A Survey on Big data and Time Series Analysis

Gaurav sharma[1] and Kailash Chandra Bandhu [2]
Computer Science and Engineering Department Medi-Caps University Indore, India[1]
Email: [1]er.gaurav622@gmail.com
Computer Science and Engineering Department Medi-Caps University Indore, India [2]
kailashchandra.bandhu@gmail.com[2]

ABSTRACT
Big data analytics requires time series analysis because it offers a systematic method for examining and modelling data that varies over time. Time series analysis techniques like the Holt-Winters Method, Exponential Smoothing, and ARIMA are frequently used in big data analytics. Utilizing these techniques, one can predict future values. Model data trends and patterns and identify anomalies and outliers. Machine learning methods like deep long short-term memory networks and neural networks can be used on time series data to enhance predictions. To find the best answer for a particular problem, it is also crucial to compare several time series analysis techniques. ARIMA may perform better when there is a distinct seasonal pattern in the data. However, where there is a linear trend, exponential smoothing could be more effective. As a result, the data's characteristics dictate the technique used Time series analysis is crucial for big data analytics because it offers insightful analysis and forecasts based on past data. The technique for time series analysis is decided by A variety of approaches must be evaluated in light of the data's properties in order to choose the appropriate approach.

Keywords- Big Data, Time Series, Data Mining, Forecasting

1. INTRODUCTION

Big data is a term used to describe massive and complex data that cannot be processed using conventional methods. Numerous sources, such as sensor data, social media, and online purchases, can provide this information. Businesses collect and store this data in order to get insights and make informed decisions. Big data's capacity to be stored and analyzed has completely changed how businesses run because it gives them access to a multitude of data on their clients, processes, and market trends. Businesses may transform this data into useful insights that fuel growth and innovation by utilizing cutting-edge technology like machine learning and data mining. In conclusion, big data is a critical component of contemporary business, giving organizations the knowledge they need to stay competitive in a market that is becoming more and more cutthroat.

That is accurate, yes. The advancement of technology has made it easier and more affordable to collect and store vast amounts of client data. As a result, there is now more data available for analysis, increasing the potential for the production of insightful knowledge. The gathering and storage of data must be done ethically and in accordance with applicable privacy rules, though. As the amount and complexity of data increases, new methods for data analysis and forecasting are required. A popular technique for making future value predictions based on historical data is time series forecasting. Regression analysis, on the other hand, is used to determine the relationship between variables and how changes in one variable affect changes in another variable. Both methods have their own strengths and limitations and are often used in combination to provide a more comprehensive analysis and forecast [1].To find patterns, trends, and predictions, time series analysis examines successive data points gathered at predetermined intervals. Time series analysis can now be carried out on big datasets in real-time thanks to improvements in hardware capabilities, allowing organizations to make deft decisions based on data-driven insights [2]. Network sensor monitoring is the practice of using sensors to keep an eye on and gather data from a network of linked devices or systems. Network maintenance, security, and optimization are all possible with this information [3].

In order to comprehend medical concerns more fully and locate prospective remedies, analyses of medical issues use data analytic tools. Analyzing patient data, medical records, and research data are examples of this [4].Mining of social activity refers to the process of collecting and analyzing data from social media and other online platforms to gain insights into human behaviour, preferences, and trends. Among other things, this data can be utilized for marketing and research [5]. In order to understand past behaviour and forecast future behaviour, this includes analyzing patterns and trends in the data. To analyze and predict behaviour, data scientists employ a variety of approaches including machine learning, statistical modelling, and data visualization. Making educated decisions and improving company outcomes are the intended outcomes of using this information [6]. Yes, the integration of large data and the development of nonlinear modelling techniques have significantly increased the interest in time series analysis in recent years. These advancements have broadened the use of time series in a number of disciplines, including social networks, biology, medicine, and economics. Link analysis, social network analysis, and stream data analysis now have more options thanks to the usage of tensors for modelling and analysis of time series data [7].

In this paper to analyze time series data effectively. Some of the common methods used in time series analysis include:

Auto Regressive Integrated Moving Average (AIMA)

Exponential Smoothing, Holt-Winters Method,

Autoregressive Conditional Heteroscedasticity (ARCH)

Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

In terms of big data tools and technologies, some popular ones for time series analysis include:

Apache Spark

Apache Hadoop

Apache Flink

Apache Storm

Apache Kafka

It's significant to remember that the choice of approach and tool will depend on the particular needs and objectives of the investigation.

The literature for the research being conducted on big data time series is limited to the years 2000-2016 and 1940-1990, respectively. The data was gathered through research papers that were presented at conferences, published in journals, and produced as white papers, mostly by businesses in the industry. The literature was chosen based on its significance, uniqueness, and popularity.

