Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The impact of applying Artificial Intelligence (AI) and Big Data Analytics on the strategic decision-making process of managers

Quang Minh Trinh
tqminh@tdu.edu.vn
Tay Do University
Thi Lan Ngo
Tay Do University
Xuan Tung Bui
Tay Do University

Research Article

Keywords: Artificial Intelligence, Big Data Analytics, Strategic Decision-Making, Machine Learning, Sentiment Analysis, Management Information Systems

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10755300/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Modern organizational management demands data-driven strategic decisions to navigate rapidly changing market conditions. This research examines the profound impact of integrating Artificial Intelligence (AI) and Big Data Analytics into managers' strategic decision-making processes. Specifically, the study explores the deployment of Machine Learning (ML) and Deep Learning (DL) technologies in key business areas: Human Resources (HR), e-commerce systems, social media sentiment analysis, financial forecasting, and large-scale customer satisfaction management. Based on common architectures and modern software platforms, the study analyzes operational choices between fully automated AI systems and interactive, customizable interfaces (checkboxes/text boxes) for different user groups (leaders vs. customers). Furthermore, using a case study of Amazon US reviews hosted on the AWS S3 cloud architecture, this study demonstrates how Natural Language Processing (NLP) and machine learning models process unstructured text to extract qualitative and quantitative cues, filter out adversarial noise, and deliver accurate business insights. The research concludes that shifting from experiential intuition to data-driven intelligence minimizes organizational risk and optimizes resource allocation, laying a solid foundation for long-term strategic advantage.

1 Research problem

Modern HR technology products are strongly leveraging the power of Machine Learning (ML) and Deep Learning (DL) on large datasets to optimize decision-making and improve management efficiency. In the recruitment process, ML models help identify talent and accurately predict candidate success rates, providing a solid foundation for screening decisions. In performance management, AI systems transform into data-driven personalization models, automatically suggesting appropriate development and training paths for each individual. Simultaneously, advanced DL algorithms are applied in sentiment analysis tools to measure engagement, supporting HR departments in proactively making decisions to improve the work environment. Finally, the predictive capabilities of AI from big data allow leaders to foresee turnover trends or skill shortages, thereby making strategic long-term HR decisions for the entire organization.

2 Reference Review

The paradigm shift from experiential management to data-driven intelligence has been widely documented in recent literature. Alam (2024) extensively mapped out AI-based data analytics applications, emphasizing how specific corporate functions can be systematically optimized through predictive modeling. The foundational structures governing how advanced analytics influence the strategic execution of top-tier executives were evaluated by Anna Ujwary-Gil and Anna Florek-Paszkowska (2025), who focused on the practical frameworks connecting algorithms with sustainable corporate choices. Complementing the operational scope, Himanshu Rai (2025) provided evidence regarding how algorithmic automation revolutionizes internal growth, specifically through the modern overhaul of HR information systems. Furthermore, the general application of machine learning

architectures inside standard corporate boundaries has been explored from both systemic and software engineering standpoints. Kimbell (2024) illustrated that business management models powered by ML algorithms yield a substantially lower error margin in asset configuration, whereas Lars Buitinck et al. (2013) contributed the foundational understanding of the software API design necessary to systematically deploy these machine learning models in real-world infrastructure. The overall transition regarding how big data serves as a disruptive, success-defining factor for executive decisions has been validated by both Mayer (2019) and Nothof (2019), both establishing that data ingestion infrastructure directly reshapes the speed, transparency, and accuracy of strategic execution.

