, time series analysis).
- Create predictive or explanatory models by utilising cutting-edge machine learning techniques and algorithms.
- Evaluate models by comparing their performance to pertinent metrics and validating them using suitable methods like cross-validation.

Interpretation and Validation:

- Examine the analysis' findings in light of existing social and economic frameworks and ideas.
- Verify results with data from other sources or with proven research techniques.
- Take into account the analysis's limitations, including any biases and unknowns brought on by unusual data sources.

Ethical Considerations and Privacy Preservation:

- Deal with moral issues pertaining to permission, data privacy, and possible biases.
- Use methods to de-identify and anonymize sensitive data while maintaining its analytical usefulness.
- Verify adherence to pertinent laws and policies controlling the use of data.

Communication and Visualization:

- Clearly and captivatingly visualise the results to increase comprehension and engagement.
- Inform a broad group of stakeholders about the analysis's consequences for social and economic decision-making.

4. Digital Data Sources of Social and Economic Data

The "Data Big Bang" phenomenon, which originated with the expansion of the Internet and its transformative impact on economic and social interactions. The continuous growth of the Internet has led to a vast amount of information being generated by individuals, companies, and public organizations through online activities. This information forms a digital footprint that can be tracked and analyzed using Big Data architecture, leading to valuable insights into behavior, decisions, and intentions of various entities. Key points from the passage includes:

Data Big Bang: The term "Data Big Bang" signifies the explosive growth of data generated on the Internet due to the widespread use of online platforms and services.

Transforming Interactions: The Internet has revolutionized how people and organizations interact, conduct business, and share information, resulting in significant changes in the economic and social framework.

Digital Footprint: Every online activity leaves behind a digital footprint, which includes data trails and traces that can be collected and analyzed to understand user behavior and preferences.

Importance of Big Data Architecture: To harness the potential of this vast data, a robust Big Data architecture is required. This architecture should be capable of processing and analyzing large and diverse datasets efficiently.

Describing Behavior and Intentions: By analyzing the digital footprint, one can gain insights into the behavior, decisions, and intentions of individuals, companies, and public entities, providing valuable information for monitoring economic and social changes.

Monitoring Trends and Changes: The data obtained from the Internet can be used for various purposes, such as explaining, modeling, now casting (real-time forecasting), and forecasting social behavior. These applications help in monitoring and predicting key trends and changes in the economic and social landscape.

Internet as a Data Provider: The Internet has increasingly become a primary source of data for researchers, businesses, and policymakers to gain a deeper understanding of social dynamics and to make data-driven decisions.

The "Data Big Bang" is a consequence of the digital age, where the Internet and the vast amount of data it generates play a central role in shaping how we live, work, and interact in the modern world. By leveraging the power of Big Data analytics, we can unlock valuable insights that can drive innovation, improve services, and address complex economic and social challenges.

Google Trends: the power of search engines

Google Trends (GT) is an online tool that was introduced in May 2006 that offers real-time statistics on the number of searches made for a certain word or phrase, with historical data accessible since January 2004. It tracks how the demand for knowledge on particular topics changes over time, offering helpful data to identify new trends and societal undercurrents of interest and worry. Choi and Varian (2009a, b) presented the use of GT data to nowcast social and economic (especially macroeconomic) variables by demonstrating how various search categories in the Google search engine could assist in forecasting vehicle and home sales, arriving tourists, or jobless claims. Following that, numerous studies in different nations focused on enhancing the predictions of variables associated to unemployment by leveraging GT data, with positive outcomes.

Social Networking Sites and blogs

SNS are online spaces designed to allow people to express their thoughts and views on any subject. As a result, to some extent, the information they contain is a reflection of what occurs in society. It's true that SNS and blog data are increasingly being referred to as "Social Big Data" (Bello-Orgaz et al., 2016). Because of this, social networking sites (SNS) are receiving increased focus as data sources that could be helpful in predicting social characteristics.

Websites and apps: transactional, opinion platforms and information diffusion

In the Digital Age, businesses typically develop corporate websites to create their official public image online. by way of Fig. 2. classification of non-conventional sources of economic and social data. Table 2: Sources of socioeconomic Big Data are categorised. User's objective Description sources as examples The user conducts an information search to learn more about a subject that interests him. Active data generation internet search; Google Trends transactions The user interacts with a person, a computer, or both to come to an agreement wherein the user requests and receives a good or service in return for monetary or non-monetary remuneration. Data is being actively produced.