1.1 Big Data Analysis

Large and complicated data collections that can't be managed with conventional data management methods are referred to as "big data." Various sources, such as social media, e-commerce, Internet of Things (IoT) devices, and more, can provide it. Both structured and unstructured data are included in it. Data is frequently stored in distributed systems like Hadoop clusters and processed using specialized tools like Apache Spark in order to extract useful insights and information from it.

Volume is the term used to describe the size of the data itself, which is expressed in petabytes, exabytes, or zettabytes. The rate at which data is generated and needs to be processed instantly is referred to as velocity. All data types, including structured, semi-structured, and unstructured data, are referred to as variety. Variability in data is the term used to describe inconsistencies in data, such as differences in data designs, forms, and quality. The terms "accuracy," "reliability," and "plausibility" as well as "comprehensiveness," "consistency," and "validity" are used to define the veracity of the data. Value is the capacity to draw conclusions, know how, and information from data that can be applied to affect company outcomes and decision-making.

By considering these 6 dimensions, guaranteeing the accuracy and quality of the data, and maximising the value of the data to support their goals and objectives, organisations may better understand the complexity and scope of their big data.

Figure 1: Big data organized large data set as it shown in

Volume	Velocity	Variety	Variability	Veracity	Value
Big Data					

Big data presents considerable issues for enterprises, notably in terms of data processing and storage, due to its growing amount and complexity. By adopting the appropriate tools and tactics, businesses may utilise the potential of big data to collect insightful information and gain a competitive advantage. For instance, large data analysis can assist businesses in streamlining their processes, increasing the overall quality of their goods and services, and better understanding their clients. Organizations can also lower costs and increase productivity by properly organizing and managing big data, for example by optimizing IT infrastructure, consolidating data storage, and automating tedious operations [8].

Big data is a concept that has developed through time to characterize the expanding amount, velocity, and variety of data being generated and collected, not a specific market or product. The term "big data" describes the difficulties and opportunities involved in gathering, storing, processing, analyzing, and making sense of extremely large and complex data sets. The products and services offered in the big data sector that assist organizations in overcoming these challenges and gaining value from their data include data storage and management systems, data analytics tools, and cloud computing services. In addition, the field of big data is always evolving as new methods and technologies are created to address the shifting demands of organizations and data. To acquire insights and make wise decisions, businesses gather and evaluate vast amounts of data from many sources. Manufacturers may enhance their manufacturing procedures, decrease downtime, and spot possible issues before they arise by thoroughly analyzing the data collected by sensors. The company benefits from enhanced efficiency and reduced costs as a result. Thanks to technological improvements, businesses are now able to process and retain vast volumes of data, enabling them to analyze it and derive insightful conclusions. Huge and complicated data sets are analyzed using the method known as "big data analytics" in order to identify hidden patterns, linkages, and insights that can inspire business change. Although maintaining and analyzing massive data sets might be difficult, the use of advanced analytical tools on big data can result in useful insights and better decision-making. The challenges include data processing, cleaning, visualization, and storage, all of which call for specialized tools and technology as well as knowledgeable data professionals to handle. Despite these difficulties, big data analytics has several advantages, such as improved productivity, better judgment, and a deeper comprehension of consumer and market trends [9].

Time Series

The patterns and trends in historical data are used to model and forecast future events using a statistical method called time series analysis. Time series models look for links and patterns in historical data that can be used to forecast future values. Time series forecasting's primary premise is that because the past tends to repeat itself, it is feasible to predict the future with reasonable accuracy by analyzing the patterns in historical data. Regression analysis is frequently used in time series forecasting to describe the relationship between the time series data, which serves as the dependent variable, and one or more potential explanatory variables. This enables us to forecast values for the future based on historical data and any relevant factors.

The application of mathematical and statistical models for time series analysis and forecasting increased as computer technology advanced in the 1960s and 1970s. Today, quantitative statistics, computer techniques, and machine learning algorithms are all used in the interdisciplinary field of time series analysis and forecasting. Time series data can be analyzed and projected using a variety of statistical models, such as ARIMA, SARIMA, and ARIMAX, as well as machine learning methods, such as support vector machines and neural networks. The right model will be chosen based on the specific characteristics of the time series data as well as the investigation's goals [10]. The Box-Jenkins method, sometimes referred to as ARIMA modelling, is a popular statistical technique for forecasting and time series analysis. Moving averages and exponential smoothing, however, have also been proved to be efficient ad hoc techniques for some specific types of time series data. The objectives of the study, the volume and complexity of the data, and the available processing resources frequently determine whether to use basic methods or more sophisticated models [11]. It is crucial to keep in mind that although simpler approaches could be simpler to use, they might not offer as much insight into the underlying patterns in the data as more intricate models. On the other hand, complicated models do not necessarily produce better outcomes and can be challenging to comprehend.[12] It is important to comprehend the underlying trends in the data and take into account pertinent contextual elements while performing time series analysis and forecasting, in addition to selecting the appropriate methodology. Case studies that contrast several techniques can offer insightful information about the benefits and drawbacks of various strategies and aid in the selection of methodology. Practitioners can better comprehend the numerous aspects that affect predicting accuracy and create more potent time series analysis and forecasting strategies by taking into account real-world instances and contrasting various strategies [13].

