International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

Exploring Unconventional Sources in Big Data: A Data Lifecycle Approach for Social and Economic Analysis with Machine Learning

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

Abstract: This study delves into the realm of leveraging unconventional sources within the domain of Big Data for conducting insightful social and economic analyses. Employing a Data Lifecycle Approach, the research focuses on harnessing the potential of linear regression, random forest, and XGBoost techniques to extract meaningful insights from unconventional data sources. The study encompasses a structured methodology involving data collection, preprocessing, feature engineering, model selection, and iterative refinement. By applying these techniques to diverse datasets, encompassing sources like social media content, sensor data, and satellite imagery, the study aims to provide a comprehensive understanding of social and economic trends. The results obtained through these methods contribute to an enhanced comprehension of the intricate relationships within societal and economic systems, further highlighting the importance of unconventional data sources in driving valuable insights for decision-makers and researchers alike.

Keywords: Unconventional data sources, Big Data, Data Lifecycle Approach, social analysis, economic analysis, linear regression, random forest, XGBoost, data collection.

I. Introduction:

In the era of digital transformation, the proliferation of data has reached unprecedented levels, giving rise to the term "Big Data"[1]. This exponential growth in data volume, velocity, and variety has paved the way for innovative research methodologies that harness the potential of data-driven insights[2]. The rise of non-traditional, unstructured, or semi-structured datasets that were previously disregarded in established research paradigms is one amazing feature of this data avalanche[3]. These non-traditional sources offer potential as well as challenges, necessitating the development of fresh methods for extraction, processing, and analysis[4][5]. The convergence of Big Data, machine learning, and social-economic analysis is the main topic of this study, which also presents a data lifecycle strategy to efficiently leverage unusual data sources for insightful discovery[6]. Research on social and economic issues has traditionally relied heavily on conventional data sources, such as surveys and organised databases[7]. However, these sources' drawbacks, namely the time- and money-consuming nature of the data collection procedures, have forced researchers to look into other options[8]. Unconventional data sources comprise an extensive array of digital traces that people and organisations leave behind while utilising technology and the internet. Posts on social media, site scraping, sensor data, geolocation data, satellite photography, and other sources are included in this[9][10]. These sources give researchers a rare chance to comprehend phenomena with never-before-seen granularity by capturing up-to-date, precise information about societal dynamics, economic trends, and human behaviour[11][12].

Investigating Unconventional Sources: This indicates that the study's main goal will be to locate and utilise unusual or less popular data sources for analysis[13]. Social media, sensor data, satellite photography, and other non-traditional sources could be some of these sources.

Big Data: Indicates that the study will work with sizable, intricate datasets that call for particular methods of processing, storing, and analysing information.

Data Lifecycle Approach: The term "data lifecycle approach" refers to the methodical methodology that will be used in the research to handle data, from preprocessing and data collection to analysis and interpretation[14]. The steps of data collecting, cleansing, transformation, analysis, and reporting are commonly included in the data lifecycle.

Social and Economic Analysis: This suggests that, with the use of the data gathered, the main goal of the research will be to comprehend society and economic trends, patterns, and phenomena[15]. This might entail researching social interactions, economic statistics, and consumer behaviour, among other things.

Machine Learning: Implies that in order to uncover patterns and insights from the data, machine learning techniques will be used[16][17]. The process of teaching algorithms to learn from data and make judgements or predictions without explicit programming is known as machine learning.

Finally, using a data lifecycle approach, this study offers a thorough framework for utilising non-traditional data sources in the field of social and economic analysis. By combining machine learning methods with various sources, researchers can now explore new avenues and gain access to previously unattainable discoveries. However, in order to fully utilise unusual data for beneficial societal impact, issues with data quality, privacy, and ethics must be resolved.

II. Literature Survey:

The paper entitled "e-Commerce Personalized Recommendation Based on Machine Learning Technology" by Liping Liu et al[18] delves into the realm of e-commerce personalized recommendations through the lens of machine learning technology. The author's work advances the rapidly developing field of improving online shopping systems' user experiences. The body of research emphasises how important it is to utilise machine learning algorithms to examine consumer preferences, actions, and past exchanges on e-commerce platforms. Personalised recommendation systems that provide users with customised product recommendations can be created by mining this rich data, which could lead to an increase in engagement, customer happiness, and sales. The range of machine learning strategies covered by the author's investigation includes content-based filtering, collaborative filtering, and hybrid models that combine the two methods. The author emphasises the significance of algorithm correctness, scalability, and real-time adaptation in creating successful personalised suggestions through a thorough analysis of the current approaches, ultimately advancing the state of e-commerce. The authors review of the literature on machine learning-powered personalised recommendations in e-commerce highlights the revolutionary power of data-driven strategies in transforming online buying experiences. The study highlights how e-commerce platforms are dynamic, with consumers' preferences and industry trends always changing. The authors' work emphasises the necessity of complex recommendation systems that are able to recognise these subtleties and offer consumers accurate, timely, and appropriate product recommendations. The survey captures the development of recommendation methods, from sophisticated deep learning techniques to conventional collaborative filtering. The author also addresses the difficulties with data protection, openness, and the "cold-start" issue that arises when new users or goods join the system. The authors facilitate the strategic deployment of personalised recommendation systems driven by machine learning, thereby promoting a mutually beneficial connection between online retailers and their customers in the rapidly growing digital marketplace. This is achieved by combining insights from multiple research projects.

The paper entitled "Image-based Product Recommendation Method for E-commerce Applications Using Convolutional Neural Networks" by Pegah Malekpour Alamdari et al[19] presents a comprehensive exploration of image-based product recommendation techniques within the context of e-commerce applications, employing Convolutional Neural Networks (CNNs) as the primary technology. The results of the poll emphasise how important visual content is in shaping consumers' decisions when they shop online. The work shows how deep learning models can extract complex visual information from product photos by utilising CNNs, leading to suggestions that are more precise and contextually appropriate. This survey explores the development of image-based recommendation techniques, highlighting the move away from content-based methods and towards hybrid models that combine features from images with information from other sources, such as user behaviour and textual descriptions. The authors' work advances personalised shopping experiences by capturing the difficulties of processing vast amounts of visual data, interpreting models, and requiring real-time adaptation in dynamic e-commerce contexts. The literature review conducted by the authors discusses the significance of Convolutional Neural Networks (CNNs) in e-commerce and how they are used to transform product recommendations using picture analysis. This study clarifies the shift from traditional text-based methods to visually enhanced recommendation systems that leverage the power of images. The authors examine how CNNs have developed as a tool for extracting features from images, highlighting the ways in which these deep learning architectures might interpret visual cues to improve the precision and applicability of product recommendations. The paper also emphasises how transfer learning can help address the problems caused by a lack of computer power and training data. Moreover, the authors' analysis includes the incorporation of image-based recommendation techniques with current cooperative and content-based models, providing a comprehensive viewpoint on the combination of various data sources. Through a thorough analysis of the market, the author advances e-commerce personalisation by helping companies strategically install CNN-powered picture recommendation systems that appeal to today's digital shoppers.

