Grenze International Journal of Engineering and Technology, June Issue

A Survey on Big Data and Time-Series Analysis

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

***Abstract*— Time series analysis is needed for big data analytics since it organizes and models time-varying data. Big data analytics uses Holt-Winters, Exponential Smoothing, and ARIMA time series analysis. One can anticipate future values using these methods. Model data patterns and find outliers. Deep long, and short-term memory networks and neural networks can improve time series forecasts. Comparing multiple time series analysis methods is essential to finding the optimal solution. When data is seasonal, ARIMA may perform better. If the trend is linear, exponential smoothing may work better. Data qualities determine the technique utilized. Big data analytics requires time series analysis for meaningful analysis and forecasts based on historical data. The technique for time series analysis is determined by Choosing the right strategy by evaluating several approaches based on data attributes.**

***Index Terms*— Big data, Time series, Data mining, Forecasting.**

I. 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 to get insights and make informed decisions. Big data's capacity to be stored and analysed has completely changed how businesses run because it gives them access to many 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.
This is true. Technology makes collecting and storing massive volumes of client data cheaper and easier. Thus, more data may be analyzed, enhancing the potential for insightful knowledge. However, data collection and storage must be ethical and compliant with privacy laws. Data analysis and forecasting require new methodologies as data volume and complexity rise. Time series forecasting is a popular method for predicting future values from previous data. However, regression analysis determines how changes in one variable affect changes in another. Both methods have pros and cons and are sometimes used together for a more complete study and forecast [1]. Time series analysis analyses data points at regular intervals to uncover patterns, trends, and forecasts. Hardware developments enable real-time time series analysis on large datasets, enabling organisations to make data-driven decisions [2]. Network sensor monitoring uses sensors to monitor and collect data from linked devices or systems. This data allows network maintenance, security, and optimisation [3]. Medical difficulties are analysed using data analytic technologies to better understand and find solutions. Examples include patient, medical, and research data analysis [4]. Mining social activity involves gathering and analysing data from social media and other online

Grenze ID: 01.GIJET.10.2.112

platforms to understand human behaviour, preferences, and trends. This data is useful for marketing and research [5]. To understand previous behaviour and predict future behaviour, analyse data patterns and trends. Data scientists use machine learning, statistical modelling, and data visualisation to predict behaviour. This information should help you make better judgements and boost company performance [6]. Indeed, big data integration and nonlinear models have boosted time series analysis attention in recent years. These advances have expanded time series use in social networks, biology, medicine, and economics. Link, social network, and stream data analysis are now more flexible thanks to tensors for time series data modelling and analysis [7].
Effective time series data analysis is presented here. Common methods for 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 important to keep in mind that the specific requirements and goals of the inquiry will determine which strategy and instrument are best. The research on large data time series is based solely on literature from the years 1940-1990 and 2000-2016, respectively. Most of the industry's companies generated white papers, research papers that were published in journals, presented at conferences, and created as white papers using the data. The selection of the literature was based on its popularity, relevance, and originality.

BIG DATA ANALYSIS

"Big data" refers to complex and sizable data sets that are too large to handle with traditional data management techniques. It can be obtained from a variety of sources, including social media, e-commerce, Internet of Things (IoT) devices, and more. It contains both structured and unstructured data. To extract valuable insights and information from data, it is commonly stored in distributed systems like Hadoop clusters and processed using specialized tools like Apache Spark. The magnitude of the data itself, measured in petabytes, exabytes, or zettabytes, is referred to as its volume. Data is created quickly and must be processed immediately. Variety encompasses unstructured, semi-structured, and structured material. Inconsistencies in data formats, designs, and quality are called "variability in data". Data veracity is described by "comprehensiveness," "consistency," "validity," "accuracy," "reliability," and "plausibility." Value is inferring knowledge, information, and conclusions from data to affect business decisions and outcomes.
These six aspects can help organizations comprehend the complexity and scope of their big data, ensure its quality and accuracy, and maximize its value to support their goals.