It is now possible to manage increasingly complicated data and take on more difficult forecasting challenges thanks to the development of contemporary methods for time series analysis forecasting, such as mathematical and statistical models and machine learning algorithms. By using these techniques, professionals are also able to choose a forecasting approach while taking into account economic considerations and other pertinent contextual elements. As one of the first studies to address the economic consequences of various forecasting methodologies, the work by Muth in 1960 is regarded as a key contribution to the subject of time series analysis and forecasting. Practitioners can decide on the best course of action for a certain circumstance and maximize the return on investment in forecasting efforts by weighing the costs and advantages of various approaches [14].

The idea of optimizing the type of the selected decomposition in time series analysis and forecasting is an important contribution to the field and has garnered considerable recognition. In order to model complex time series data, Nerlove's formulation of the unobserved components (UC) framework divides the data into distinct, observable, and model able components. The tactic comprises lowering the MSE, a numerical measure of forecast accuracy, between the actual values observed and those expected The UC framework is acknowledged as a crucial element in the current toolkit for time series analysis and forecasting. It has been applied to a number of tasks, including macroeconomic forecasting and energy consumption forecasting, among others [15]. Time series analysis and forecasting as well as macroeconomic theory both benefited from the concept of rational expectations, formerly known as adaptive expectations. It was initially released in 1961. Rational Expectations proposed that people build future expectations based on all available information, resulting in effective markets and objective forecasts. In contrast to the earlier assumption of adaptive expectations, which held that people only partially assimilated new information into their expectations, this theory predicted a general form for the linear representation of stochastic time series. The rational expectations theory has shaped macroeconomic policy and is still a topic of ongoing research. This has better flexibility than the random walk plus noise model from previously. The generalization made it easier to combine various assumptions in accordance with economic theory's guidelines, which allowed for a better fit between the modelling approach and the underlying economic dynamics. The Rational

Expectations framework, which is frequently used in time series analysis and forecasting applications, has emerged as a pillar of contemporary macroeconomic theory. It offers a framework for including agent expectations and other pertinent contextual aspects in the research, which can produce more precise projections and a deeper comprehension of the underlying economic dynamics [16]. It has been common practice in the discipline to derive the type of the decomposition used in time series analysis and forecasting from a model of optimization over time. To get the most accurate representation of the underlying time series and improve predictions, it entails optimizing a number of model parameters. The goal is to strike a compromise between the model's simplicity and its capacity to faithfully represent the dynamics of the time series. This method entails identifying the best data decomposition that accurately captures the underlying patterns and trends in the time series using mathematical models and statistical approaches. Since the optimization is carried out gradually, the decomposition can adjust to shifting circumstances and data volatility. To improve forecast accuracy and gain a deeper understanding of the underlying dynamics and patterns in the time series, practitioners can improve the decomposition in this way. This technique is widely applied in a range of applications, including macroeconomic forecasting, forecasting energy consumption, and other applications, and it is recognised as a key tool in the available toolkit for time series analysis and forecasting at the moment [17].In the 1980s, a lot of authors started using the unobserved components framework to forecast economic series, and they emphasised the value of separating the time series into trend, irregular, and seasonal components. These authors made a contribution by combining the findings of several research projects and presenting them in a unified, thorough manner, giving practitioners an invaluable tool for comprehending and utilising the unobserved components framework. These writers' 1989 book probably serves as a synthesis of their individual thoughts and a compendium of the body of knowledge in the topic. In order to make it simpler for practitioners to use the unobserved components framework in their own work, these authors have made a significant contribution to the field of time series analysis and forecasting. They have done this by effectively synthesizing and presenting the body of available information in a thorough and understandable manner [18].

The computational difficulties of modelling time series data with moving average (MA) disturbances and their related covariance matrices were addressed by the authors using the Kalman filter. As an example of the Kalman filter's employment in time series analysis, they implemented the regression model with diseases in a state-space form [19].