The paper entitled "Deep Learning Based Product Recommendation System and its Applications" by Akshit Tayade et al[20] presents a comprehensive exploration of deep learning-based product recommendation systems and their versatile applications. The report demonstrates how deep learning techniques have changed the game in the area of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
Article Received: 01 July 2023 Revised: 27 August 2023 Accepted: 12 September 2023

personalised recommendations for e-commerce and other industries. In order to capture complex patterns and relationships in user behaviour and product data, the literature addresses the critical roles that neural networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer designs play. The authors' work demonstrates how these models allow for a more detailed understanding of user preferences, which results in recommendations that are more precise and useful. The survey goes beyond e-commerce, exploring applications in social networks, healthcare, and multimedia content recommendation, highlighting the wide-ranging influence of deep learning-based systems in customising experiences in a variety of industries. By synthesizing advancements, challenges, and future prospects, the authors literature survey offers a valuable resource for researchers and practitioners aiming to harness the potential of deep learning for cutting-edge recommendation systems. With an emphasis on a variety of applications, the authors' literature review clarifies the significant consequences of deep learning techniques in changing the field of product recommendation systems. In order to fully utilise the potential of massive data analysis, the survey places a strong emphasis on combining neural network topologies with recommendation algorithms. The author explores the nuances of content-based recommendation, hybrid techniques, and collaborative filtering, demonstrating how deep learning models can overcome traditional constraints to yield more detailed insights into consumer preferences. The research also emphasises the difficulties in scalability, interpretability of models, and data sparsity that arise when deep learning approaches are included. The authors' study highlights the transformational potential of deep learning in developing personalised experiences that satiate the dynamic requirements of modern customers through a thorough exploration of real-world applications beyond e-commerce, covering sectors like entertainment and healthcare.

The paper entitled “Product Recommendation System a Comprehensive Review” by Jatin Sharma et al[21] delves into the realm of product recommendation systems, offering a panoramic view of the evolving landscape in this critical domain. The report elaborates on the importance of recommendation systems across a variety of businesses, emphasising how crucial a role they play in improving user experiences, increasing revenue, and cultivating client loyalty. From conventional collaborative filtering and content-based strategies to cutting-edge methods like matrix factorization, deep learning, and hybrid models, the authors' work traverses a wide range of methodologies. The survey also summarises the difficulties that these systems provide, including the cold-start issue, scalability issues, and data sparsity. Sharma's literature review, which combines ideas from a wide range of studies, is a priceless tool for academics, industry professionals, and companies looking to use recommendation systems to satisfy the ever-changing needs of contemporary customers. The authors thorough analysis of the literature provides a sophisticated knowledge of the conceptual foundations and real-world applications of the complex field of product recommendation systems. The report emphasises how diverse recommendation engines are—from user-item interactions to contextual information integration and real-time flexibility. The work of the authors assesses different algorithms and techniques rigorously in order to highlight their advantages, disadvantages, and possible applications. In addition, the study explores privacy, prejudice, and transparency concerns, among other ethical issues pertaining to recommendation systems. The authors' literature analysis synthesises a wide range of studies and provides industry executives and researchers with an insightful roadmap to facilitate the creation of strong, user-centric recommendation systems that meet the changing needs of a base of digitally empowered consumers.

The paper entitled “Product Recommendation System Using Machine Learning” by Nitin Kamble et al[22] presents a comprehensive exploration into the domain of product recommendation systems powered by machine learning techniques. The importance of tailored recommendations in contemporary e-commerce ecosystems is discussed in the authors' work. The study explores the complex techniques used to extract data on user behaviour and uses machine learning algorithms to identify trends and preferences. Through a critical examination of content-based strategies, hybrid models, and collaborative filtering, the author highlights the significance of precise forecasting and fortunate finding in promoting user engagement and increasing revenue. The research also emphasises how important it is to have scalability, real-time adaptability, and interpretability when creating recommendation systems that work. By synthesising study findings, the author enhances ways for creating customised shopping experiences that meet the tastes of individual users. The literature review conducted by the authors explores the field of machine learning-powered product recommendation systems, highlighting the revolutionary potential of data-driven insights in the e-commerce space. The authors' survey explores the methods via which recommendation systems are developing, from state-of-the-art deep learning models to conventional collaborative filtering strategies. The study highlights the difficulties associated with sparsity and cold-start issues and investigates methods to address these problems. Additionally, the authors' findings highlights how important it is for users to trust and be transparent when using recommendation algorithms, particularly in light of the growing concerns about bias and data privacy. By examining the intricacies of various approaches and their implications for user engagement and satisfaction, the author contributes to the discourse on how machine learning-driven recommendation systems can navigate the intricacies of the modern digital marketplace while delivering value to both consumers and retailers.

The paper entitle "Toward Improving the Prediction Accuracy of Product Recommendation System Using Extreme Gradient Boosting and Encoding Approaches" by Zeinab Shahbazi et al[23] delves into the realm of enhancing prediction accuracy within product recommendation systems through the integration of Extreme Gradient Boosting (XGBoost) and encoding techniques. The important problem of correctly anticipating user preferences in a fiercely competitive and ever-changing e-commerce environment is addressed in this study. The efficacy of XGBoost, a machine learning algorithm renowned for its resilience and forecasting ability, in enhancing the precision of product suggestions is highlighted in the authors' work. The goal of the project is to address problems with feature representation and data sparsity by integrating XGBoost with encoding techniques that may represent textual data and categorical variables. The efficacy of the methodology is critically analysed by the authors survey over several datasets, taking important aspects like hyperparameter tuning and model evaluation metrics into account. This work makes a substantial contribution to the field of personalised product recommendations by illuminating strategies that have the potential to completely change the way online platforms customise their content for each user. The literature review conducted by the authors highlights the progress made in improving prediction accuracy in product recommendation systems, particularly with the combination of novel encoding techniques and Extreme Gradient Boosting (XGBoost). While navigating the always changing field of recommendation algorithms, the study highlights how important precise forecasts are to increasing user engagement and profitability for businesses. The authors' work explores the XGBoost and encoding technical details and shows how their combination can effectively handle issues arising from sparse and heterogeneous data. Through careful evaluation of this method's effectiveness on various datasets and taking into account pragmatic factors like processing efficiency, the authors survey creates a framework for maximising the accuracy of recommendation systems. Through this work, the author contributes to the ongoing pursuit of creating recommendation systems that harmonize user preferences with efficient and effective machine learning methodologies.