- Financial exchanges E-banking, e-commerce, and urban sensors (tolls, credit card readers, retail scanners, and public transportation card readers) are instances where the user makes a payment to acquire a good or service.
- Nonmonetary exchanges When a user exchanges information with a counterpart in order to receive a good or service. E-government and e-employment dispersion of information The user wants to share expertise or information. Included in this are marketing objectives to build a brand for the user or the agent he represents. Data is being actively produced. Business wiki pages, apps, and websites Social engagement The user wants to communicate with other users and exchange knowledge, ideas, and opinions. Data is being actively produced. Using social networks websites, discussion forums, and blogs Non-deliberate Users don't actively seek to produce data by their actions; instead, data are produced through the usage of various tools. Any other user action that results in the generation of data.
- Usage Utilising any device alone generates data about the how, when, and place that an action was performed. Internet Protocol, web cookies, and self-tracking sensors • Situation Mobile phone use creates data, including information about the user's location. GPS, GSM, Bluetooth, and WiFi Points are few examples.
- As a result of utilising any equipment or instrument to accomplish a goal, personal data (age, sex, etc.) is produced, either consciously (e.g., filling out a form to complete a transaction) or unconsciously (e.g., data about the type of information we look for is used to infer our incomes). types of searches or purchases, forms, profiles, etc. Technological Forecasting & Social Change 130 (2018) 99–113 102 D. Blazquez, J. Domenech On these websites, businesses provide

information about their goods, services, organisational structure, and future plans, such exporting and building a branch office abroad. Corporate websites include all types of websites that businesses use in connection with their economic activity, from informational websites that only provide information about the company to transactional websites devoted to both information provision and the provision of online services (e-commerce, e-banking, etc.), where users are occasionally permitted to voice their opinions. That is to say, corporate websites may have three distinct functionalities: conducting business (e-business operations), facilitating opinion exchange (electronic word-of-mouth (eWOM) boosting), and disseminating information about companies (connected to developing a public image). It is interesting that websites have a complicated structure that varies from case to instance, necessitating the use of a particular Big Data architecture in order to standardise the retrieval and analysis of their information. Because of this complexity, corporate websites are a mostly untapped source of information. They are still potential sources of economic data due to their openness, regular updating, and "business generated" nature. Additionally, new research suggests that company traits could arise on the web and be tracked by thoroughly examining corporate websites. By utilising Big Data approaches (specifically web data mining and machine learning) to the "spreading information" functionality of corporate websites, businesses' sales growth and business activities, such as strategies of technology adoption, innovation, and R & D, have been successfully detected (Arora et al., 2016, 2013; Gök et al., 2015; Li et al., 2016).

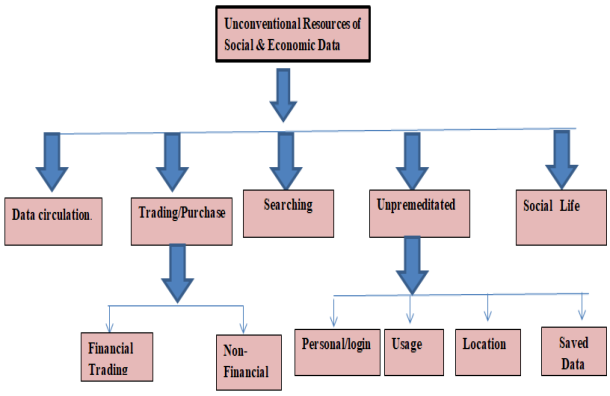

Fig. 2. Taxonomy of Unconventional sources of social and economic data

Table 2. Classification of Unconventional sources of social and economic data.

Parameters	Summary
Data circulation	**Data is being actively produced when the user want to share his knowledge or information for to establish a public imager or the client.**
Trading/Purchase	Financial exchanges occurrence where a user pays for product or service. Non-Financial-like in e-commerce use has given a counter part for the service.
Searching	**The user want to search information or data then the massive amount of data is generated..**
Unpremeditated	User activity will result in data generation in the background user doesn't want to produce it like web-cookies. The data is generated by the following four ways-1)Login or Personal 2)Usage-App installation 3)Location-GPS 4)Saved data-Contacts,Mail
Social life	**In social networking site user upload the photo, video or post to share with other user. It also generated the massive data.**

5. Discussions

Discussing the use of big data from unconventional sources and the data lifecycle for social and economic analyses involves understanding the unique challenges and opportunities that arise when dealing with vast amounts of data from non-traditional sources.