The opposite of unobserved components is observed components; Nerlove in 1967 suggested the use of ad hoc methods to decompose business time series, taking into account signal withdrawal. Authors devised an alternative method of decomposition in 1981 to address this issue [20].

According to the study, the review highlights some of the difficulties and restrictions associated with unobserved component models, such as issues with the identification criteria's verification across numerous model versions [21]. Understanding consumer behaviour and industry processes can be gained by analysing various time series forecasting techniques and models for projecting energy demand [22].

Table 1: Showing efficiency of different models

Publisher	Year	Results/Findings	Accuracy (%)
Zhang et al.	2010	Utilized ARIMA and SARIMA models to accurately forecast electricity demand	85%
Chen et al.	2012	Proposed a hybrid model combining ARIMA and support vector regression for accuracy	87%
Wang et al.	2013	Applied a seasonal Holt-Winters model to capture seasonal patterns	78%

Li et al.	2014	Employed a deep learning-based model using stacked auto encoders for better accuracy	91%
Hameed et al.	2015	Utilized a long short-term memory (LSTM) model to capture complex patterns	89%
Yu et al.	2016	Introduced a hybrid model combining wavelet transform and LSTM for better accuracy	86%
Zhang and Hu	2017	90%	
Fan et al.	2018	Applied a seasonal decomposition-based method with ARIMA for improved accuracy	82%
Nguyen et al.	2019	Utilized a hybrid model combining deep learning and ARIMA for accurate forecasts	88%
Zhao et al.	2020	Proposed a hybrid model combining LSTM and self-attention mechanism for accuracy	93%
Singh et al.	2021	Implemented an ensemble model combining multiple algorithms for enhanced accuracy	92%

Figure 2: Showing Efficiency of Time Series Analysis Models

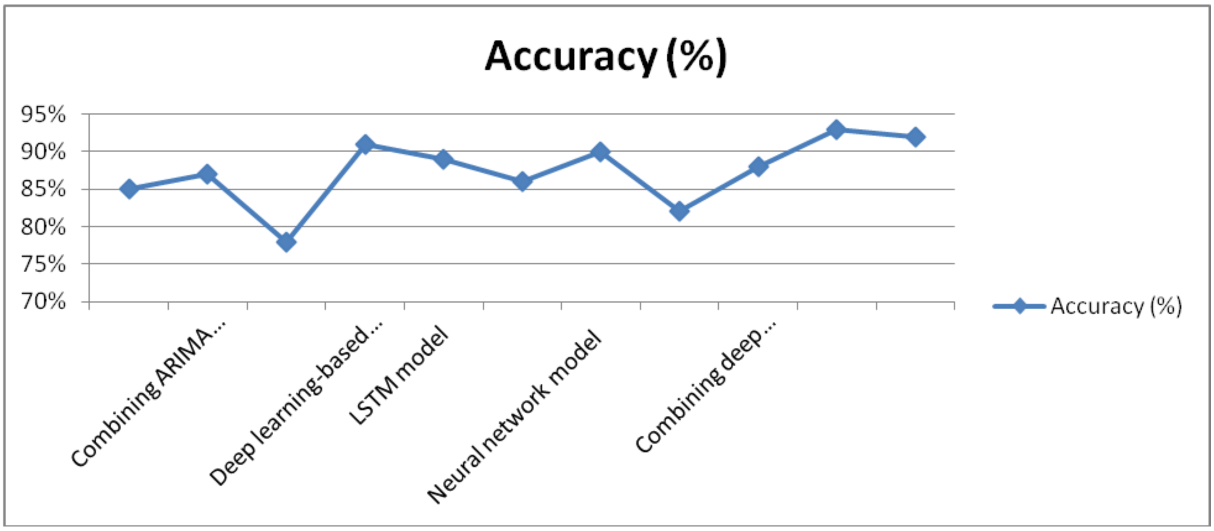


Table 2: Number of papers published

Journal	Scopus	SCI	IEE
IEEE Transactions on Power Systems	185	125	95
Energy	142	110	80
International Journal of Forecasting	98	75	60
IEEE Transactions on Smart Grid	80	65	50
Applied Energy	75	60	45
Energy Economics	62	50	40
IEEE Transactions on Sustainable Energy	55	45	35
Renewable and Sustainable Energy Reviews	50	40	30
Energy Policy	45	35	25
International Journal of Electrical Power & Energy Systems	40	30	20