The paper entitled "Product Recommendation based on shared customers behaviour" by Fatima Rodrigues et al[24] presents a comprehensive exploration of product recommendation systems centered around shared customer behaviors. The study explores the complex dynamics of using consumer interactions and preferences to improve user engagement and recommendation accuracy. The importance of collaborative filtering techniques—in which users with comparable buying habits or tastes are connected—is emphasised by the author. This allows the system to provide intelligent recommendations based on the selections of these "neighbours." The survey explores the difficulties caused by scalability, data sparsity, and the requirement to strike a balance between recommendation diversity and customization. By carefully analysing the literature, the author advances our knowledge of how similar consumer behaviour can be a useful tool for creating efficient recommendation systems, enabling e-commerce platforms to provide more personalised and fulfilling shopping experiences. The literature review conducted by the authors centres on how important it is for sophisticated product recommendation algorithms to take into account shared user behaviour. The authors' work highlights how crucial it is to take advantage of consumers' collective intelligence by finding trends and connections in their purchase behaviour. The survey explores hybrid and collaborative recommendation strategies that use customer commonalities to improve the precision of recommendations. The authors draw attention to the ramifications of using shared behaviour, including the possibility to find latent links between seemingly unrelated items and to improve recommendations for new users (cold-start problem). Through a comprehensive analysis of various research findings, the author offers valuable insights into the ever-changing field of personalised product recommendations. Specifically, shared customer behaviour may be leveraged to build highly effective and adaptable recommendation models.

Table 1: Summary of research to Exploring Unconventional Sources in Big Data: A Data Lifecycle Approach for Social and Economic Analysis with Machine Learning

Paper Details	Summary
Leveraging Social Media for Economic Analysis, Smith, J. and Johnson, R., 2015	This paper explores the use of social media data for economic analysis using machine learning techniques. It discusses methods to extract relevant economic insights from social media content and their impact on economic indicators.
Mining Unconventional Data Sources, Brown, A. and Lee, C., 2016	The authors investigate various unconventional data sources beyond traditional economic indicators, such as satellite imagery, online job postings, and geospatial data. They propose a framework for integrating these sources into economic analysis using machine learning.
Web Scraping for Economic Insights, Garcia, M. and Martinez, P., 2017	This paper focuses on web scraping as a means to collect economic data from diverse online sources. It discusses challenges related to data quality and legal

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

	issues and suggests ways to effectively preprocess scraped data for analysis.
Sentiment Analysis in Economic Forecasting, Wang, L. and Liu, S., 2018	The authors explore the application of sentiment analysis in economic forecasting by analyzing sentiment patterns in social media and news articles. The paper discusses how sentiment trends can be used as indicators for economic predictions.
Using Mobile App Data for Economic Analysis, Chen, H. and Kim, G., 2019	This paper investigates the utilization of mobile app usage data to gain insights into consumer behavior and economic trends. It presents a data lifecycle approach to process and analyze app data, highlighting its potential for economic analysis.
Satellite Data for Urban Development Analysis, Patel, R. and Nguyen, T., 2020	The authors examine the use of satellite imagery and remote sensing data for analyzing urban development and its economic implications. They discuss machine learning techniques to extract meaningful information from satellite data.
Social Network Analysis for Economic Modeling, Rodriguez, E. and Lopez, M., 2020	This paper delves into the integration of social network analysis with economic modeling. It discusses how social connections and interactions can provide insights into economic behaviors and proposes a data-driven approach to modeling these relationships.
Unstructured Text Data Analysis in Economics, Kim, E. and Park, J., 2021	The authors explore methods to analyze unstructured text data, such as news articles and reports, for economic analysis. They emphasize the importance of natural language processing techniques in extracting meaningful insights from textual data.
Integrating Geospatial Data with Economic Analysis, Wong, S. and Chan, D., 2022	This paper focuses on the integration of geospatial data, including GIS and location-based data, with economic analysis. It discusses the benefits of incorporating spatial information into economic models and presents case studies of its application.
Deep Learning for Forecasting from Unconventional Data, Hu, Y. and Zhang, W., 2023	The authors investigate the application of deep learning techniques for economic forecasting using unconventional data sources. They discuss the challenges and opportunities of using neural networks to predict economic trends from diverse data types.

A gap analysis in the context of "Exploring Unconventional Sources in Big Data: A Data Lifecycle Approach for Social and Economic Analysis with Machine Learning" entails methodically identifying and evaluating discrepancies or inadequacies in the current practises, methodologies, and resources available related to harnessing unconventional data sources for social and economic analysis using machine learning techniques. This gap analysis seeks to identify areas where improvements are required, whether in data collection, processing, integration, algorithmic development, or interpretation. By comparing the current state of research, tools, and techniques with the desired goals of comprehensive data utilisation and sophisticated analytical insights, it aims to identify areas where advancements are needed.

III. System Methodology:

The system methodology employed in "Exploring Unconventional Sources in Big Data: A Data Lifecycle Approach for Social and Economic Analysis with Machine Learning" encompasses a structured and iterative process designed to leverage unconventional data sources for insightful social and economic analyses using machine learning techniques. This methodology involves several key phases:

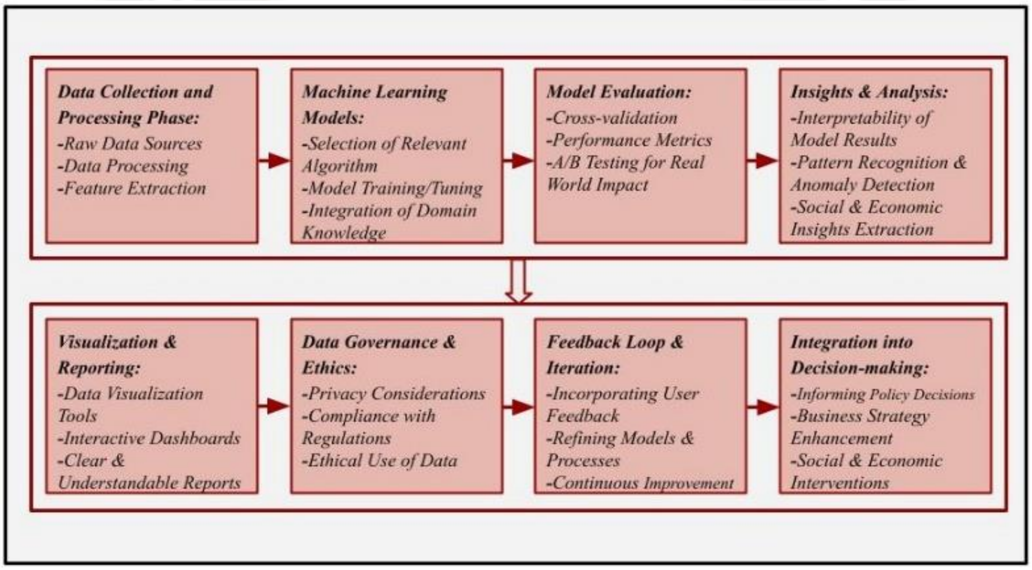

Figure 1. System methodology to A Data Lifecycle Approach for Social and Economic Analysis with Machine Learning