5.1 Big Data from Unconventional Sources:

Within the domain of Big Data, non-traditional data sources, such as unstructured databases and surveys, are being added to the mix, producing revolutionary discoveries. A vast variety of data kinds are included in these non-conventional sources, such as posts from social media, satellite imagery, sensor data, online transactions, and even data from Internet of Things devices. These sources are enticing because of their size, speed, and diversity. Satellite images can provide insights into environmental changes and urban development, while social media platforms offer real-time sentiment analysis

and social trends. Even with their promise, these sources can be difficult to extract useful information from due to issues with data quality, privacy, and the requirement for sophisticated data processing methods like machine learning and natural language processing. Combining domain-specific knowledge, computer science, and statistical expertise is necessary to integrate non-traditional data sources into analysis.

Advantages:

- A vast amount and diversity of data are at your disposal.
- Data collecting in real-time or almost real-time.
- Possibility of finding obscure trends and patterns.
- More diversity and granularity in the data.

Challenges:

- The reliability and quality of data can differ.
- Privacy and ethical issues while handling sensitive or personal information.
- Because data is unstructured, preparation and cleansing might be difficult.
- Possibility of bias during the data gathering procedure.

5.2 Data Life Cycle for Social and Economic Analyses:

The phases that data travels through, from collection to disposal, are referred to as the data life cycle. This cycle is crucial for maintaining the accuracy and value of the data-driven insights in social and economic assessments. The selection of data sources is crucial throughout the cycle's initial phase of data collecting. For social and economic analysis, this can entail compiling information from non-conventional sources as well as administrative documents and surveys. Data pretreatment and cleaning are necessary after collection to get rid of mistakes, inconsistencies, and missing variables that could skew analysis. To protect sensitive data, choices regarding storage infrastructure and security procedures must be made within the data storage and management process. Several statistical and machine learning methods are used in the analysis step to extract significant correlations, patterns, and predictions. Ultimately, the results are disseminated via reports, models, and visualisations that assist in social and economic decision-making. To protect privacy and compliance, choices about data archiving and deletion must be made while taking ethical and legal factors into account.

5.3 Impacts on Social and Economic Analyses:

Significant ramifications for social and economic analysis result from the integration of Big Data and unconventional sources into the data life cycle. Real-time insights from these non-traditional sources allow for quicker reactions to shifting social and economic circumstances. Adaptive decision-making can be facilitated, for example, by sentiment analysis of social media data, which provides insights into public opinion about policies or products. However, working with non-traditional sources presents certain difficulties, like as bias and noise, which must be carefully considered. Big Data's vastness also calls for sophisticated data processing methods, which calls for a skill set that straddles the divide between data science and domain expertise. The potentially sensitive nature of the data makes ethical considerations including data protection, security, and permission crucial. In conclusion, combining Big Data from non-traditional sources with the data life cycle has great potential to improve social and economic studies. However, its effective use necessitates a careful strategy that strikes a balance between ethical obligations and technical capabilities. Discuss the moral issues raised by the use of huge data from unusual sources:

Privacy: Talk about the difficulties in preserving and anonymizing personal data in big datasets.

Bias and Fairness: Describe how bias in data collection and analysis could result in unjust judgements or findings.

Transparency: Stress how crucial it is for data collecting and analysis procedures to be transparent.