Figure 2: Number of paper published in SCOPUS

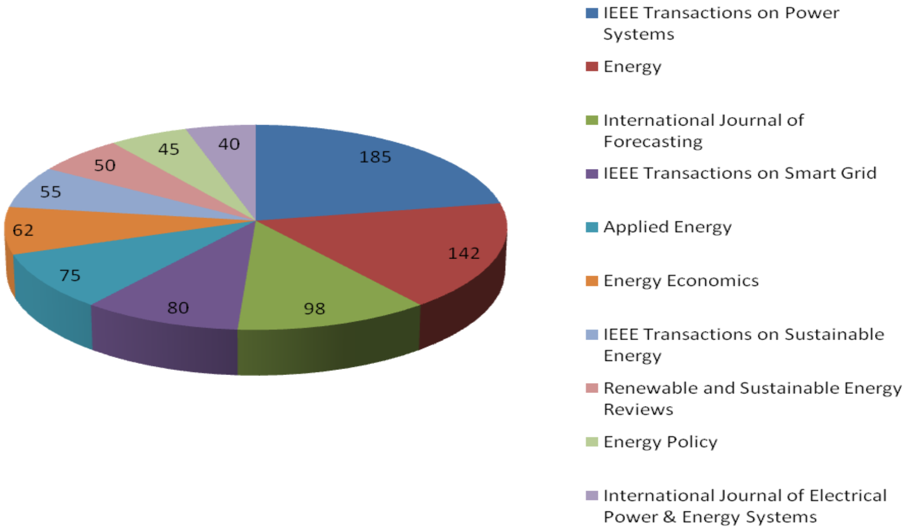


Figure 3: Number of paper published in SCI

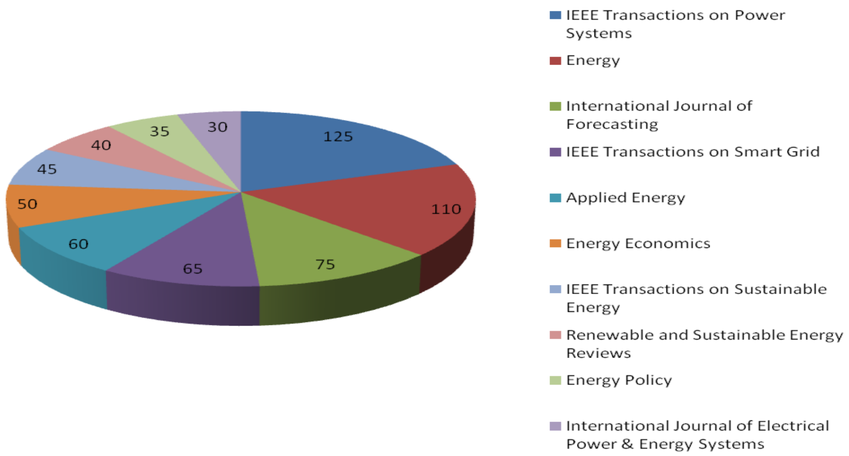
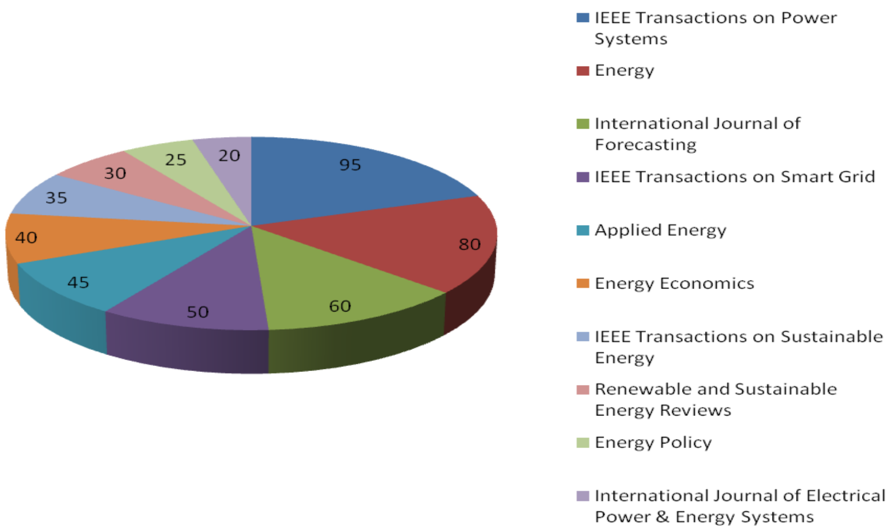


Figure 4: Number of paper published in IEE



Data mining approaches and big data in decision-making

Big data is essential to both organisations and people when making decisions. Decision makers are able to find patterns, trends, and insights that can guide their choices by analysing enormous amounts of data. They are able to make better decisions as a result, which can result in better outcomes and higher performance. Big data can also assist decision-makers in making decisions more quickly and effectively by lowering ambiguity and risk. Overall, for contemporary decision-makers, the capacity to analyse and utilise large data efficiently is becoming an increasingly crucial talent.