3 Approach and solution methodology

Approach and solution methodologyTo comprehensively analyze the impact of AI and Big Data on managerial decision-making, this study adopts a multi-faceted methodological framework incorporating both supervised and unsupervised architectural models.For qualitative text streams, such as social media and financial discussions, the methodology establishes two primary computational pipelines:Supervised Sentiment Analysis Pipeline: Initiated by systematic Data Collection & Labeling, followed by Text Pre-processing (tokenization, stop-word removal, and stemming). The data is divided using a standard Data Splitting protocol into distinct training and testing sets to execute supervised models, including Linear/Logistic Regression, Support Vector Machines (SVMs), Decision Trees, and advanced Deep Learning architectures (RNNs, LSTMs, and Transformers). The models undergo strict Optimization via hyperparameter tuning before targeted deployment. Unsupervised Clustering and Representation Learning: To handle data streams devoid of manual tags, an alternative unsupervised pipeline is applied. This utilizes semantic Clustering to group text based on word co-occurrences and Word Embedding techniques (e.g., mapping words into high-dimensional vector spaces). Unsupervised Feature Learning via deep CNNs and RNNs is used to autonomously capture rich textual structures from massive unlabeled datasets. The methodology also evaluates user interaction mechanisms within Explainable AI (XAI) models. We formulate a comparative analytical framework contrasting Fully Automated AI against Interactive AI (utilizing customizable combo boxes and textboxes). This framework tests operational scenarios based on target constraints: strategic "what-if" analyses for corporate leaders requiring risk control, versus frictionless automation versus high-transparency configuration interfaces for consumer-facing environments. Finally, the technical integration is validated through an enterprise data pipeline utilizing the Amazon US Reviews dataset hosted on AWS S3. The methodology maps out data digestion across critical structural variables—including marketplace, customer_id, review_date, product_parent, and product_id—to run real-time sentiment analysis on the text fields (review_headline, review_body) while neutralizing adversarial data through validation metrics like helpful_votes, vine, and verified_purchase

4 Results and evaluation

4.1 AI-Based Data Analytics Applications for Business Management

Core applications of AI and Machine Learning (ML) that optimize business efficiency and enhance customer experience in e-commerce include: *Recommendations & Products*: AI understands individual needs to accurately predict potential customers with high conversion rates, supporting users in making quick purchases and helping businesses build effective marketing strategies. *Virtual Assistants & Chatbots*: Automating customer list creation, shipping arrangements, and quick information responses. This technology creates convenience, drives purchasing decisions, and enhances brand reputation. *SEO & Supply Chain Automation*: AI helps accurately reach target customers through continuous innovation based on market needs, while automating the supply chain based on actual demand to maximize profits and minimize operating costs. *Data Analysis & Future Forecasting (Predictive Power of AI/ML)*: Leveraging large datasets (purchase history, warranties, inventory) to accurately predict market trends and consumer preferences even before customers realize their needs (e.g., automatically analyzing and suggesting computer configurations that fit budgets and usage requirements). *Mutual Benefits*: The combination of AI and e-commerce offers a double benefit: it allows consumers to shop conveniently from home while enabling businesses to personalize products to enhance competitiveness. Major platforms now utilize advanced AI-powered algorithms to store data, monitor, and accurately evaluate performance to achieve strategic goals.

(i). The supervised learning process for social media sentiment analysis involves the following key steps

Supervised learning functions most effectively when a sizable, varied, and representative dataset is available for training. Since the effectiveness of the model can be greatly impacted by the labels' quality, it is crucial to carefully evaluate how to label the text data when performing sentiment analysis.

Data Collection & Labeling

Gather a substantial dataset of social media text samples. These samples can be labeled manually by human annotators, automatically via existing datasets, or through a hybrid approach.

Text Pre-processing

Clean and prepare the text data using techniques such as tokenization, stop-word removal, and stemming.

Data Splitting

Divide the dataset into a training set to build the model and a test set to evaluate its performance.

Model Training

Train a supervised learning model—such as Support Vector Machines (SVMs), decision trees, or deep learning architectures—on the training data.

Evaluation

Assess the model's performance on the test set using standard metrics, including accuracy, precision, and recall.

Optimization

Fine-tune the model to enhance performance by adjusting hyperparameters or applying feature selection and dimensionality reduction techniques.

Deployment

Apply the finalized model to classify the sentiment of new, unseen social media text samples in real-time.