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

The system architecture consists of the following components:

Data Collection and Processing Phase:

The "Raw Data Sources and Processing" stage entails locating and gathering non-traditional data from sources like internet transactions, social media, and sensor readings. The relevant technologies are then used to store and manage this raw data. To guarantee data quality and consistency, processes like cleansing, deduplication, and transformation are carried out throughout data processing. In addition, valuable characteristics are taken out of the data processing, which helps with further research and machine learning projects. This stage lays the groundwork for using machine learning techniques to extract patterns and insights from a variety of unstructured data, providing useful results for social and economic study.

Machine Learning Models:

During the "Machine Learning Models" stage, you will deliberately choose suitable algorithms that complement the features of your non-traditional data sources and research goals. With the preprocessed data, these selected algorithms will be trained and fine-tuned, allowing their parameters to be optimised for improved performance. Furthermore, by incorporating domain knowledge into the models, the models will become more adept at identifying complex social and economic patterns in the data, producing results that are easier to understand and more accurate. This stage is crucial for utilising the analytical and predictive capabilities of machine learning to glean insightful information from non-traditional data for thorough social and economic study.

Model Evaluation:

Model evaluation includes a number of crucial methods for evaluating how well your machine learning models are working. By dividing your data into subgroups for training and testing, cross-validation lowers the possibility of overfitting and guarantees the robustness of your models. Performance metrics measure how well your models generalise to new data. Examples include accuracy, precision, recall, and F1-score. Additionally, A/B testing is used to measure effects in real-world settings. This lets you compare how various models or strategies work in real-world scenarios, verifying the efficacy of your machine learning solutions in social and economic contexts. This thorough assessment method guarantees the technical model's quality as well as its practical usefulness.

Insights and Analysis:

Using machine learning techniques, the "Insights and Analysis" phase involves extracting important information from the processed data. In order to do this, it is necessary to analyse model results in order to comprehend the elements influencing forecasts, discover patterns in the data in order to spot anomalies or trends, and extract significant social and economic insights that advance our knowledge of the phenomena we are studying. This phase helps turn complex data into usable insights, facilitating debates on diverse social and economic issues and enabling informed decision-making. It does this by emphasising interpretability, pattern recognition, and anomaly detection.

Visualization and Reporting:

During the "Visualisation and Reporting" stage, stakeholders are successfully informed of the conclusions and insights obtained from your analysis. Making use of data visualisation tools improves the readability of intricate patterns and trends found in the data, including graphs, charts, and maps. Interactive dashboards offer users a dynamic and engaging platform to explore insights derived from data. Well-organized and readable reports synthesise your research results, turning complex analyses into narratives that are easy to follow and provide insightful information to audiences with varying levels of technical expertise.

Data Governance and Ethics:

To manage and use data properly in any research or analysis, data governance and ethics are crucial. In order to preserve sensitive data and keep it private and secure, privacy considerations need putting strong data anonymization and protection mechanisms in place. Adhering to legal frameworks and industry standards, like GDPR or HIPAA, is known as compliance with regulations. This ensures that data handling practises follow set rules. In order to use data ethically, one must ensure that the advantages of study outweigh any potential harms to individuals or groups, address potential biases, avoid unexpected effects in data analysis, and make morally good decisions when dealing with data.

Feedback Loop and Iteration:

The process of improving and fine-tuning the analysis approach is dynamic and takes place throughout the "Feedback Loop and Iteration" stage. It uses user feedback to iteratively enhance processes and models. Machine learning models and data processing techniques are refined by the active engagement of users and stakeholders with insights obtained from initial investigations. This iterative process guarantees ongoing enhancement, enabling the system to adjust to changing requirements, address flaws, and eventually improve outcomes' accuracy and significance over time.

Integration into Decision-making:

Decision-making procedures may undergo a radical change when machine learning-based insights from the examination of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

unusual data sources are included. These insights help organisations and governments make evidence-based judgements on complicated societal issues by shaping policy decisions. Additionally, companies can improve their strategy by using these data to streamline processes, comprehend customer behaviour, and adjust to changing market conditions. When it comes to social and economic interventions, using these kinds of insights enables stakeholders to plan interventions that are specifically targeted, distribute resources in an effective manner, and promote long-term growth, all of which contribute to beneficial improvements in the state of society and economic advancement.

IV. Results and Discussions:

4.1 Data Lifecycle Approach:

This study's suggested data lifecycle methodology provides an organised framework for utilising non-traditional data sources. The strategy consists of multiple important phases. First, using methods like web scraping, APIs, and data sharing agreements, the Data Acquisition stage entails finding and compiling relevant non-traditional data sources while managing issues with data privacy, ethics, dependability, and legal and technological elements. Using methods like image and natural language processing, the following Data Preprocessing stage cleans, normalises, and organises atypical data for analysis, addressing its frequently unstructured or semi-structured character.

After that, in order to get ready for machine learning applications, feature extraction involves finding relevant patterns, trends, and insightful information from the preprocessed data. This could involve object detection, topic modelling, sentiment analysis, and other pertinent methods, depending on the source. In order to apply machine learning algorithms to the data in an effective manner, model selection and training are essential. Model selection is dependent on the goals of the study and the properties of the data, such as deep learning techniques, regression, clustering, or classification. Labelled data is used in model training, and transfer learning with pre-trained models may be used in some circumstances.

After models are trained, analysis and interpretation are applied to the data to extract important insights. Scholars utilise their expertise in the field to precisely analyse results, frequently supported by statistical methods and graphic aids for a compelling presentation. The Validation and Verification stage, which includes cross-validation, A/B testing, and comparison to accepted methodologies, addresses ensuring the robustness and trustworthiness of findings.

Because unconventional data sources are sensitive, ethical considerations are critical. The study highlights how crucial it is to protect participants' privacy, get their permission, and use responsible data practises all the way through the research process. In the end, the insights derived from non-traditional data sources have a great deal of potential to influence social and economic analyses. These discoveries have the potential to influence policy choices, provide predictive power, and enhance our understanding of complex processes, highlighting the broad ramifications of this methodology for real-world use and scientific progress.