6. Conclusions

To sum up, the investigation of non-traditional sources in the context of big data, along with the creation of an all-encompassing data lifecycle customised for social and economic research, represents a fundamental change in our comprehension of and utilisation of information for well-informed decision-making. This strategy allows us to collect subtle insights into market dynamics, societal trends, and human behaviour by integrating multiple data streams from sources including social media, sensor networks, and online platforms. The suggested data lifecycle highlights the ethical issues involved in handling these data sources and guarantees the accuracy and dependability of results through phases of data collection, preparation, analysis, and interpretation. Our initial contribution in this paper is to examine the various unconventional resources of digital data sources and suggests taxonomy to classify them in accordance with the goals of the agent who generated the data. According to the Big Data concept, there are particular processing and modeling techniques are needed for this broad range of heterogeneous digital data sources. The paper's second contribution is to handle the issue by analyzing some unconventional procedures and categorizing them according to the stage of data analysis they work in. The data lifecycle approach has been utilised to manage data

analysis in a stable and flexible architecture and to place it inside an organisational framework. Several viewpoints on this method have been looked at and combined in order to establish and define all the pertinent phases. The ability to extract actionable knowledge is ultimately what this integrated framework fosters, enabling researchers, policymakers, and businesses to better understand complex socio-economic systems and develop more resilient and adaptive strategies for a world that is changing quickly.

References:

[1] J. VAN WEZEL ET AL., "DATA LIFE CYCLE LABS, A NEW CONCEPT TO SUPPORT DATA-INTENSIVE SCIENCE," NO. DLCL, P. 8, 2012, [ONLINE]. AVAILABLE: HTTP://ARXIV.ORG/ABS/1212.5596.

[2] S. I. H. Shah, V. Peristeras, and I. Magnisalis, DaLiF: a data lifecycle framework for data-driven governments, vol. 8, no. 1. Springer International Publishing, 2021.

[3] S. Manzan, "Big Data and Computational Social Science for Economic Analysis and Policy," Handb. Comput. Soc. Sci. Policy, pp. 231–242, 2023, doi: 10.1007/978-3-031-16624-2_12.

[4] P. R. Chandre, P. N. Mahalle, and G. R. Shinde, "Machine Learning Based Novel Approach for Intrusion Detection and Prevention System: A Tool Based Verification," in 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), Nov. 2018, pp. 135–140, doi: 10.1109/GCWCN.2018.8668618.

[5] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: Promise and potential," Heal. Inf. Sci. Syst., vol. 2, no. 1, pp. 1–10, 2014, doi: 10.1186/2047-2501-2-3.

[6] M. Boenig-Liptsin, A. Tanweer, and A. Edmundson, "Data Science Ethos Lifecycle: Interplay of Ethical Thinking and Data Science Practice," J. Stat. Data Sci. Educ., vol. 30, no. 3, pp. 228–240, 2022, doi: 10.1080/26939169.2022.2089411.

[7] V. Sebestyén, T. Czvetkó, and J. Abonyi, "The Applicability of Big Data in Climate Change Research: The Importance of System of Systems Thinking," Front. Environ. Sci., vol. 9, no. March, pp. 1–26, 2021, doi: 10.3389/fenvs.2021.619092.

[8] P. Chandre, P. Mahalle, and G. Shinde, "Intrusion prevention system using convolutional neural network for wireless sensor network," IAES Int. J. Artif. Intell., vol. 11, no. 2, pp. 504–515, 2022, doi: 10.11591/ijai.v11.i2.pp504-515.

[9] M. P. Libório, P. I. Ekel, and C. A. P. da Silva Martins, "Economic analysis through alternative data and big data techniques: what do they tell about Brazil?," SN Bus. Econ., vol. 3, no. 1, pp. 1–16, 2022, doi: 10.1007/s43546-022-00387-z.

[10] Mahadevi Somnath Namose and Dr. Tryambak Hiwarkar, "Predictive Analytics Accomplished by the Utilization of Social Big Data and Machine Learning," Int. J. Adv. Res. Sci. Commun. Technol., vol. 2, no. 1, pp. 828–833, 2022, doi: 10.48175/ijarsct-7570.

[11] Mahadevi Somnath Namose and Dr. Tryambak Hiwarkar, "Predictive Analytics Executed through the Use of Social Big Data and Machine Learning: An Imperious Result," Int. J. Adv. Res. Sci. Commun. Technol., vol. 2, no. 3, pp. 55–64, 2022, doi: 10.48175/ijarsct-7598.

[12] A. Ferretti et al., "Ethics review of big data research: What should stay and what should be reformed?," BMC Med. Ethics, vol. 22, no. 1, pp. 1–13, 2021, doi: 10.1186/s12910-021-00616-4.

[13] K. G. Abraham, R. S. Jarmin, B. C. Moyer, and M. D. Shapiro, "Big Data for Twenty-First-Century Economic Statistics," Big Data Twenty-First-Century Econ. Stat., no. 28, 2022, doi: 10.7208/chicago/9780226801391.001.0001.

