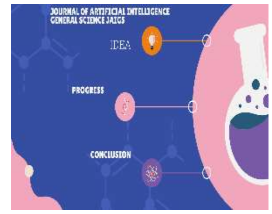
Vol.3, Issue 1, February 2024
Journal of Artificial Intelligence General Science JAIGS

https://ojs.boulibrary.com/index.php/JAIGS



AI-Driven Analytics: Transforming Data Platforms for Real-Time Decision Making

Chandrashekar Althati[1], Jesu Narkarunai Arasu Malaiyappan[2], Lavanya Shanmugam[3]

[1]Medalogix, USA

[2]Meta Platforms Inc, USA

[3]Tata Consultancy Services, USA

ARTICLEINFO
Article History:
Received:01.03.2024
Accepted:
15.03.2024
Online: 30.03.2024
Keyword: AI-driven analytics, data platforms, real-time decision-making, machine learning algorithms, insights generation, operational efficiency.

ABSTRACT

AI-driven analytics represents a transformative paradigm shift in data platforms, enabling real-time decision-making capabilities across various domains. This paper explores the integration of artificial intelligence (AI) technologies into data platforms, elucidating their role in accelerating insights generation and facilitating agile decision-making processes. By harnessing advanced machine learning algorithms, AI-driven analytics empower organizations to extract actionable insights from vast datasets, thereby driving innovation and enhancing operational efficiency. Through a comprehensive analysis, this study delineates the key components and methodologies underlying AI-driven analytics and examines their impact on transforming traditional data platforms into agile, adaptive ecosystems conducive to real-time decision-making.

Introduction

In today's fast-paced business landscape, the ability to make data-driven decisions in real-time is imperative for organizations striving to stay competitive. The advent of AI-driven analytics represents a groundbreaking shift in the realm of data platforms, promising transformative capabilities in extracting actionable insights and enabling agile decision-making processes. By leveraging advanced artificial intelligence (AI) technologies, organizations can harness the power of machine learning algorithms to unlock valuable insights from vast and complex datasets.

The integration of AI-driven analytics into data platforms holds the potential to revolutionize traditional approaches to data analysis, empowering businesses to respond swiftly to changing market dynamics and emerging opportunities. This paper delves into the intricacies of AI-driven analytics, exploring its role in transforming data platforms into dynamic and adaptive ecosystems capable of delivering real-time insights.

Through a comprehensive examination of the underlying methodologies and components of AI-driven analytics, this study aims to shed light on its transformative impact on organizational decision-making processes. By harnessing the potential of AI-driven analytics, organizations can not only enhance operational efficiency but also drive innovation and strategic growth initiatives.

In the subsequent sections, we delve deeper into the mechanisms and applications of AI-driven analytics, elucidating its potential to revolutionize data platforms and facilitate real-time decision-making across various domains.

Objectives:

1. Explore the Role of AI-Driven Analytics:

Investigate the role of AI-driven analytics in transforming traditional data platforms, elucidating how advanced machine learning algorithms enhance the speed and accuracy of insights generation for real-time decision-making.

2. Examine Methodologies and Components:

Delve into the methodologies and components underlying AI-driven analytics, analyzing how machine learning techniques are integrated into data platforms to enable efficient data processing, analysis, and interpretation.

3. Evaluate Impact on Decision-Making Processes:

Assess the impact of AI-driven analytics on organizational decision-making processes, examining how real-time insights derived from AI-powered data platforms empower businesses to respond swiftly to market dynamics, optimize operations, and drive strategic initiatives for growth and innovation.

Research Methodology:

This study employs a multi-faceted research methodology aimed at comprehensively exploring the role and impact of AI-driven analytics in transforming data platforms for real-time decision-making. The research methodology comprises the following key components:

1. Case Studies and Use Cases:

Analyzing real-world case studies and use cases of organizations that have successfully implemented AI-driven analytics solutions. By examining these practical examples, insights will be gained into the challenges, benefits, and best practices associated with integrating AI technologies into data platforms for real-time decision-making.

