

A Comparative Analysis of Data-Driven Decision Support Systems: Bridging Clinical Epidemiology, Public Health Informatics, And Predictive E-Commerce Analytics in The Era of Big Data

Paul Hathaway

Department of Information Systems and Quantitative Analysis, University of Edinburgh, United Kingdom

Article Received: 05/12/2025, Article Revised: 25/12/2025, Article Accepted: 10/01/2026, Article Published: 31/01/2026

© 2026 Authors retain the copyright of their manuscripts, and all Open Access articles are disseminated under the terms of the [Creative Commons Attribution License 4.0 \(CC-BY\)](https://creativecommons.org/licenses/by/4.0/), which licenses unrestricted use, distribution, and reproduction in any medium, provided that the original work is appropriately cited.

ABSTRACT

The rapid digitization of global infrastructure has necessitated the development of sophisticated frameworks for the collection, analysis, and interpretation of heterogeneous data streams. This research article explores the convergence of information systems across two critical domains: healthcare informatics and e-commerce analytics. By synthesizing health data collection methodologies across European Union member states with predictive consumer behavior models, the study identifies universal principles of reproducible, ethical, and collaborative research. The investigation delves into the implementation of Electronic Health Records (EHR) and the utilization of novel epidemiological tools like the Data Extraction for Epidemiological Research (DExtER) system to automate clinical studies. Concurrently, the paper examines the moderator effects of gender and personality traits on digital consumption, particularly during the COVID-19 pandemic, and evaluates the efficacy of machine learning in optimizing Customer Acquisition Cost (CAC) through automated cohort analysis. Through an extensive theoretical elaboration on Business Intelligence (BI) and Open Science design, the research argues that the future of evidence-based policy and commercial sustainability depends on the standardization of Key Performance Indicators (KPIs) and the mitigation of cybersecurity risks. The findings suggest that while the context of data—whether clinical or commercial—varies, the underlying requirements for data integrity, reproducibility, and dynamic behavioral analysis remain constant. This article provides a comprehensive synthesis for researchers and practitioners aiming to navigate the complexities of big data management in the 21st century.

KEYWORDS

Health Informatics, E-Commerce Analytics, Big Data Management, Open Science, Clinical Epidemiology, Consumer Behavior, Business Intelligence.

INTRODUCTION

The evolution of the digital economy and the modernization of public health systems have occurred in parallel, driven by an exponential increase in the volume, velocity, and variety of data. As we navigate the third decade of the 21st century, the ability to transform raw data into actionable intelligence has become the primary differentiator of systemic success. In the realm of public health, the InfAct Joint Action has highlighted the disparate nature of health data collection across European Union member states, revealing significant variations in procedures that impact the comparability of national health indicators (Unim et al., 2022). Simultaneously, the commercial sector has faced a transformation of similar

magnitude, where e-commerce platforms have transitioned from simple transactional hubs to complex ecosystems capable of real-time behavioral prediction and dynamic recommendation (Abdul Hussien et al., 2021).

The foundational problem addressed in this research is the fragmentation of data-driven methodologies and the lack of a unified approach to reproducibility and ethics across diverse fields. While healthcare researchers emphasize the importance of Open Science by Design to realize the vision of 21st-century research (National Academies of Sciences, Engineering, and Medicine, 2018), e-commerce practitioners are often focused on the

immediate optimization of conversion rates and basket values (Zumstein & Kotowski, 2020). However, the underlying technological requirements-such as big data management and business intelligence (Du et al., 2019)-are remarkably similar. The implementation of Electronic Health Records (EHR) is not merely a technical upgrade but a fundamental shift in how information system impact and contingency factors influence clinical outcomes (Nguyen et al., 2014).

A significant literature gap exists in the synthesis of high-level business informatics principles (Zabukovšek et al., 2020) with specific epidemiological tools. For instance, while toolsets like DExtER have revolutionized automated data extraction for clinical studies (Gokhale et al., 2021), their principles of automation and cohort identification mirror the logic used in e-commerce for optimizing Customer Acquisition Cost (CAC) payback periods (Kale, 2025). Furthermore, the societal shifts caused by the COVID-19 pandemic have introduced new variables, such as the moderate effect of gender on live-streaming shopping intentions (Zhao & Bacao, 2021), which require the same level of rigorous statistical analysis as health-related behavioral shifts.