A variety of machine learning algorithms can be applied to supervised sentiment analysis, including:

Linear Regression

Predicts continuous values, making it suitable for regression-based sentiment scoring (e.g., mapping a post's sentiment onto a scale from − 1 to 1).

Logistic Regression

A foundational classification algorithm used to predict binary outcomes (e.g., classifying sentiment as strictly positive or negative).

Support Vector Machines (SVMs)

A powerful classifier capable of handling both binary and multi-class sentiment classification tasks by finding optimal decision boundaries.

Decision Trees

This technique constructs a tree-like model of decisions and their potential outcomes, applicable to both sentiment classification and regression tasks.

Deep Learning Models

Neural network-based architectures (such as RNNs, LSTMs, or Transformers) that excel at capturing complex semantic patterns in text for advanced sentiment analysis.

(ii). **Unsupervised learning** (*No Labeled Data, Clustering Approach, Sentiment Inference*) offers an alternative strategy where the system operates without labeled training data. Instead, the model must independently discover underlying patterns, structures, and relationships within the dataset. A common approach is clustering, which groups similar text data points together based on shared semantic features. The system can then infer the sentiment of new, unseen text samples by mapping them to these automatically generated clusters.

Several unsupervised learning techniques can be applied to social media sentiment analysis:

Clustering

This technique groups social media posts based on semantic or topical similarities (e.g., matching word co-occurrences). Once clustered, the dominant emotion of each group is identified, allowing users to label the clusters as positive, negative, or neutral. These newly labeled data points can subsequently be used to train a supervised model for classifying future text samples.

Word Embedding

This method maps words into a high-dimensional continuous vector space, preserving semantic relationships so that words with similar meanings sit closer together. While techniques like PCA, SVD, or Factor Analysis help reduce dimensionality for visualization (via scatter plots or t-SNE), the dense embeddings themselves serve as powerful features for downstream sentiment classifiers.

Deep Learning (Unsupervised Feature Learning)

Deep learning architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), can be trained on vast unlabeled social media datasets to learn rich textual representations. These self-learned features are then leveraged for subsequent classification tasks.

Because unsupervised learning (the challenge, the core objective, the practical value) requires the system to autonomously discover hidden structures without human-annotated examples, it is inherently more complex than supervised learning. This independence makes the performance of unsupervised algorithms exceptionally dependent on the inherent quality and organization of the incoming data. Without pre-labeled training samples, deciphering the nuanced sentiment of social media posts becomes a demanding task. In summary, though the paths differ, the ultimate goal of any sentiment analysis workflow is to accurately classify text into positive, negative, or neutral categories. For social media platforms, mastering this allows them to effectively monitor the collective tone of user discussions and identify critical red flags in real-time.

Data Augmentation, or Dataset Expansion, is a crucial step in Machine Learning (both supervised and unsupervised) to help the model "learn better," increase generalization capabilities, and avoid overfitting. For text data (*Text/NLP*) in social media sentiment analysis, augmenting samples is not as simple as

rotating or flipping images in image processing; it requires specific techniques to avoid losing the original meaning of the sentence.

AI-powered sentiment analysis tools and platforms for social media include:

Large cloud platforms (with integrated APIs/Machine Learning): IBM Watson (Tone Analyzer, Natural Language Understanding), Google Cloud Natural Language API, Microsoft Azure (Text Analytics API, Azure ML), and SAP Leonardo Machine Learning Foundation.

Specialized social media management and monitoring tools (Social Listening): Hootsuite Insights, Brand24, Talkwalker, Brandwatch, NetBase, and Mention. All integrate sentiment analysis features to evaluate user data.

Finance: the explosion of the internet has increased the volume of text data streams in the financial sector. Although unstructured and more difficult to manage than digital data, this data reflects the collective voice of the community and holds significant value for financial decision-making. Financial social networks connect individuals and organizations, creating a crucial source of big data. In the context of a constantly fluctuating financial market, applying sentiment analysis is essential to understanding and exploiting this unstructured text data. This method helps identify market trends and directions. To address this problem, AI solution providers offer two main approaches: firstly, developing search tools capable of filtering large volumes of social media content (*news, blogs, tweets, etc*.) and classifying relevant industry information.