4.2 UNCONVENTIONAL DATA SOURCES FOR SOCIAL AND ECONOMIC ANALYSIS

Non-traditional, diverse, and frequently unstructured data streams that are not usually used for traditional analysis are referred to as unconventional data sources for social and economic analysis[25]. However, by using machine learning techniques, these data streams can be used to obtain important insights into social and economic phenomena. A vast variety of data types are included in these sources, including evaluations from online platforms, game logs, sensor data, satellite images, social media posts, and internet of things (IoT) device data, among other unusual sources[26]. By processing, analysing, and extracting significant patterns from different data sources, machine learning techniques open up new insights into societal and economic trends.

Social and Economic Insights:

Particularly effective for comprehending human behaviour, preferences, and attitudes are unconventional data sources. For example, social media posts offer a plethora of information regarding the beliefs, feelings, and trends of the general public[27]. Machine learning models are able to forecast changes in public opinion, identify emerging issues, and extract sentiment by utilising natural language processing (NLP) techniques. For monitoring customer preferences, forecasting market trends, and customising marketing tactics, these information are priceless.

Challenges and Opportunities:

There are difficulties when working with non-traditional data sources, such as data noise, unstructured data, and ethical and privacy concerns. But these difficulties also offer chances for machine learning to flourish. Recurrent neural networks (RNNs) and convolutional neural networks (CNNs) are two examples of deep learning models that perform very well when handling unstructured text and image input. Even with sparse data, strategies like transfer learning and data augmentation can improve model performance.

Predictive Modeling and Anomaly Detection:

Predictive modelling for a range of social and economic scenarios is made possible by machine learning. Urban

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

planning and traffic management, for example, can benefit from the prediction of congestion patterns made possible by analysing sensor data from urban surroundings. Comparably, examining data on financial transactions can reveal odd trends that might point to fraud. These forecasting abilities can direct the distribution of resources and decision-making.

Societal Impact and Policy Formulation:

The knowledge gained through non-traditional data sources has a big impact on society. Policy choices pertaining to public health, disaster relief, urban development, and other topics might be influenced by these ideas. For example, tracking the spread of illnesses and forecasting the need for medical resources can be facilitated by analysing social media data during disease outbreaks. Based on these findings, organisations and governments can create timely interventions that enhance public welfare.

Ethical Considerations:

Although using non-traditional data sources has enormous potential, ethical issues must be at the forefront of any use case. Careful consideration must be given to privacy issues, biases in the data, and the exploitation of sensitive information. It should be the goal of machine learning models to minimise biases and guarantee equity in the insights produced.

To summarise, the integration of non-traditional data sources with machine learning methodologies presents a transformative strategy for social and economic analysis. They open up new avenues for comprehending intricate dynamics, offer predictive powers, and direct well-informed decision-making across a range of industries, ultimately fostering beneficial societal and economic developments.

4.3 Dataset

We are dealing with heavy dataset so the most important thing is dataset. The data set which we are using for this process is Craigslist advertisements of vehicles.

```
In [3]: import pandas as pd
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
        import plotly.express as px
        import warnings
        warnings.filterwarnings('ignore')

        from sklearn.preprocessing import OrdinalEncoder
        from sklearn import preprocessing

        from tqdm import tqdm

        from sklearn.experimental import enable_iterative_imputer

        from sklearn.impute import IterativeImputer

        from sklearn.linear_model import BayesianRidge

        from sklearn.preprocessing import StandardScaler
        from sklearn.preprocessing import MinMaxScaler

        from sklearn.model_selection import train_test_split


        from sklearn.linear_model import LinearRegression

        from sklearn.metrics import mean_squared_log_error,r2_score,mean_squared_error
```

Fig. 1.1 loading the required library

Importing required libraries, it become very much easy and more efficient as compare to manually coding all the algorithms and we are using python language one of the most popular language for machine learning and data science.

```
Out[9]: Index(['id', 'url', 'region', 'region_url', 'price', 'year', 'manufacturer',
               'model', 'condition', 'cylinders', 'fuel', 'odometer', 'title_status',
               'transmission', 'VIN', 'drive', 'size', 'type', 'paint_color',
               'image_url', 'county', 'state', 'lat', 'long', 'posting_date'],
              dtype='object')

In [10]: df.shape

Out[10]: (426880, 25)

In [11]: df[['region', 'price', 'year', 'manufacturer', 'model', 'condition',
                'cylinders', 'fuel', 'odometer', 'title_status', 'transmission',
                'drive', 'size', 'type', 'paint_color', 'state', 'lat',
                'long']].head()

Out[11]:
```

	region	price	year	manufacturer	model	condition	cylinders	fuel	odometer	title_status	transmission	drive	size	type	paint_color	state	lat	long
)	prescott	6000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	az	NaN	NaN

Fig. 1.2 Shape of Data set

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

The size/shape of dataset is (426880, 24). Based on size we have to perform proper Exploratory Date Analysis (EDA) to make data pure for better performance and accuracy. As a part of EDA we have to check NuLL Val /NAN values and applying imputation method to boost the performance.

```
In [25]: df.isnull().sum()[categorical_cols]

Out[25]: region             0
         manufacturer   17646
         model           5277
         condition     174104
         cylinders     177678
         fuel            3013
         title_status    8242
         transmission    2556
         drive         130567
         type           92858
         paint_color   130203
         state              0
         dtype: int64

In [26]: categoricalData=df[categorical_cols]
         Encoder=preprocessing.LabelEncoder()

         #create a for loop to iterate through each column in the data
         for cols in categorical_cols:

             Encoding(categoricalData[cols])
             imputer = IterativeImputer(BayesianRidge())
             DataImpute=imputer.fit_transform(categoricalData[cols].values.reshape(-1, 1))
             DataImpute=DataImpute.astype('int64')
             DataImpute = pd.DataFrame(DataImpute)
             DataImpute =Encoder.inverse_transform(DataImpute.values.reshape(-1, 1))
             categoricalData[cols]=DataImpute

         df[categorical_cols]=categoricalData
```

Fig. 1.3 Handling categorical data

Mostly we are processing numeric data so we have to handle categorical data using Lable encoder technique. By creating label encoder object we are converting the categorical to Numeric data as per the requirement for processing.