2. Expert Interviews:

Conducting interviews with industry experts, data scientists, and professionals specializing in AI-driven analytics and data platform transformations. These interviews will provide valuable insights into emerging trends, challenges, and opportunities in the field, as well as practical strategies for implementing AI-driven analytics solutions effectively.

3. Data Analysis and Interpretation:

Utilizing data analysis techniques to examine the performance metrics, efficiency gains, and qualitative outcomes associated with AI-driven analytics implementations. By analyzing data from various sources, including case studies, interviews, and experimental data, this research aims to quantify the impact of AI-driven analytics on decision-making processes.

4. Qualitative and Quantitative Evaluation:

Employing a mixed-methods approach to evaluate the qualitative and quantitative aspects of AI-driven analytics adoption. This includes assessing the qualitative feedback from interviews and case studies, as well as quantifying the measurable outcomes such as cost savings, productivity improvements, and decision-making accuracy.

By integrating these research methods, this study endeavors to provide a comprehensive understanding of the role of AI-driven analytics in transforming data platforms for real-time decision-making, and to offer practical insights for organizations seeking to leverage AI technologies for strategic advantage.

Literature Review

AI-driven analytics, particularly through Augmented Analytics (AA), is revolutionizing data platforms for real-time decision-making by combining Business Intelligence (BI) with advanced Artificial Intelligence (AI) features [1]. This transformation enables automation of the business analytics cycle using Machine Learning (ML) and natural language comprehension, enhancing analysis, reducing time, and supporting data preparation, visualization, modeling, and insight generation [2]. Additionally, cognitive computing and AI empower enterprises to swiftly translate vast data into valuable insights, improving decision-making efficiency and accuracy [3]. Furthermore, in the public sector, AI-driven transformation and smart data management are enhancing decision-making processes by providing value to citizens and assisting decision-makers, emphasizing the need for operational transformation and efficient services that benefit society [4]. Overall, AI-driven analytics is reshaping data platforms, enabling real-time decision-making across various sectors with increased efficiency and effectiveness.

Background

The rapid advancement of big data technology has significantly impacted various industries, profoundly altering the landscape of enterprise management decision-making. This evolution presents both opportunities and challenges, as decision-making data becomes increasingly complex and decision-makers face heightened difficulty in identifying pertinent information [1]. Consequently, there is a pressing need for innovation in traditional decision-making methodologies to mitigate risks and ensure the smooth operation of enterprises.

In response to these challenges, organizational leadership must cultivate data awareness and harness the power of big data in management decision-making processes. This necessitates the adoption of a comprehensive and diversified approach to human resources management, enabling organizations to effectively navigate environmental shifts and mitigate risks associated with decision-making. Moreover, it calls for innovation in data mining and analysis technologies to gather a wealth of information relevant to management decision-making.

By embracing new methodologies and technologies, organizations can proactively address decision-making risks and leverage data-driven insights to drive strategic initiatives and foster sustainable growth.

The concept of big data encompasses both a broad and narrow interpretation. In its broad sense, big data refers to the vast collection of data that not only enables rapid and accurate analysis but also holds significant implications for decision-making. In a narrower sense, big data involves the aggregation of massive volumes of information. Big data exhibits several key characteristics: firstly, it possesses a large capacity, often measured in terabytes (TB), petabytes (PB), and exabytes (EB), facilitating extensive data integration. Secondly, big data technology boasts fast processing efficiency, offering significant advantages over previous processing technologies by swiftly identifying valuable resources within vast datasets and showcasing enterprise capabilities. Thirdly, big data encompasses diverse data types, including text, images, and video [2]. This diversity strengthens data capacity expansion and facilitates the collection of relevant information for a broader range of users, thereby enhancing data processing and analysis efficiency and positively influencing scientific management decisions.