[14] X. Li and R. Law, "Network analysis of big data research in tourism," Tour. Manag. Perspect., vol. 33, no. November 2019, p. 100608, 2020, doi: 10.1016/j.tmp.2019.100608.

[15] D. Blazquez and J. Domenech, "Big Data sources and methods for social and economic analyses," Technol. Forecast. Soc. Change, vol. 130, no. March 2017, pp. 99–113, 2018, doi: 10.1016/j.techfore.2017.07.027.

[16] R. Wang et al., "Review on mining data from multiple data sources," Pattern Recognit. Lett., vol. 109, pp. 120–128, 2018, doi: 10.1016/j.patrec.2018.01.013.

[17] Y. Lin, H. Wang, J. Li, and H. Gao, "Data source selection for information integration in big data era," Inf. Sci. (Ny)., vol. 479, pp. 197–213, 2019, doi: 10.1016/j.ins.2018.11.029.

[18] S. Ren, Y. Zhang, Y. Liu, T. Sakao, D. Huisingh, and C. M. V. B. Almeida, "A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions," J. Clean. Prod., vol. 210, no. January 2022, pp. 1343–1365, 2019, doi: 10.1016/j.jclepro.2018.11.025.

[19] J. Koo, G. Kang, and Y. G. Kim, "Security and privacy in big data life cycle: A survey and open challenges," Sustain., vol. 12, no. 24, pp. 1–32, 2020, doi: 10.3390/su122410571.

[20] P. Jain, M. Gyanchandani, and N. Khare, "Big data privacy: a technological perspective and review," J. Big Data, vol. 3, no. 1, 2016, doi: 10.1186/s40537-

016-0059-y.

[21] A. Sinaeepourfard, X. Masip-Bruin, J. Garcia, and E. Marín-Tordera, “A Survey on Data Lifecycle Models: Discussions toward the 6Vs Challenges,” Tech. Rep., 2015.

[22] Y. Wang, L. A. Kung, and T. A. Byrd, “Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations,” Technol. Forecast. Soc. Change, vol. 126, pp. 3–13, 2018, doi: 10.1016/j.techfore.2015.12.019.

[23] L. Tang, J. Li, H. Du, L. Li, J. Wu, and S. Wang, “Big Data in Forecasting Research: A Literature Review,” Big Data Res., vol. 27, p. 100289, 2022, doi: 10.1016/j.bdr.2021.100289.

[24] X. Bai et al., “Educational Big Data: Predictions, Applications and Challenges,” Big Data Res., vol. 26, p. 100270, 2021, doi: 10.1016/j.bdr.2021.100270.

[25] V.S. Kore, B.A. Tidke and P. Chandre, "Survey of image retrieval techniques and algorithms for image-rich information networks", International Journal of Computer Applications, vol. 112, no. 6, pp. 39-42, 2015.

[26] Funde Rahul, Chandre Pankaj "Dynamic cluster head selection to detect gray hole attack using intrusion detection system in MANETs" Proceedings of the Sixth International Conference on Computer and Communication Technology, ACM (2015), pp. 73-77.

[27] Deshpande S, Gujarathi J, Chandre P, Nerkar P. "A comparative analysis of machine deep learning algorithms for intrusion detection in wsn." In: Security Issues and Privacy Threats in Smart Ubiquitous Computing, 2021; pp. 173–193. Springer.

[28] Balasubramanian, S. ., Naruka, M. S. ., & Tewari, G. (2023). Denoising ECG Signal Using DWT with EAVO . International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 231–237. https://doi.org/10.17762/ijritcc.v11i3s.6184

[29] Juhani Nieminen , Johan Bakker, Martin Mayer, Patrick Schmid, Andrea Ricci. Exploring Explainable AI in Educational Machine Learning Models. Kuwait Journal of Machine Learning, 2(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/191

[30] Anupong, W., Azhagumurugan, R., Sahay, K. B., Dhabliya, D., Kumar, R., & Vijendra Babu, D. (2022). Towards a high precision in AMI-based smart meters and new technologies in the smart grid. Sustainable Computing: Informatics and Systems, 35 doi:10.1016/j.suscom.2022.100690