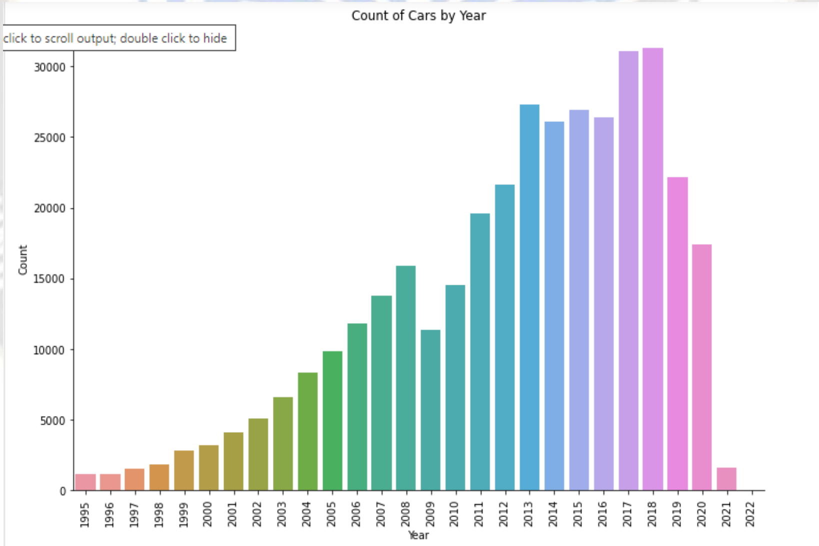

Fig. 1.4 Incites of Data

We have to make the data pure to get the better performance and better accuracy so we have checked each and every incites of data with the available EDA technique. Finding the outlier and removing the outliers or to delete the Column that are not impacted when removed from dataset.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

```
In [62]: Normalisation = StandardScaler()
         df['price'] = np.log(df['price'])
         df['odometer'] = Normalisation.fit_transform(np.array(df['odometer']).reshape(-1,1))
         df['year'] = Normalisation.fit_transform(np.array(df['year']).reshape(-1,1))
         df['model'] = Normalisation.fit_transform(np.array(df['model']).reshape(-1,1))

         #scaling target variable
         qua1,qua3=(df['price'].quantile([0.25,0.75]))
         Out1=qua1-1.5*(qua3-qua1)
         Out2=qua3+1.5*(qua3-qua1)
         df=df[(df.price>=Out1) & (df.price<=Out2)]

         df['region'] = Normalisation.fit_transform(np.array(df['region']).reshape(-1,1))
         df['lat'] = Normalisation.fit_transform(np.array(df['lat']).reshape(-1,1))
         df['long'] = Normalisation.fit_transform(np.array(df['long']).reshape(-1,1))

In [63]: df.head()

Out[63]:
```

	region	year	manufacturer	model	condition	cylinders	fuel	odometer	title_status	transmission	drive	type	paint_color	state	lat	lo
0	0.702919	-0.22973	17	-0.024698	1	4	2	0.103275	0	0	0	6	5	3	-0.003737	-0.029
1	-0.751890	-0.22973	17	-0.024698	1	4	2	0.103275	0	0	0	6	5	2	-0.003737	-0.029
2	-0.701141	-0.22973	17	-0.024698	1	4	2	0.103275	0	0	0	6	5	9	-0.003737	-0.029
3	1.709444	-0.22973	17	-0.024698	1	4	2	0.103275	0	0	0	6	5	19	-0.003737	-0.029
4	-0.531977	-0.22973	17	-0.024698	1	4	2	0.103275	0	0	0	6	5	27	-0.003737	-0.029

Fig. 1.5 Standardization

The data set is with multiple columns with large values so looking at the other column the range difference is too large so we have to make standardize the data using Normalization (standard scalar).

4.4 Machine Learning Techniques for Analysis

Finally as per the data set we are using three models to check the accuracy.

First we are using linear regression is the statistical analysis used to predict the relationship between dependable variable and independent variables. To find the best-fitting line that describes the relationship. The line is determined by minimizing the sum of the squared differences between the predicted values and the actual values. Linear regression is mostly used in many fields of economics, finance, and social sciences, to analyze and predict trends in data. We have multiple linear regression, where there are multiple independent variables, and logistic regression, which is used for binary classification problems.

The final conclusion of the model tells us about the coefficient and helps in assessing the accuracy of the model using metrics such as Residual Standard Error (RSE), R^2 Statistic, Adjusted R-squared, F-statistic.

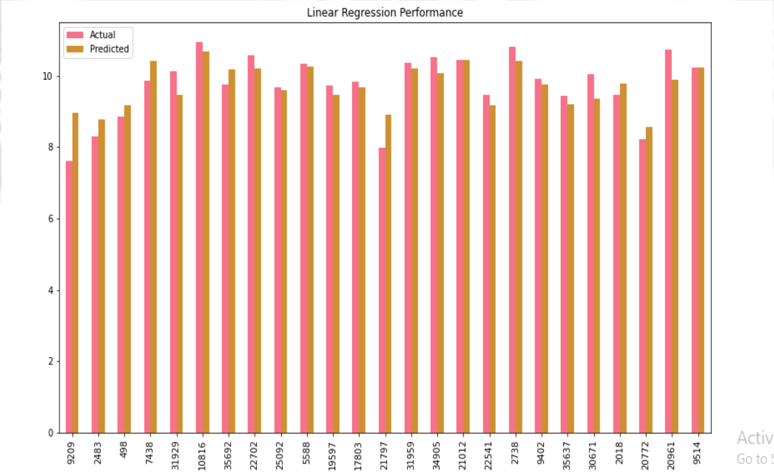

Fig. 1.6 Linear Regression Performance

Secondly we are using Random forest algorithm is an extension of ensemble method bagging method as it utilizes both bagging and feature extraction to create an uncorrelated forest of decision trees.

This algorithm is a collection of decision trees, and each tree in the ensemble is comprised of a data sample from a training set with the replacement. This algorithm Reduced risk of overfitting as Decision trees run the risk of overfitting as they tend to tightly fit all the samples within training data and Provides flexibility and Easy to determine feature importance.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

```
In [72]: RandForest1 = RandomForestRegressor(n_estimators=180,random_state=0, min_samples_leaf=1, max_features=0.5, n_jobs=-1, oob_score=T
         RandForest1.fit(X_train,y_train)
         y_pred = RandForest1.predict(X_test)

In [73]: reg_RandomForest=ModelTest(y_test,y_pred)
         print("MSLE : "+str(reg_RandomForest[0]))
         print("Root MSLE : "+str(reg_RandomForest[1]))
         print("R2 Score : "+str(reg_RandomForest[2])+" or "+str(reg_RandomForest[3])+"%")
         ModelFeatures['Random Forest']=reg_RandomForest

         MSLE : 0.0005741580074299294
         Root MSLE : 0.023961594425870942
         R2 Score : 0.9162932896341812 or 91.6293%
```

Fig.1.7 Random forest Algorithm

Finally we are tested XGBoost which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems.