Decision-making lies at the heart of enterprise management, encompassing strategic and operational decisions that significantly impact the state and trajectory of organizational operations and development. The decision-making environment undergoes dynamic changes influenced by the era of big data, characterized by a dramatic increase in data volume and storage capacity, ranging from TB to PB and EB. This evolution reshapes decision-making paradigms under the influence of data-driven approaches [3]. However, despite the potential of big data, many enterprises face challenges related to low information processing efficiency, impacting the effectiveness of big data utilization.

Furthermore, decision-making data has evolved significantly over time, with advancements in type, quantity, and structure. In business operations, data collected through information platforms require organization and optimization to ensure clarity and purposefulness. This entails selective and strategic data screening, continuous optimization of data and information, and comprehensive upgrades to existing information processing systems [4]. Leveraging strong technical assistance can facilitate real-time data processing and enhance the connection between big data and information, thereby uncovering information closely linked to enterprise objectives and promoting sound development goals.

Impact on Decision-Making Participants:

The role of decision-making participants has undergone significant transformation with the advent of big data. Traditional decision-making approaches have proven inadequate in adapting to evolving enterprise needs and the changing times. Consequently, decision-making now demands more refined and evidence-based analyses. Senior leaders are tasked with abandoning outdated decision-making methods reliant solely on past experiences. Instead, they must adopt a comprehensive approach by gathering data and information and aligning task deployment with the enterprise's actual situation. This shift ensures optimal utilization of human resources through carefully planned and orchestrated decision-making processes.

Impact on Decision-Making Systems:

The decision-making system comprises two crucial elements: the foundation for decision-making and the decision-making process itself. Previously, enterprise decision-making relied heavily on internal information system data and report data, often resulting in a narrow and subjective view that merely reflected operational and financial aspects [5]. However, with the proliferation of network technologies, enterprises can now swiftly collect information from various sources, enabling a more comprehensive understanding of market dynamics, consumer preferences, and other pertinent factors. Incorporating this diverse information into decision-making processes enhances objectivity and enables enterprises to better navigate market risks, thereby enhancing their core competitiveness.

Moreover, enterprises can develop decision management systems empowered by big data, establishing integrated systems tailored to different departments to showcase practicality, scalability, and comprehensive functionality [6]. Leveraging these integrated systems, enterprises can tap into diverse data sources, understand user behaviors and feedback, and optimize product design accordingly. This enables products to better align with consumer expectations, drive sales, and generate increased economic benefits. However, the complexity of decision-making, fueled by big data, necessitates a focused data analysis approach that maximizes the benefits of big data technology. The effectiveness of data analysis outcomes is contingent upon the analytical abilities of employees. Staff with strong analytical skills can effectively extract valuable insights from data, while those lacking such skills may struggle, resulting in a waste of data resources [7]. To address this, enterprises must establish platforms for data analysis during their development journey, striving for higher efficiency at lower costs [8].