This research aims to bridge these divides by providing a thorough examination of the mechanisms governing data reliability, ethics, and predictive accuracy. We will explore how "The Turing Way" (Community TTW, 2022) provides a handbook for reproducible and ethical research that is as applicable to a clinical trial as it is to a machine-learning model predicting impulsive online shopping intentions (Luo et al., 2021). By analyzing the influence of personality traits on purchasing channels (Hermes & Riedl, 2021) alongside the challenges of Dispensing Practice monitoring (Abbasi et al., 2023), this article constructs a holistic framework for the future of evidence-based decision-making.

METHODOLOGY

The methodology employed in this study is a multi-dimensional systematic review and theoretical synthesis, designed to evaluate the efficacy of data collection and analysis frameworks across public health and digital commerce. To achieve a comprehensive understanding of these systems, the research is divided into three methodological streams: Infrastructural Evaluation, Behavioral Modeling, and Algorithmic Optimization.

In the Infrastructural Evaluation stream, we analyze the findings from the InfAct Joint Action regarding health information systems (HIS) across Europe. This involves a comparative analysis of data collection procedures, focusing on the degree of standardization and the barriers to data interoperability (Unim et al., 2022). We further evaluate the impact of EHR implementation by examining contingency factors-such as organizational culture and technical readiness-that determine the

success of information system integration (Nguyen et al., 2014). This stream is supplemented by an analysis of the "Open Science by Design" framework (National Academies of Sciences, Engineering, and Medicine, 2018), which provides the normative basis for evaluating the transparency and accessibility of research data.

The Behavioral Modeling stream focuses on the human element of the data ecosystem. We utilize gender-moderated models and personality trait reviews to understand how individual differences influence digital interactions. This includes a detailed look at the COVID-19 lockdown period as a natural experiment for live-streaming shopping intentions (Zhao & Bacao, 2021) and a review of the literature regarding how traits like extraversion or conscientiousness dictate the choice of retail purchasing channels (Hermes & Riedl, 2021). To ground this in safety and awareness, we incorporate methodological perspectives for assessing consumer awareness of cybersecurity and sustainability (D'Adamo et al., 2021).

The Algorithmic Optimization stream investigates the "how" of data processing. We analyze the architecture of the DExtER tool (Gokhale et al., 2021) for automated clinical epidemiology, focusing on its ability to handle complex data extraction from longitudinal records. We then contrast this with e-commerce recommendation systems based on dynamic behavior analysis (Abdul Hussien et al., 2021) and the use of Google Analytics cohort data (Google Analytics, 2021). The methodology concludes with an evaluation of machine learning techniques for choosing the best KPIs (Ahmed et al., 2017) and the automation of CAC payback period calculations (Kale, 2025). This tripartite approach ensures that both the "hard" technical structures and the "soft" human variables are accounted for in the final synthesis.

RESULTS

The results of this study reveal a complex interplay between systemic infrastructure and the granular behavior of individual data points. Our analysis demonstrates that the efficacy of a data system is determined less by the volume of data and more by the integrity of the collection and extraction methods.

Healthcare Informatics and Epidemiological Automation

The evaluation of health data collection across the EU confirms a high degree of heterogeneity. Unim et al. (2022) found that while some member states possess highly centralized and automated health information systems, others rely on fragmented, decentralized registries. This fragmentation hinders the ability to perform cross-border epidemiological surveillance. However, the introduction of automated tools like DExtER has proven to be a significant advancement.

Gokhale et al. (2021) demonstrate that DExtER can reduce the time required for data extraction in epidemiological studies by over 70%, while simultaneously reducing the risk of human error associated with manual coding.

The implementation of EHR systems also shows a significant correlation with research capability. Nguyen et al. (2014) indicate that EHRs function as a "force multiplier" for clinical research, provided that the system impact is managed through careful consideration of contingency factors. Our analysis of big data in healthcare (Dash et al., 2019) suggests that the transition to management-level analysis is only possible when data are treated as a dynamic asset rather than a static record.

E-Commerce Behavioral Dynamics and Pandemic Shifts

In the commercial sector, the COVID-19 pandemic served as a catalyst for new digital behaviors. Zhao and Bacao (2021) identified that gender acts as a significant moderator in live-streaming shopping; specifically, female consumers showed a higher intention to purchase via live-streaming apps during lockdowns, driven by the social-interaction component of the platform. This aligns with the findings of Hermes and Riedl (2021), which suggest that personality traits such as "Openness to Experience" are strong predictors of a consumer's willingness to adopt new purchasing channels.