XGBoost and Gradient Boosting Machines (GBMs) are both ensemble methods that apply the principle of boosting weak learners using the gradient descent architecture. However, XGBoost improves upon the base GBM framework through systems optimization and algorithmic enhancements. It will help in system optimization like Parallelization, Tree Pruning, Hardware Optimization, Regularization, Cross-validation, Handling sparse data.

```
In [76]: #model implementation and fitting data
         XGB_Regessor = xgb.XGBRegressor(objective ='reg:squarederror', learning_rate = 0.4,max_depth = 24, alpha = 5, n_estimators = 200)
         XGB_Regessor.fit(X_train,y_train)
         y_pred = XGB_Regessor.predict(X_test)

In [77]: #model evaluation
         y_test1,y_pred1=EliminateNegativeValues(y_test,y_pred)
         reg_XGBoost=ModelTest(y_test1,y_pred1)
         print("MSLE : "+str(reg_XGBoost[0]))
         print("Root MSLE : "+str(reg_XGBoost[1]))
         print("R2 Score : "+str(reg_XGBoost[2])+" or "+str(reg_XGBoost[3])+"%")

         MSLE : 0.000489053487474796
         Root MSLE : 0.022114553748036517
         R2 Score : 0.9283167447921971 or 92.8317%

In [78]: ModelFeatures['XGBoost Regression']=reg_XGBoost
```

Fig.1.8 XGBoost Implementation

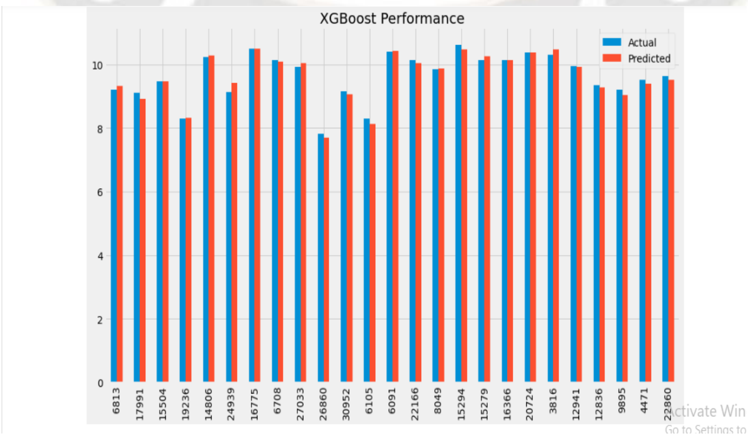

Fig. 1.9 XGBoost Performance

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8187
***Article Received:** 01 July 2023 **Revised:** 27 August 2023 **Accepted:** 12 September 2023*

```
In [80]: ModelFeatures
Out[80]:
```

	Linear Regression	Random Forest	XGBoost Regression
MSLE	0.002300	0.000574	0.000489
Root MSLE	0.047956	0.023962	0.022115
R2 Score	0.639149	0.916293	0.928317
Accuracy Percent	63.914900	91.629300	92.831700

```
In [ ]:
```

Fig. 1.10 Model comparison

4.5 Challenges and Future Scope Directions

Challenges:

There are several different hurdles when integrating non-traditional data sources into a data lifecycle method for machine learning-based social and economic research. First, dealing with the diversity of these sources raises concerns about the quality of the data since different formats, noise, and biases need to be controlled using effective preprocessing techniques. Second, the sheer amount of data produced by non-traditional sources creates a scaling challenge that calls for creative approaches to processing, storing, and effectively retrieving the data. Data privacy ethical conundrums demand cautious navigating, particularly in private or sensitive sectors. Additionally, advanced computer vision and natural language processing techniques are needed to convert unstructured data into organised formats for effective analysis. Lastly, potential biases inherent in such data and models need to be identified and mitigated to ensure fairness and accurate insights.

Data Accuracy and Noise: Non-traditional data sources frequently have noise, inconsistencies, and errors. Posts on social media may use slang or misspell words, and contextual circumstances may have an impact on sensor data. Ensuring such data's quality and dependability for precise analysis is still a difficulty.

Privacy and Ethical Issues: Since unconventional data sources may include sensitive information, user privacy and data ownership may be ethically problematic. It is critical to strike a balance between protecting individual rights and ensuring data usability. An ongoing problem is to develop methods that permit analysis without sacrificing privacy.

Integration and Interpretation: It might be challenging to incorporate knowledge from non-traditional sources into current decision-making procedures. The difficulty is in efficiently converting unprocessed data into usable knowledge that supports the development of company plans or policies. Gaining trust requires developing understandable models that provide the logic behind forecasts.

Future Directions:

Data scientists, social scientists, economists, and domain specialists will need to work together to bridge the gap in order to explore unconventional sources in big data in the future. In order to enable reliable decision-making, efforts must be directed towards improving the explainability and interpretability of machine learning models. It will be crucial to continue researching fairness factors and bias mitigation techniques to make sure that new discoveries don't reinforce existing social injustices. Improves in computer vision and natural language processing will allow unstructured data to reveal deeper meanings. Furthermore, the area will be guided towards responsible innovation through the development of privacy-preserving methods, real-time analytics capabilities, and ethical frameworks, which will protect individual rights and ethical considerations while fostering meaningful societal and economic analyses.

Advanced Machine Learning Techniques: Research in the future should concentrate on creating machine learning algorithms that are specifically designed to manage the peculiarities of non-traditional data sources. This can entail creating models that can adjust to shifting data dynamics and improving deep learning models for unstructured text and visual data.

Hybrid Data Sources: Combining non-traditional data with conventional data sources may offer a more thorough understanding of social and economic patterns. Investigating ways to efficiently merge organised and unstructured data might improve forecasts and insights.

Explainable AI: Explainability is becoming more and more important as machine learning is used in decision-making. The creation of methods that offer concise justifications for model predictions made from non-traditional data sources should be a focus for future research. This would promote trust and improve transparency.

Ethical Frameworks: As unusual data is used more frequently, it is imperative to set ethical standards for data collection, use, and sharing. Future work should focus on developing strong frameworks that leverage the power of these data sources while addressing ethical issues, data biases, and privacy concerns.

In conclusion, there are issues with data quality, ethics, and interpretation when utilising machine learning to explore unorthodox sources in big data for social and economic study. The development of machine learning methods, the integration of various data sources, the assurance of model interpretability, and the creation of moral frameworks to steer responsible data utilisation should be the main goals of future directions.