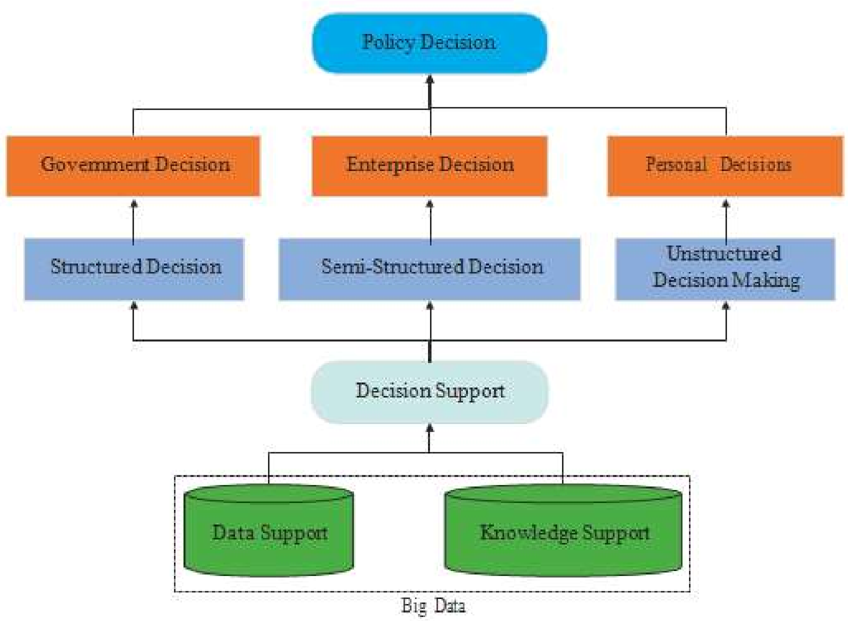


Figure 1: Enterprise decision based on big data.

Algorithm Design: Anomaly Detection Method

Detecting anomalies in data involves considering various factors such as timing, context, and numerical information of the data itself. Unlike conventional classification problems, anomaly detection poses unique challenges, including severely unbalanced positive and negative samples and the presence of unlabeled anomalous data.

Recent research in time series anomaly detection has increasingly utilized deep learning-based modeling methods. This study focuses on analyzing anomaly detection algorithms based on recurrent neural networks (RNNs) and variational autoencoders (VAEs).

Given that anomaly detection problems typically involve significantly more normal data than anomalous data, traditional classification approaches are inadequate. Thus, researchers have turned to unsupervised learning algorithms for anomaly detection. Time series can be decomposed into several components using methods like Seasonal and Trend decomposition using Loess (STL), represented by the formula:

$$x_t = \tau_t + c_t + s_t + i_t$$

Here, x_t denotes the numerical vector of the time series at moment t, τ_t is the trend part, c_t is the periodic part, s_t is the seasonal part, and i_t is the irregular part. Anomaly detection models process the retained parts of the time series after decomposition. Decomposing time series simplifies the problem, making it easier for models to detect anomalies by removing noisy and irrelevant parts.

Unsupervised learning allows models to learn the complex distribution of data. Recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are suitable for sequence-based scenarios. LSTM networks excel at capturing time series-dependent information. Therefore, this study incorporates RNNs for time series anomaly detection. Previous research has demonstrated the effectiveness of LSTM in this context. To enhance the model's ability to capture time series-dependent information, additional neural network layers are considered for processing time series in Variational Autoencoder (VAE). However, the presence of too many LSTM layers can increase the model's training difficulty. Hence, RNN layers are introduced to balance model complexity. The anomaly detection model follows a VAE-like structure and undergoes unsupervised training using normal time series data to determine anomalies based on reconstruction errors. The general process involves time series decomposition followed by anomaly detection using the cyclic VAE model.

FIGURE 2: Decision-making process of uncertainty increase and human-machine cooperation.

Figure 3 illustrates the overarching structure of D-R-VAE, which consists of three main components. The anomaly detection model combines recurrent neural networks and variational autoencoders. Each module is detailed below.

For non-periodic time series, we employ the HP filtering method. This approach decomposes time series data into two components: the trend part and the residual part. We retain the trend part of the decomposed time series for subsequent anomaly detection. Compared to the original time series, the trend part becomes less sensitive to short-term fluctuations, effectively filtering out noise and other interfering factors to some extent. The sensitivity of the trend part to short-term fluctuations can be adjusted using parameters in the decomposition function, tailored to the dataset's characteristics.

D-R-VAE decomposes time series into multiple components using standard time series decomposition methods, discarding unnecessary components for anomaly detection tasks. The cyclic VAE model conducts anomaly detection on the retained time series, leveraging a combination of VAE, LSTM, and RNN. Figure 4 depicts the algorithm's flowchart framework. The algorithm's execution unfolds in two main stages: first, time series are decomposed using classical statistical and mathematical time series decomposition methods. Time series with evident periodicity undergo STL decomposition, while those lacking periodicity undergo HP filtering. The residual part after STL decomposition and the trend part after HP filtering are retained. These retained segments form the basis of the new time series. In the subsequent phase, the D-R-VAE model conducts anomaly detection on the processed time series. Initially, the D-R-VAE model is trained using normal time series data, after which the anomaly score output from each subsequence in the training set is used to calculate anomaly thresholds. During testing, time series in the test set are reconstructed using the D-R-VAE model, and abnormal scores are computed. These scores are then compared with the anomaly thresholds to determine abnormality status.