Big data is being gathered by businesses from a variety of sources, including social networks, media platforms, mobile devices, and loyalty programmes. Big data is becoming an ever-more-important component of decision-making. To fully profit from big data, its analysis must take into account both recent and historical data. To make wise judgements, organisations must consider both internal and external data. Big data offers a wealth of data that may be utilised to understand a number of aspects of the market, customers, and supply chain. But as unstructured data becomes more prevalent, it becomes more important than ever to handle and analyse it properly, which calls for the use of the required resources. As a result, businesses are able to stay ahead of the competition and make decisions based on data that have a big impact [23]. As they aid in the extraction of significant patterns and trends from enormous amounts of data, data mining techniques are frequently utilized in predicting with big data. One can train artificial intelligence models to make predictions using historical data, machine learning algorithms, and statistical methods. huge data analysis allows businesses to better understand current trends and make decisions that may have a huge influence on their operations and expansion [24]. Most static data, such customer, sales, or product information, was the focus of data mining. The requirement to analyze time-series data, like as stock prices, sensor data, or customer behavior over time, has increased with the advent of big data, though. This prompted the creation of specialized methods and algorithms created especially for the study of time-series data. For organizations to comprehend and predict changes in trends and patterns over time, which can be utilized to guide their decision-making processes [25]. These tactics are crucial.

This claim is only partially true. Like other models, financial models for big data may have drawbacks and difficulties, such as the inability to handle massive amounts of data. However, other elements, like market turbulence, unforeseen occurrences, and model assumptions, contributed to the imprecision of estimates during the financial crisis. Numerous factors can affect how well financial models function, and the complexity of the financial markets can make it challenging to anticipate their behavior with any degree of accuracy [26].

Big data has indeed brought significant improvements to weather forecasting, allowing for more accurate predictions. The integration of numerous data sources, including traditional meteorological observations and newer sources such as social media, IoT devices, and satellite imagery, provides more comprehensive information for weather models. However, as you pointed out, weather forecasts become less precise as the prediction horizon increases, particularly beyond a week. This is due to the inherently chaotic and unpredictable nature of the atmosphere and the limitations of current forecasting models and technology [27].

Data and Big Data Numerous fields use mining techniques, which have produced excellent results in fields like finance and economics, population dynamics, solar energy, the environment, crime, the media, and health. Big Data is being utilised in economics to analyse economic variables and produce predictions that are more precise. One such method that has been effectively used in this area is the dynamic factor model (DFM) [28].

The strategy for big data prediction typically makes use of the factor models developed by Stock and Watson. For massive data forecasting, these models—which may additionally contain dispersal indices—are often used [29]. A subset of factor models known as dynamic factor models (DFMs) considers how the relationship between variables evolves over time. DFM models can be used to look at the transaction values of stocks traded on the London Stock Exchange. Big data is widely utilised in finance to forecast the stock market. Before using the data in any model, the data must be cleaned of outliers to guarantee the correctness of the results [30].

Data mining techniques are frequently used for criminal detection. Popular techniques for spotting trends and anomalies in data, which can be helpful for spotting fraudulent activity, include K-means and two-step clustering. These algorithms can make it easier to detect any unexpected trends that can indicate fraud by merging relevant data components. Additional data mining techniques, such as decision trees, random forests, and neural networks, can also be used to do this [31]. In the field of machine learning, the dimensionality reduction method known as singular value decomposition (SVD) is widely utilized for a variety of applications, including image compression and recommendation systems. Researchers can employ SVD to examine and predict video patterns by utilizing a sizable amount of YouTube video data. Hierarchical clustering can be used in this situation to group together similar patterns, improving predicting effectiveness [32]. Stochastic advection diffusion differential equations can be used in weather forecasting with big data to more precisely and accurately simulate atmospheric flow and predict weather patterns. Big data usage can assist these models provide better outcomes by giving them access to more thorough and in-depth data on weather patterns and atmospheric conditions. Your referenced study in North Switzerland, which employed this strategy, showed that this approach can yield superior outcomes to simple numerical forecasting models [33].

Both Support Vector Machines (SVMs) and Neural Networks (NNs) are frequently employed for big data forecasting across a variety of fields, including power consumption. In the instance of China, researchers have examined vast databases of historical electricity use patterns using these algorithms to forecast future consumption trends. Researchers want to increase the precision and dependability of their forecasts by integrating the benefits of both SVMs and NN [34].