An explainable artificial intelligence model in e-commerce is composed of a combination of core elements, beginning with leveraging input data including user preferences and behavior, combined with market trend information, product categories, and customer data. This aggregated data is then processed through Machine Learning and Deep Learning algorithms to make purchasing decisions or recommendations. A key characteristic of this model is that these decisions must be clearly explainable so that humans can easily understand them. Finally, the entire process is continuously optimized through a feedback loop, where user feedback is collected to help the AI system learn and improve its accuracy over time. The close integration of input data, AI algorithms, explainability, and the feedback loop is key to unlocking the full power of the technology, thereby enhancing and optimizing the customer experience in the e-commerce environment.

4.2 Fully Automated AI and Interactive AI (with customizable combo boxes/textboxes)

4.2.1. For Leaders (Management & Strategy Perspective):

For leaders, decisions are often strategic, high-risk, and require flexibility. Therefore, solutions with combo boxes and text boxes (AI Decision Support Systems) are always superior. Why do leaders need interaction? Leaders need to consider different scenarios (what-if analysis). For example: "If the budget

is reduced by 10% (enter text box) and the target market is Southeast Asia (select combo box), what will the revenue be?" Risk control: Leaders are ultimately responsible legally and for business performance; they cannot delegate 100% of decision-making power to an AI "black box" without understanding the reasons why. Explainable AI: As a core principle in management, algorithm-based decisions need to be clearly explainable and involve human participation (Human-in-the-loop) to optimize effectiveness. Conclusion for Leaders: Combo box/textbox input is better. AI acts as an intelligent assistant providing predictions based on parameters that leaders actively control.

4.2.2. For Customers (User Experience - UX Perspective):

For customers, the answer will depend on the type of service they are experiencing, but the general trend leans towards a balance or intelligent automation.

Case 1

Prioritizing fully automated AI

Context: Fast, personalized, entertainment, or everyday utility tasks.

Examples: Song recommendation systems on Spotify, movie recommendations on Netflix, or automatic photo grouping.

Reason: Customers are lazy about data entry. They want a smooth, fast process without having to think about what to choose or enter. AI automatically recognizes past behavior to make decisions for them.

Case 2

Prioritizing Combo Box and Textbox

Context: Services related to finance, high-value purchases, or deep personalization (such as designing travel tours, choosing insurance packages, or configuring computers).

Example

A customer buying insurance wants to select their age (combo box) and enter the amount they want to save monthly (textbox) for AI to calculate the optimal package.

Reason: When paying for important services, customers need a sense of control and security. If AI makes all the decisions automatically, they will be skeptical about transparency.

(*). Conclusion for Customers: Fully automated AI is best for convenient and quick entertainment services. AI offers the best data entry interaction for large-scale purchasing decisions requiring transparency and personalization based on real-world needs.

(**). In short: The "Hybrid" trend is optimal. To achieve the highest efficiency for both parties, modern systems often adopt a hybrid model: Fully automated by default: AI automatically pre-fills combo boxes and textboxes based on historical data or the most intelligent predictions to save time for both customers and management. Allows for advanced customization: Retains selection buttons and input fields so users (especially management or demanding customers) can edit parameters as needed.

Popular Data Analytics Platforms and Vendors

In a constantly evolving technology market, the group of popular data analytics platforms and vendors is asserting its strong position through groundbreaking cloud and artificial intelligence solutions. Alibaba Cloud dominates the Asia-Pacific market thanks to its flexible Quick BI platform, supporting both SaaS and on-premises models and deep integration with office applications like DingTalk. Competing in niche market segments, AWS attracts customers with Amazon QuickSight thanks to its cost-effectiveness, ability to integrate with existing data ecosystems, and advanced Q natural language querying capabilities. Simultaneously, Domo has become an attractive option for businesses outside the Microsoft ecosystem thanks to its robust native cloud architecture and Sandbox solution applying DevOps processes to accelerate content creation speed. Finally, Google continues to be a major challenger with its ownership of Looker—a platform notable for its standardized multi-cloud analytics capabilities, utilizing the LookML semantic layer, and continuously integrating Google Data Studio to unify the user experience while providing a flexible, extensible framework for data-driven application developers.