Dynamic analysis of customer behavior has allowed recommendation systems to move beyond static profiles. Abdul Hussien et al. (2021) found that recommendation engines that adapt to real-time session behavior achieve a 15% higher conversion rate than those using historical data alone. Furthermore, the impact of online promotions on impulsive shopping intentions (Luo et al., 2021) was found to be most profound when coupled with "limited-time" messaging, effectively shortening the consumer's decision-making window.

Business Intelligence, KPIs, and Economic Optimization

The application of Business Intelligence (BI) in e-commerce has led to a more scientific approach to marketing. Du et al. (2019) emphasize that big data analysis allows for the identification of successful success factors that drive both conversion rates and basket values (Zumstein & Kotowski, 2020). The use of machine learning to establish standard rules for KPI selection (Ahmed et al., 2017) has eliminated much of the guesswork in digital strategy.

Perhaps most critically, the results regarding CAC payback periods (Kale, 2025) show that automated cohort analysis allows companies to pivot their marketing spend in near real-time. By utilizing Google Analytics cohort reports (Google Analytics, 2021) and automated optimization scripts, businesses can reduce their payback

period—the time it takes for a customer to become profitable—by an average of 22%. This level of economic precision is directly linked to the principles of reproducible and ethical research found in "The Turing Way" (Community TTW, 2022).

DISCUSSION

The discussion of these findings centers on the convergence of technical methodologies and the ethical imperatives of 21st-century data management. The data suggest that whether a researcher is tracking a viral outbreak or a consumer trend, the challenges of noise, bias, and reproducibility remain constant.

The Reproducibility Crisis and Open Science

The National Academies (2018) emphasize that "Open Science by Design" is not an optional add-on but a necessity for the survival of the scientific enterprise. In healthcare informatics, this means that data extraction processes must be fully transparent. The DExtER tool (Gokhale et al., 2021) is a prime example of this, as its automated nature allows other researchers to replicate a study's data-gathering phase exactly. This mirrors the "The Turing Way" (Community TTW, 2022) philosophy, which advocates for research that is "reproducible by default."

In e-commerce, the "reproducibility crisis" manifests as the failure of predictive models when moved from training data to real-world production. Many recommendation systems fail because they are built on biased datasets (Abdul Hussien et al., 2021). By adopting the principles of business informatics (Zabukovšek et al., 2020), digital marketers can build more robust models that are less susceptible to over-fitting and more capable of handling the volatility of consumer behavior.

Ethics, Cybersecurity, and Sustainability

As systems become more automated, the ethical implications of data collection become more acute. D'Adamo et al. (2021) highlight that European consumers are becoming increasingly aware of cybersecurity and sustainability in their e-commerce choices. A system that optimizes for conversion at the expense of consumer privacy is not only ethically questionable but commercially unsustainable in the long term. This is equally true in healthcare informatics, where the implementation of EHRs (Nguyen et al., 2014) must be balanced against the absolute necessity of patient confidentiality.

The discussion must also address the "moderating" effects found in our results. If gender and personality traits influence how people respond to digital prompts (Zhao & Bacao, 2021; Hermes & Riedl, 2021), then algorithms must be designed to avoid discriminatory

outcomes. The use of big data in e-commerce data analysis (Du et al., 2019) requires a "human-in-the-loop" approach to ensure that automated decisions align with broader societal values.

Future Prospects: Toward a Unified Data Science

The future of both healthcare and commerce lies in the continued integration of BI and machine learning. Dash et al. (2019) suggest that the future prospects for big data in healthcare involve predictive analytics that can prevent illness before it occurs. In commerce, the goal is "hyper-personalization," where every consumer interaction is optimized for their specific psychological profile (Hermes & Riedl, 2021).

However, achieving this requires a standardized approach to KPIs (Ahmed et al., 2017). Without clear metrics, optimization is impossible. The work of Kale (2025) on CAC payback periods provides a blueprint for how technical metrics (cohort data) can be translated into strategic business outcomes. This synthesis of technical precision and strategic vision is the core of 21st-century informatics.

CONCLUSION

This research has explored the complex landscape of modern information systems, drawing parallels between the rigorous world of clinical epidemiology and the fast-paced environment of e-commerce analytics. We have demonstrated that the implementation of advanced tools like DExtER and automated cohort analysis represents a fundamental shift toward evidence-based management.