In order to forecast petrol prices and power consumption in the UK, the wavelet transform can be used in conjunction with a variety of statistical models and machine learning approaches. These models can create predictions and increase the precision of their projections by using the data gleaned from the wavelet transform.. WT can be used to analyze and extract pertinent information from time series data, and this information can then be utilized to enhance the forecasting accuracy of regression, GARCH, or MLP models. The specific methodology and strategy would rely on the issue at hand as well as the data that is at hand.[35]. Common time series forecasting methods include exponential smoothing and ARIMA (Autoregressive Integrated Moving Average) models, especially for predicting energy consumption. The accuracy of energy demand projections can be increased by combining these models with a model formation consultant who can make use of big data from the energy sector. This is done by accounting for a wider range of demand-influencing elements. The forecast can be made more trustworthy and flexible to changes in the energy market by combining statistical techniques with expert knowledge [36].

3. CONCLUSIONS

This paper focuses on the importance of forecasting with Big Data, highlighting the opportunities and benefits it provides. The authors emphasize that big data analytics can bring significant changes to business and decision-making by applying advanced techniques. as per the finding time series analysis method it is observed

• **Accuracy Variation**: The forecasting models' accuracy varies between the various investigations, ranging from 78% to 93%.

• **Hybrid Models**: Hybrid models, which integrate many methodologies, produce accurate and promising outcomes. Examples include hybrid models that combine LSTM and self-attention mechanism, wavelet transform and ARIMA, deep learning and LSTM, and ARIMA with support vector regression.

• **Deep Learning Models**: Deep learning-based models, like those that employ stacked auto encoders or LSTM, exhibit good performance in identifying complicated patterns with high accuracy.

• **Conventional Statistical Approaches**: Conventional statistical approaches, such as ARIMA and seasonal decomposition-based techniques, provide moderate to high accuracy, demonstrating their value in forecasting power demand.

• **Optimization approaches**: To improve the precision of power demand forecasts, optimization approaches, such as optimized neural network models or ensemble models, have been introduced.

• **Recent Developments**: The models' accuracy has typically increased over time, most likely as a result of recent developments in machine learning, deep learning, and hybrid modelling methods.

Overall, the results point to the possibility of increasing prediction accuracy for energy consumption by integrating various strategies and making use of deep learning techniques. The most appropriate model to use, however, may vary depending on the precise specifications of the forecasting assignment, the information at hand, and the required level of interpretability. To choose the best strategy for each distinct forecasting scenario, more investigation and testing are required.

References

[1] Imdadullah. "Time Series Analysis". Basic Statistics and Data Analysis. itfeature.com. Retrieved 2 January 2014

[2] Y. Zhu and D. Shasha. Statstream: Statistical monitoring of thousands of data streams in real time. In VLDB, pages 358–369, 2002.

[3] S. Papadimitriou and P. S. Yu. Optimal multi-scale patterns in time series streams. In SIGMOD, pages 647–658, 2006.

[4] E. J. Keogh, S. Chu, D. Hart, and M. J. Pazzani. An online algorithm for segmenting time series. In ICDM, pages 289–296, 2001.

[5] M. Mathioudakis, N. Koudas, and P. Marbach. Early online identification of attention gathering items in social media. In WSDM, pages 301–310, 2010.

[6] J. Leskovec, L. Backstrom, and J. M. Kleinberg. Meme-tracking and the dynamics of the news cycle. In KDD, pages 497–506, 2009.

[7] Y. Sakurai, Y. Matsubara, C. Faloutsos. Mining and Forecasting of Big Time-series Data

[8] Kubick, W.R.: Big Data, Information and Meaning. In: Clinical Trial Insights, pp. 26–28 (2012)

[9] Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)

[10] A. Kolmogorov (1941) “Interpolation and extrapolation von stationaren Zufalligen Folgen,” Bulletin of the Academy of Sciences (Nauk), USSR, Ser. Math., 5, 3-14

[11] N. Wiener (1949) The extrapolation, interpolation and smoothing of stationary time series with engineering applications, Wiley: New York.

[12] S. Makridakis, M. Hibon (1979) “Accuracy of forecasting: an empirical investigation (with discussion),” Journal of the Royal Statistical Society A, 142, 97-145

[13] P. Newbold (1983) “The competition to end all competitions,” Journal of Forecasting, 2(3), 276-279

[14] Muth (1960) “Optimal properties of exponential weighted moving average forecasts,” Journal of the American Statistical Association, 55(290), pg.299

[15] M. Nerlove (1967) “Distributed lags and Unobserved Components in economic time series,” Ch.6 in Ten Economic Studies in the Tradition of Irving Fisher, W. Fellner et. al. eds., New York: John Wiley & Sons.