4.3 Applying AI and Big Data from Amazon US Reviews in Strategic Decision Making

4.3.1. Optimizing Strategic Vision Through Big Data Analytics: The Amazon dataset provides a diverse and comprehensive information structure. Centralized data storage on cloud platforms like AWS S3 as clearly separated data files allows businesses to process data at an extremely large scale (Gigabytes to Terabytes) across various industries – from Technology and Toys to Books and Fashion. Through core information fields such as geographic location (marketplace), product category, customer identifier (customer_id), and timeline (review_date), managers can: Identify market trends: Track changes in product perception across different geographic regions. Accurately segment customers: Synthesize purchasing behavior and loyalty levels of each customer group. Portfolio Strategy Planning: Analyze the parent-child product structure (product_parent, product_id) to make appropriate investment or product divestment decisions.

4.3.2. Automating and Improving Decision Quality with Artificial Intelligence (AI): If Big Data provides the "raw materials," then AI – specifically Natural Language Processing (NLP) and Machine Learning (ML) – is the "brain" that transforms that data into strategic actions. By applying AI to qualitative data fields such as review headlines and review body, combined with quantitative data such as star ratings, managers gain superior value: Real-time Sentiment Analysis: AI automatically categorizes millions of customer opinions into positive, negative, or neutral states. This helps managers detect product quality

crises early or anticipate improvement opportunities that competitors haven't yet recognized. Filtering noise and validating strategic information: Thanks to machine learning algorithms that analyze metrics such as helpful votes (total votes), experience programs (vine), or verified purchase status (verified_purchase), AI can expose and eliminate fake reviews, negative advertising trends, or competitive fraud. Managers' strategic decisions are then based on a "clean" and highly reliable data foundation.

Conclusion

Applying AI and Big Data Analytics to datasets like Amazon US Reviews is not simply a technical problem. This is a strategic shift, helping managers change their approach from "guessing based on experience" to "data-driven decision-making," minimizing risks and optimizing resources for businesses in the global market.

The deployment of the proposed AI and Big Data frameworks yields significant improvements across multiple organizational domains. In HR applications, the integration of ML and DL models successfully shifts talent acquisition from subjective bias to predictive precision, accurately calculating candidate success rates and dynamically mapping personalized employee training trajectories. Regarding sentiment analysis architectures, the evaluation reveals distinct performance tradeoffs:Supervised pipelines provide high classification metrics (accuracy, precision, and recall) but remain heavily constrained by the quality and availability of human annotations. Unsupervised approaches offer superior adaptability across large, unstructured financial and social data streams, though they exhibit a higher baseline complexity due to their reliance on the inherent organization of incoming data. The interactive design evaluation proves that a Hybrid Model delivers optimal enterprise utility. Fully automated pipelines perform optimally for low-risk, high-velocity interactions (e.g., e-commerce product recommendations on Spotify/Netflix). However, for high-stakes strategic choices managed by corporate executives, or high-value personalized consumer purchases (e.g., computer configurations or insurance plans), the inclusion of parameter-driven combo boxes/textboxes provides necessary explainability, ensuring transparent human-in-the-loop control. Our evaluation of cloud-based vendor ecosystems shows that platforms like Alibaba Cloud (Quick BI), AWS (Amazon QuickSight with Q natural language querying), Domo (Sandbox DevOps), and Google (Looker using the LookML semantic layer) successfully scale these capabilities across global enterprises. Finally, empirical application on the Amazon US Reviews dataset proves that utilizing NLP to parse unstructured customer text fields while filtering out competitive fraud via validation metrics enables managers to entirely replace traditional intuition with robust, empirical strategic planning.