Volume	Velocity	Variety	Variability	Veracity	Value
Big Data					

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

Due to its volume and complexity, big data poses challenges for organizations in processing and storage. Businesses may use big data to obtain insights and a competitive edge by using the right tools and strategies. Large data analysis can help firms streamline procedures, improve product quality, and better understand customers. Optimising IT infrastructure, consolidating data storage, and automating laborious tasks can help organisations organise and manage big data to reduce costs and boost productivity [8].
Big data is a concept that describes the growing amount, velocity, and variety of data being generated and collected, not a market or commodity. Gathering, storing, processing, analysing, and making sense of massive and complicated data sets is called "big data" and involves many challenges and opportunities. Data storage and management systems, data analytics tools, and cloud computing services help organisations overcome these obstacles and derive value from their data. Big data also evolves as new approaches and technologies meet

changing organisational and data needs. Businesses analyse massive volumes of data from numerous sources to gain insights and make smart decisions. By analysing sensor data, manufacturers may improve processes, reduce downtime, and anticipate faults. The company gains efficiency and lowers costs. Technology allows firms to handle and store massive amounts of data, allowing them to analyse and draw conclusions. Large, complex data sets are analysed using "big data analytics" to find hidden patterns, relationships, and insights that might inspire business change. Advanced analytical techniques on big data can improve decision-making and insights, despite the challenges of storing and analysing enormous data sets. Data processing, cleaning, visualisation, and storage require specialised tools and highly skilled data specialists. Big data analytics improves efficiency, judgement, and consumer and market trends despite these challenges [9].

TIME SERIES

Time series analysis is a statistical technique used to model and estimate future occurrences based on patterns and trends in historical data. Time series models search through historical data for relationships and trends that can be applied to project values in the future. The main idea behind time series forecasting is that as history repeats itself, it is possible to estimate the future very accurately by looking for patterns in previous data. Time series forecasting commonly uses regression analysis to characterize the relationship between the dependent variable, the time series data, and one or more possible explanatory variables. This gives us the ability to predict values for the future using past data and any pertinent variables.

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 analysed 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]. Box-Jenkins, or ARIMA modelling, is a common statistical tool for predicting and time series analysis. Moving averages and exponential smoothing are also effective ad hoc methods for some time series data. The study's goals, data amount and complexity, and processing resources often determine whether to utilise simple or more advanced models [11]. Although simpler techniques may be easier to employ, they may not provide as much insight into data patterns as more complex models. However, complex models may not generate superior results and are harder to understand.[12] Time series analysis and forecasting require understanding data trends and contextual factors, as well as choosing the right methodology. Case studies that compare methods can help choose a method by revealing their pros and cons. Real-world examples and strategy comparisons can help practitioners understand the many factors that affect predicting accuracy and develop more effective time series analysis and forecasting procedures [13].

Modern time series analysis forecasting methods like mathematical and statistical models and machine learning algorithms may handle more complex data and forecasting issues. These methods allow professionals to choose a forecasting approach that takes economic factors and other contextual factors into account. Muth's 1960 study on time series analysis and forecasting was one of the first to examine the economic effects of forecasting methods. Weighting the costs and benefits of several forecasting methods helps practitioners choose the optimal one for a given situation and maximize ROI [14].