Experimentation

Abnormal Data Detection Experiment

Experimental Parameter Configuration: For comparison, three algorithms are chosen alongside the algorithms proposed in this study. The benchmark algorithms are outlined briefly as follows:

1. LSTM-based anomaly detection algorithm: Utilizes LSTM for time series prediction and anomaly determination based on prediction error.

2. VAE-based anomaly detection algorithm: Employs VAE to reconstruct input data.

Datasets:

1. SMD (Machine Dataset): Collected by researchers from an internet company, comprising time series data from three sets of physical machines. This dataset consists of 28 multivariate time series datasets from various machines, each needing independent analysis.

2. Yahoo Benchmark Dataset (Yahoo): Comprising four benchmark datasets, with the first selected to evaluate the model algorithm. The remaining three datasets are synthetic datasets, designed for easy identification of anomalies by the model.

Since anomaly detection problems typically involve a dataset with more normal data than anomalous data, resulting in unbalanced data, evaluation criteria for algorithm performance under such conditions are chosen. Abnormal data is classified as positive, and normal data as negative. The data judged by the model are categorized into four groups based on their nature and the model's judgment: True Positive (TP) for abnormal data correctly identified as such, False Negative (FN) for abnormal data misclassified as normal, False Positive (FP) for normal data incorrectly flagged as abnormal, and True Negative (TN) for normal data correctly classified as such.

Experimental Results and Analysis

We begin by examining the impact of parameters and window length on the anomaly detection performance of the D-R-VAE model. The parameters are increased incrementally from 0.2 to 1.0 (in increments of 0.2), and the model's performance is evaluated using the F1 score metric on the SMD dataset. Figure 8 illustrates the experimental results, where the red curve depicts the model's performance. It is observed that the model achieves optimal performance when the parameters fall within the range of 0.2 to 0.6. Beyond this range, as the parameters increase, the model's anomaly detection performance tends to decrease.

Furthermore, for model training and anomaly detection using the D-R-VAE model, a fixed-length sub-series is required as input, known as the time series window length. This window length significantly impacts the model's anomaly detection performance. Figure 8 also displays the experimental results of the D-R-VAE model with varying window lengths for anomaly detection on the SMD dataset. It is evident that the model's anomaly detection performance gradually improves as the window length increases from 80 to 140 and then stabilizes.

Table 1 demonstrates the effect of time series decomposition on the anomaly detection performance of the D-R-VAE model. Prior to anomaly detection by the D-R-VAE algorithm, the time series is decomposed using specific time series decomposition methods. The D-R-VAE model then performs anomaly detection on the retained portion of the decomposed time series, while the R-VAE algorithm conducts anomaly detection on the original time series. The experimental results indicate a significant enhancement in the performance of the D-R-VAE algorithm compared to the R-VAE algorithm. This suggests that the anomaly detection performance of the D-R-VAE model can be effectively enhanced by appropriately processing the time series using time series decomposition methods.

Conclusions

In the era of big data, the landscape of enterprise management and decision-making has undergone significant transformation, particularly concerning the decision-making environment, data, participants, and systems. This evolution imposes greater demands on leadership and managers, necessitating the establishment of robust data awareness. Leaders must leverage big data to comprehensively understand market dynamics and formulate scientifically informed decision-making strategies tailored to their specific contexts. To address these challenges, this paper proposes an artificial intelligence decision-making platform to assist decision-makers in tackling major decision-making dilemmas.

Firstly, to address the anomaly detection of time series data, the D-R-VAE algorithm is introduced. This algorithm decomposes time series data and retains components relevant to anomaly detection. Leveraging the model structure of Variational Autoencoder (VAE), the algorithm learns the normal patterns of time series data. Additionally, incorporating LSTM and RNN layers enhances the model's ability to process sequence-like data and extract time series-dependent information during data processing.