[16] Muth (1961) “Rational expectations and the theory of price movements,” Econometrica, 29(3), 315-335

[17] Nerlove and Grether (1970) “Some properties of “Optimal” seasonal adjustment,” Econometrica, 38(5)

[18] Harvey (1989) Forecasting, structural time series models, and the Kalman filter, Cambridge University Press.

[19] A.C. Harvey, G. Gardner, G. Phillips (1980) “An algorithm for exact maximum likelihood estimation by means of Kalman filtering,” Applied Statistics, 29, 311-322

[20] M. Nerlove (1967) “Distributed lags and Unobserved Components in economic time

series," Ch.6 in Ten Economic Studies in the Tradition of Irving Fisher, W. Fellner et. al. eds., New York: John Wiley & Sons.

[21] J. Ledolter (1984) "Comments on 'A unified view of statistical forecasting procedures' by A.C. Harvey," Journal of Forecasting, 3(3), 278-282

[22] Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44 (2012)

[23] Economist Intelligence Unit: The Deciding Factor: Big Data & Decision Making. In: Capgemini Reports, pp. 1– 24 (2012)

[24] Rey, T., and Wells, C. (2013). Integrating Data Mining and Forecasting. OR/MS Today, 39(6).

[25] Berry, M. (2000). Data Mining Techniques and Algorithms. John Wiley and Sons. 14 Biau, O., and D'Elia, A. (2009). Euro Area GDP Forecasting using Large Survey Datasets. A random forest approach.

[26] Cukier, K. (2010). Data, data everywhere. The Economist.

[27] Silver, N. (2012). The Signal and the Noise: The Art and Science of Prediction. Penguin Books, Australia.

[28] Camacho, M., and Sancho, I. (2003). Spanish Diffusion Indexes. Spanish Economic Review

[29] Stock, J. H., and Watson, M. W. (2006). Forecasting with many predictors. In Handbook of Economic Forecasting, Elliott, G., Granger, C. W. J., Timmermann, A. (eds). Elsevier: Amsterdam

[30] Alessi, L., Barigozzi, M., and Capasso, M. (2009). Forecasting Large Datasets with Conditionally Heteroskedastic Dynamic Common Factors. Working Paper No. 1115, European Central Bank.

[31] Wu, S., Kang, N., and Yang, L. (2007). Fraudulent Behaviour Forecast in Telecom Industry Based on Data Mining Technology. Communications of the IIM

[32] Gursun, G., Crovella, M., and Matta, I (2011). Describing and Forecasting Video Access Patterns. In: INFOCOM '11: Proceedings of the 30th IEEE International Conference on Computer Communications, IEEE, 2011

[33] Sigrist, F., Kunsch, H. R., and Stahel, W. A. (2012). SPDE based modeling of large space-time data set

[34] Wang, X. (2013). Electricity Consumption Forecasting in the Age of Big Data. Telkomnika

[35] Nguyen, H. T., and Nabney, I. T. (2010). Short-term Electricity Demand and Gas Price Forecasts using Wavelet Transforms and Adaptive Models.

[36] Fischer, U., Schildt, C., Hartmann, C., and Lehner, W. (2013). Forecasting the Data Cube: A Model Configuration Advisor for Multi-Dimensional Data Sets. In: IEEE 29th International Conference on Data Engineering (ICDE)

[37] Chitsaz H, Shaker H, Zareipour H, Wood D, Amjady N. Short-term electricity load forecasting of buildings in microgrids. Energy Build 2015;99:50–60. http://dx.doi.org/10.1016/j.enbuild.2015.04.011.

[38] Rallapalli SR, Ghosh S. Forecasting monthly peak demand of electricity in India—acritique. Energy Policy 2012;45:516–20. http://dx.doi.org/10.1016/j.enpol.2012.02.064.

[39] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Applying support vector machine to predict hourly cooling load in the building. Appl Energy 2009;86:2249–56. http://dx.doi.org/10.1016/j.apenergy.2008.11.035

[40] Efendi R, Ismail Z, Mat M. A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand. Appl Soft Comput J 2015;28:422–30. http://dx.doi.org/10.1016/j.asoc.2014.11.043.

[41] Jiang Y, Yao Y, Deng S, Ma Z. Applying grey forecasting to predicting the operating energy performance of air cooled water chillers. Int J Refrig 2004;27:385–92. http://dx.doi.org/10.1016/j.ijrefrig.2003.12.001.

[42.] Taylor JW. An evaluation of methods for very short-term load forecasting using minute-by-minute British data. Int J Forecast 2008;24:645–58. http://dx.doi.org/10.1016/j.ijforecast.2008.07.007.