Optimizing decomposition type in time series analysis and forecasting is a significant contribution to the field. The unobserved components (UC) paradigm of Nerlove splits complex time series data into distinct, observable, and model able components. The strategy lowers the MSE, a numerical measure of forecast accuracy, between observed and expected values. UC is a key tool for time series research and forecasting. Macroeconomic and energy consumption predictions are among its uses [15]. Rational expectations (previously adaptive expectations) helped time series analysis, forecasting, and macroeconomic theory. It debuted in 1961. Rational Expectations recommended that people base future expectations on all available knowledge, creating effective markets and objective projections. Unlike adaptive expectations, which assumed that people only partially integrated new information into their expectations, this theory predicted a general form for stochastic time series linear representation. The rational expectations theory influences macroeconomic policy and is still studied. This is more flexible than the random walk plus noise model. The generalisation made it easy to combine assumptions according to economic theory, which improved the modelling approach's fit with economic dynamics. A cornerstone of modern macroeconomic theory is the Rational Expectations framework, employed in time series analysis and forecasting. It provides a framework for adding agent expectations and other contextual factors in research, which can improve projections and understanding of economic dynamics [16]. Most time series analysis and prediction decomposition use optimisation models. To accurately characterise time series and enhance predictions, tune model parameters. The model must balance time series dynamics accuracy and simplicity. This technique uses mathematical models and statistics to discover the proper data breakdown for time series patterns and trends. Decomposition adapts to data volatility and changing situations with incremental optimisation. Decomposition improves forecast accuracy and helps practitioners grasp time series patterns. Time series analysis and macroeconomic and energy consumption forecasting are possible with this method [17]. Many 1980s authors projected economic series using the unobserved components framework and underlined the significance of dividing time series into trend, irregular, and seasonal components. These writers helped practitioners comprehend and implement the unobserved components paradigm by synthesising several research programmes' findings. Their 1989 book may be a compilation of their views and relevant expertise. By simplifying the unobserved components paradigm, these authors advanced time series analysis and forecasting. Using data synthesis and explanation [18], they did this. 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, considering 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 THE 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 autoencoders 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%
Hamid, Ghous	2023	ANN, evaluation, matrices, ARIMA	86 %