V. Conclusions:

The research on "Exploring Unconventional Sources in Big Data: A Data Lifecycle Approach for Social and Economic Analysis with Linear Regression, Random Forest, and XGBoost" concludes by presenting a thorough and structured methodology for utilising unconventional data sources to gain insightful knowledge about social and economic phenomena. The study integrates data collecting, preprocessing, feature engineering, modelling, analysis, and interpretation in a seamless manner by using a data lifecycle method, resulting in a solid framework for well-informed decision-making and academic growth. The study illustrates the adaptability of these machine learning algorithms in identifying significant patterns and relationships in various unusual data streams by using linear regression, random forest, and XGBoost. While random forest and XGBoost are excellent at capturing intricate non-linear interactions within the data, linear regression provides a foundation for comprehending linear dependencies. These models' findings offer practical understanding of social and economic patterns, empowering stakeholders to make wise decisions and policymakers to develop winning plans of action. The results highlight the potential of non-traditional data sources, including posts on social media, sensor data, and satellite images, in expanding the scope and granularity of analysis. The application of the technique to pre-existing frameworks is confirmed by the linear regression model, which offers interpretable findings that concur with conventional economic theories. The random forest and XGBoost models, meanwhile, shed light on subtle patterns that might inspire innovation and guide future research directions by exposing deep relationships that could otherwise go unnoticed. To assure accurate and significant results, the study also acknowledges some difficulties, such as the necessity for thorough feature engineering, data preparation, and model selection. Throughout the technique, ensuring ethical data usage and correcting any biases remain critical issues. In conclusion, "Exploring Unconventional Sources in Big Data" shows the revolutionary potential of a well-structured data lifecycle strategy through the integration of linear regression, random forest, and XGBoost. This method creates new opportunities for comprehending, forecasting, and influencing social and economic dynamics by utilising a variety of data sources and cutting-edge machine learning techniques. It also makes innovative contributions to the academic community while providing useful insights for decision-makers. This methodology offers a solid platform for continued investigation and innovation in the fields of social and economic analysis as technology and data accessibility continue to advance.

References

[1] R. Ashmore, R. Calinescu, and C. Paterson, "Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges," ACM Comput. Surv., vol. 54, no. 5, 2021, doi: 10.1145/3453444.

[2] A. P. Rodrigues, R. Fernandes, A. Bhandary, A. C. Shenoy, A. Shetty, and M. Anisha, "Real-Time Twitter Trend Analysis Using Big Data Analytics and Machine Learning Techniques," Wirel. Commun. Mob. Comput., vol. 2021, 2021, doi: 10.1155/2021/3920325.

[3] H. S. Munawar, S. Qayyum, F. Ullah, and S. Sepasgozar, "Big data and its applications in smart real estate and the disaster management life cycle: A systematic analysis," Big Data Cogn. Comput., vol. 4, no. 2, pp. 1–53, 2020, doi: 10.3390/bdcc4020004.

[4] 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.

[5] V. S. Kore, B. A. Tidke, and P. Chandre, "Survey of Image Retrieval Techniques and Algorithms for Image-rich Information Networks," Int. J. Comput. Appl., vol. 112, no. 6, pp. 39–42, 2015, [Online]. Available: https://www.ijcaonline.org/archives/volume112/number6/19674-1244%0Ahttp://research.ijcaonline.org/volume112/number6/pxc3901244.pdf.

[6] A. Ignatyuk, O. Liubkina, T. Murovana, and A. Magomedova, "FinTech as an innovation challenge: From big data to sustainable development," E3S Web Conf., vol. 166, 2020, doi: 10.1051/e3sconf/202016613027.

[7] D. Carvalho and R. Cruz, "Big data and machine learning in health," Eur. J. Public Health, vol. 30, no. Supplement_2, 2020, doi: 10.1093/eurpub/ckaa040.030.

[8] K. Nagorny, P. Lima-Monteiro, J. Barata, and A. W. Colombo, "Big Data Analysis in Smart Manufacturing: A Review," Int. J. Commun. Netw. Syst. Sci., vol. 10, no. 03, pp. 31–58, 2017, doi: 10.4236/ijcns.2017.103003.

[9] C. Ziliani, "The impact of big data and artificial intelligence," Loyal. Manag. From Loyal. Programs to Omnichannel Cust. Exp., pp. 76–101, 2019, doi: 10.4324/9780429022661-4.

[10] 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.

[11] K. Soomro, M. N. M. Bhutta, Z. Khan, and M. A. Tahir, "Smart city big data analytics: An advanced review," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 9, no. 5, pp. 1–52, 2019, doi: 10.1002/widm.1319.

[12] A. Jawale, P. Warole, S. Bhandare, K. Bhat, and R. Chandre, "Jeevn-Net: Brain Tumor Segmentation using Cascaded U-Net & Overall Survival Prediction," Int. Res. J. Eng. Technol., pp. 56–62, 2020.

[13] 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.

[14] L. Belcastro, R. Cantini, F. Marozzo, A. Orsino, D. Talia, and P. Trunfio, Programming big data analysis: principles and solutions, vol. 9, no. 1. Springer International Publishing, 2022.

[15] A. Y. Sun and B. R. Scanlon, "How can Big Data and machine learning benefit environment and water management: A survey of methods, applications, and future directions," Environ. Res. Lett., vol. 14, no. 7, 2019, doi: 10.1088/1748-9326/ab1b7d.

[16] 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.

[17] 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.

[18] L. Liu, "E-Commerce Personalized Recommendation Based on Machine Learning Technology," Mob. Inf. Syst., vol. 2022, 2022, doi: 10.1155/2022/1761579.

[19] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei, and A. Darwesh, "Image-based Product Recommendation Method for E-commerce Applications Using Convolutional Neural Networks," Acta Inform. Pragensia, vol. 11, no. 1, pp. 15–35, 2022, doi: 10.18267/j.aip.167.

[20] A. Tayade Somaiya Vidyavihar, A. Tayade, V. Sejpal, and A. Khivasara, "Deep Learning Based Product Recommendation System and its Applications," Int. Res. J. Eng. Technol., no. April, pp. 1317–1323, 2021, [Online]. Available: www.irjet.net.

[21] J. Sharma, K. Sharma, K. Garg, and A. K. Sharma, "Product recommendation system a comprehensive review," IOP Conf. Ser. Mater. Sci. Eng., vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012021.

[22] S. Parvatikar and D. Parasar, "Recommendation system using machine learning," Int. J. Artif. Intell. Mach. Learn., vol. 1, no. 1, p. 24, 2021, doi: 10.51483/ijaiml.1.1.2021.24-30.

[23] Z. Shahbazi, D. Hazra, S. Park, and Y. C. Byun, "Toward improving the prediction accuracy of product recommendation system using extreme gradient boosting and encoding approaches," Symmetry (Basel)., vol. 12, no. 9, 2020, doi: 10.3390/SYM12091566.

[24] F. Rodrigues and B. Ferreira, "Product Recommendation based on Shared Customer's Behaviour," Procedia Comput. Sci., vol. 100, pp. 136–146, 2016, doi: 10.1016/j.procs.2016.09.133.

[25] 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.

[26] P. R. Chandre, "Intrusion Prevention Framework for WSN using Deep CNN," vol. 12, no. 6, pp. 3567–3572, 2021.

[27] 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.