Subsequently, an artificial intelligence decision platform is constructed based on width learning. This platform includes a decision algorithm grounded in coded width learning networks and a parallelized training algorithm based on width learning models. The former integrates denoising self-coding and width learning to achieve high timeliness and accuracy in decision-making, while the latter combines width learning features with integrated learning methods for parallelized training, thereby enhancing efficiency and mitigating memory constraints.

Experimental results demonstrate that the D-R-VAE model effectively detects anomalous data, furnishing valuable data references for subsequent decision-making processes. Furthermore, the width learning-based decision platform exhibits short practice times and high accuracy, meeting real-time decision requirements. Looking ahead, future endeavors will explore the application of graph convolutional neural networks in building the artificial intelligence decision platform and leveraging big data analysis technology.

References List:

[1]. Soni, J., Gangwani, P., Sirigineedi, S., Joshi, S., Prabakar, N., Upadhyay, H., & Kulkarni, S. A. (2023). Deep Learning Approach for Detection of Fraudulent Credit Card Transactions. In Artificial Intelligence in Cyber Security: Theories and Applications (pp. 125-138). Cham: Springer International Publishing.

[2]. Talati, D. (2023). AI in healthcare domain. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 256-262.

[3]. Talati, D. (2023). Telemedicine and AI in Remote Patient Monitoring. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 254-255.

[4]. Talati, D. (2024). Virtual Health Assistance–AI-Based. Authorea Preprints.

[5]. Talati, D. (2023). Artificial Intelligence (Ai) In Mental Health Diagnosis and Treatment. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 251-253.

[6]. Talati, D. (2024). Ethics of AI (Artificial Intelligence). Authorea Preprints.

[7]. Talati, D. V. AI Integration with Electronic Health Records (EHR): A Synergistic Approach to Healthcare Informatics December, 2023.

[8]. Singla, A., & Malhotra, T. (2024). Challenges And Opportunities in Scaling AI/ML Pipelines. Journal of Science & Technology, 5(1), 1-21.

[9]. Singla, A., & Chavalmane, S. (2023). Automating Model Deployment: From Training to Production. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 340-347.

[10]. Gehrmann, S., & Rončević, I. (2015). Monolingualisation of research and science as a hegemonial project: European perspectives and Anglophone realities. Filologija, (65), 13-44.

[11]. Roncevic, I. (2021). Eye-tracking in second language reading. Eye, 15(5).

[12]. Šola, H. M., Gajdoš Kljusurić, J., & Rončević, I. (2022). The impact of bio-label on the decision-making behavior. Frontiers in sustainable food systems, 6, 1002521.

[13]. Sirigineedi, S. S., Soni, J., & Upadhyay, H. (2020, March). Learning-based models to detect runtime phishing activities using URLs. In Proceedings of the 2020 4th international conference on compute and data analysis (pp. 102-106).

[14]. Verma, V., Bian, L., Ozecik, D., Sirigineedi, S. S., & Leon, A. (2021). Internet-enabled remotely controlled architecture to release water from storage units. In World Environmental and Water Resources Congress 2021 (pp. 586-592).

[15]. Soni, J., Sirigineedi, S., Vutukuru, K. S., Sirigineedi, S. C., Prabakar, N., & Upadhyay, H. (2023). Learning-Based Model for Phishing Attack Detection. In Artificial Intelligence in Cyber Security: Theories and Applications (pp. 113-124). Cham: Springer International Publishing.

[16]. Verma, V., Vutukuru, K. S., Divvela, S. S., & Sirigineedi, S. S. (2022). Internet of things and machine learning application for a remotely operated wetland siphon system during hurricanes. In Water Resources Management and Sustainability (pp. 443-462). Singapore: Springer Nature Singapore.

[17]. Talati, D. (2024). AI (Artificial Intelligence) in Daily Life. Authorea Preprints.