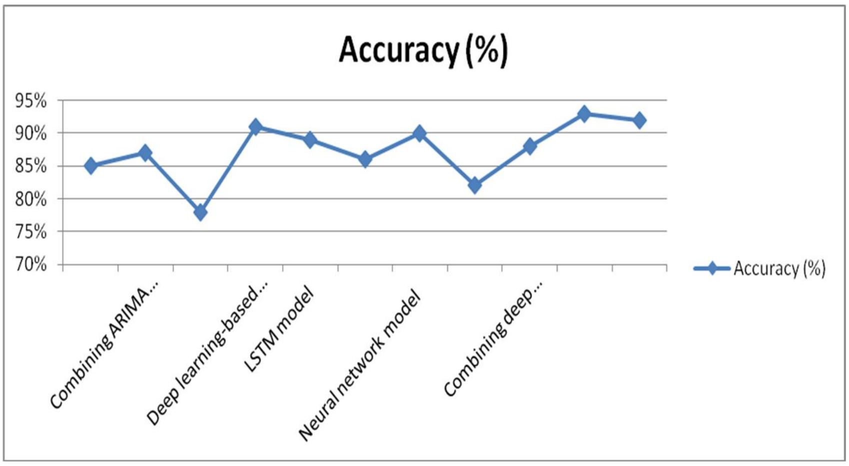

Figure 2: Showing Efficiency of Time Series Analysis Models

TABLE II: 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

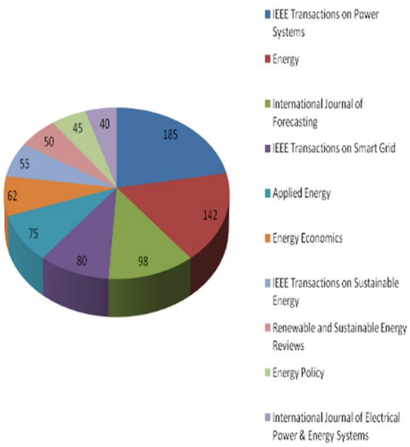

Figure 3: Number of papers published in SCOPUS

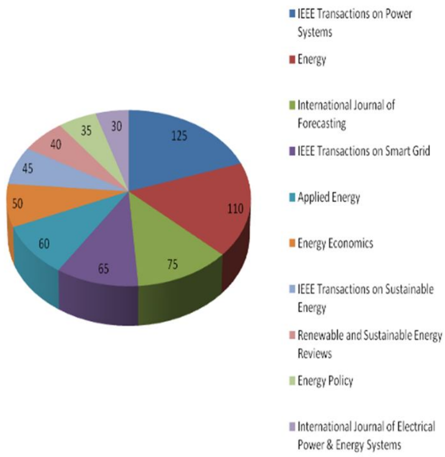

Figure 4: Number of papers published in SCI

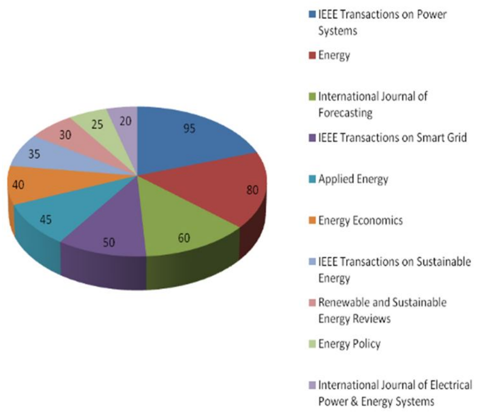

Figure 5: Number of papers published in IEE

II. 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 can make better decisions, which improves results and performance. Big data reduces uncertainty and risk, helping decision-makers make faster, better judgements. Large data analysis and use are becoming increasingly important for modern decision-makers. Businesses collect big data from social media, mobile devices, media platforms, and loyalty programmes. Big data is becoming more crucial in decision-making. Big data analysis must include recent and historical data to be effective. Organisations must use internal and external data to make good decisions. 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 In dynamic factor models (DFMs), variables' relationships change over time. London Stock Exchange transaction values can be examined using DFM models. Big data is used to predict stock prices in finance. Before incorporating data in a model, outliers must be removed to ensure accuracy [30].

For criminal detection, data mining is often used. Two-step clustering and K-means are popular methods for detecting data trends and anomalies, which can help identify fraud. Merging data components can help these algorithms spot fraud-related trends. Decision trees, random forests, and neural networks can also be employed [31]. Singular value decomposition (SVD) is used in machine learning for picture compression and recommendation systems. Researchers can use a lot of YouTube video data to anticipate video patterns using SVD. Hierarchical clustering groups similar patterns to improve prediction [32]. In big data weather forecasting, stochastic advection diffusion differential equations can better model air flow and predict weather patterns. These models can improve by using big data to obtain more detailed weather and atmospheric data. This strategy outperformed simple numerical forecasting models in your North Swiss study [33].

Big data forecasting in several sectors, including power consumption, uses SVMs and NNs. Researchers in China have used these algorithms to predict future electricity use trends from massive databases. By using SVMs and NN, researchers hope to improve forecast accuracy [34]. The wavelet transform can be used with statistical models and machine learning to predict UK petrol and power prices. Wavelet transform data helps these models predict and improve projections. WT can analyse and extract useful information from time series data to improve regression, GARCH, and MLP forecasting. Methodology and strategy depend on the topic and data.[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]. Additionally referred to as time series databases, TSDBs are utilised for the purpose of maintaining and analysing enormous amounts of data that is arranged in a time series fashion. This is accomplished through the utilisation of TSDBs. The benchmarks that have been established might not be an accurate representation of the performance of TSDBs in the real world [37]. This is a possibility. Within the confines of this study, a total of twelve distinct dimensionality reduction algorithms are offered for consideration. These algorithms can be classified into a variety of different categories, including supervision, linearity, time and memory complexity, hyper-parameters, and disadvantages [38]. There is a wide range of distinguishing qualities that can be investigated in great detail in order to build these categories, which is the goal that needs to be accomplished. This inquiry is being conducted with the intention of delivering a full examination of data stream frameworks, statistical analysis, and algorithmic approaches. What a —! The purpose of this article is to study a wide variety of approaches that are currently being utilised for real-time operations and forecasting. These strategies encompass a variety of different techniques. For the purpose of this research study, deep learning techniques such as LSTMs, CNNs, AEs, and GNNs, in addition to machine learning techniques such as LightGBM, were utilised. AEs and GNNs were two more approaches that were utilised. What a —! [39] During the course of the conversation, auto-regression and classical statistical approaches were brought up as additional points of interest that were brought up.

III. 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] Performance Study of Time Series Databases 2022. doi: 10.48550/arxiv.2208.13982.
[38] A Survey on Dimensionality Reduction Techniques for Time-series Data. IEEE Access 2023, doi: 10.1109/access.2023.3269693.
[39] Ana, Almeida., Susana, Brás., Susana, Sargento., Filipe, Cabral, Pinto. (2023). Time series big data: a survey on data stream frameworks, analysis and algorithms. Journal of Big Data, doi: 10.1186/s40537-023-00760-1.