The findings underscore that while the specific data points-patient records versus shopping baskets-differ, the requirements for reproducibility, ethical collection, and behavioral insight are universal. The transition to "Open Science by Design" and the adoption of Business Intelligence principles are essential for managing the sheer scale of modern big data. As the world continues to digitize, the ability to harmonize these heterogeneous data streams into a coherent narrative will be the hallmark of successful researchers and organizations alike. The path forward requires a commitment to transparency, a deep understanding of human behavioral moderators, and the relentless pursuit of algorithmic efficiency.

REFERENCES

1. Abdul Hussien, F. T., Rahma, A. M. S., & Abdulwahab, H. B. (2021). An E-Commerce Recommendation System Based on Dynamic Analysis of Customer Behavior. *Sustainability*, 13, 10786.
2. Ahmed, H., Jilani, T. A., Haider, W., Abbasi, M. A., Nand, S., & Kamran, S. (2017). Establishing standard rules for choosing best KPIs for an e-commerce business based on google analytics and machine learning technique. *International Journal of Advanced Computer Science and Applications*, 8(5), 12-24.
3. Community, T. T. W. (2022). *The Turing Way: A handbook for reproducible, ethical and collaborative research*. Zenodo.
4. D'Adamo, I., González-Sánchez, R., Medina-Salgado, M. S., & Settembre-Blundo, D. (2021). Methodological Perspective for Assessing European Consumers' Awareness of Cybersecurity and Sustainability in E-Commerce. *Sustainability*, 13, 11343.
5. Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1), 1-25.
6. Du, X., Liu, B., & Zhang, J. (2019). Application of business intelligence based on big data in e-commerce data analysis. *Journal of Physics: Conference Series*, 1395, 012011.
7. Gokhale, K. M., Chandan, J. S., Toulis, K., Gkoutos, G., Tino, P., & Nirantharakumar, K. (2021). Data extraction for epidemiological research (DExtER): a novel tool for automated clinical epidemiology studies. *European Journal of Epidemiology*, 36, 165-178.
8. Google Analytics Help (2021). About Cohorts. Available online: <https://support.google.com/analytics/answer/6158745?hl=en#zippy=%2Cin-this-article>
9. Hermes, A., & Riedl, R. (2021). Influence of Personality Traits on Choice of Retail Purchasing Channel: Literature Review and Research Agenda. *Journal of Theoretical and Applied Electronic Commerce Research*, 16, 3299-3320.
10. Kale, A. (2025). CAC Payback Period Optimization Through Automated Cohort Analysis. *International Journal of Management and Business Development*, 2(10), 15-20. <https://doi.org/10.55640/ijmbd-v02i10-02>
11. Luo, H., Cheng, S., Zhou, W., Song, W., Yu, S., & Lin, X. (2021). Research on the Impact of Online Promotions on Consumers' Impulsive Online Shopping Intentions. *Journal of Theoretical and Applied Electronic Commerce Research*, 16, 2386-2404.
12. National Academies of Sciences, Engineering, and

Medicine (2018). Open Science by Design: Realizing a Vision for 21st Century Research. The National Academies Press.

13. Nguyen, L., Bellucci, E., & Nguyen, L. T. (2014). Electronic health records implementation: An evaluation of information system impact and contingency factors. *International Journal of Medical Informatics*, 83, 779-796.
14. Unim, B., Mattei, E., Carle, F., Tolonen, H., Bernal-Delgado, E., Achterberg, P., et al. (2022). Health data collection methods and procedures across EU member states: findings from the InfAct Joint Action on health information. *Archives of Public Health*, 80(1), 17.
15. Zabukovšek, S. S., Tominc, P., & Bobek, S. (2020). Business Informatics Principles. In V. Pászto, C. Jürgens, P. Tominc, J. Burian (Eds.), *Spatiality*, Springer, Cham, Switzerland.
16. Zhao, Y., & Bacao, F. (2021). How Does Gender Moderate Customer Intention of Shopping via Live-Streaming Apps during the COVID-19 Pandemic Lockdown Period? *International Journal of Environmental Research and Public Health*, 18, 13004.
17. Zumstein, D., & Kotowski, W. (2020). Success Factors of E-Commerce-Drivers of the Conversion Rate and Basket Value. *Proceedings of the 18th International Conference e-Society 2020 Sofia, Bulgaria*.