5 Conclusion

Research confirms that the application of Artificial Intelligence (AI) and Big Data Analytics is not merely a technical upgrade, but a core strategic shift in modern business management. By deploying sophisticated machine learning models, businesses can effectively uncover hidden semantic patterns

from massive streams of unstructured text data (*operations, finance, customers*). The research also highlights distinct differences in application environments: Fully automated: Optimizes performance for standardized, high-speed, consumer-oriented tasks. Interactive & Explainable AI: Essential for strategic management environments, helping leaders build hypothetical scenarios ("what-if") and maintain human-in-the-loop control. Ultimately, making decisions based on clean, validated data platforms (such as large-scale cloud-based rating systems) is key for leaders to mitigate structural risks, accurately predict market trends, and transform raw data into sustainable global competitive advantages.

6 Future Work

Future research directions will focus on expanding the scalability and real-time processing capabilities of the proposed algorithmic frameworks. First, we aim to investigate the integration of multi-modal generative AI models that can simultaneously process unstructured text, financial numbers, and visual datasets to provide unified strategic summaries for executives. Second, future studies will evaluate the runtime performance and cost-efficiency tradeoffs of training large language models (LLMs) locally versus utilizing third-party cloud APIs across diverse multi-cloud environments. Finally, we intend to implement and test specific adversarial defense mechanisms within our data pipelines to further protect Explainable AI frameworks against sophisticated automated data poisoning, fake review generation, and algorithmic market manipulation.

Declarations

Ethics approval and consent to participate Not applicable (This study uses secondary data from publicly available databases and published research abstracts, and does not involve humans or animals.) .

Consent for publication Not applicable.

Availability of data and material The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request. The datasets generated and analyzed during the current study are available in the ereader.perlego.com repository.

Competing interests The authors declare that they have no competing interests.

Funding This research received no external funding.

Author contributions

- **Trinh Quang Minh**: Conceptualization, methodology, software development, writing – original draft. Algorithm design, evaluation, visualization.
- **Ngo Thi Lan**: Data curation, validation, writing – review & editing. Data preprocessing, statistical analysis. Literature review, interpretation of results.

- **Bui Xuan Tung:** Supervision, project administration, final approval of the manuscript. System architecture, technical support. Dataset preparation, experimental testing.

References

1. Alam KC (2024) *AI-Based Data Analytics Applications for Business Management, DOI: 10.1201/9781032614083, ISBN: 978-1-032-41176-7 (hbk).* Boca Raton (Florida, FL), USA. and Abingdon (Oxfordshire, UK).: CRC Press is an imprint of Taylor & Francis Group, LLC
2. Anand S (2000) *Recommender System Using Amazon Reviews*. Retrieved from Kaggle.com: https://www.kaggle.com/code/saurav9786/recommender-system-using-amazon-reviews
3. Anna Ujwary-Gil and Anna Florek-Paszkowska (2025) Artificial Intelligence, Analytics, and Strategic Decision-MakingConcepts, Practice, and Future Research Directions, DOI: 10.4324/9781003507840-1. Routledge, London, England & Wales
4. Florek-Paszkowska AU-G (2025) Artificial Intelligence, Analytics, and Strategic Decision-MakingConcepts, Practice, and Future Research Directions, DOI: 10.4324/9781003507840-1. Routledge,, London, England & Wales,
5. Himanshu RaiIndian Institute of Management, Indore I-A (2025) *Impact of Artificial Intelligence on Data-Driven Decision Making in HR for Revolutionizing Organizational Growth, ISBN: 978-1-83662-983-2 (Print), ISBN: 978-1-83662-982-5 (Online), ISBN: 978-1-83662-984-9 (Epub).* The UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center.: Emerald Publishing Limited
6. Kimbell J (2024) Decision-Making in Business Management through Artificial Intelligence and Machine Learning. GRIN, The German National Library
7. Lars Buitinck (ILPS), G. L.-I.-I (2013) 9 01). *API design for machine learning software: experiences from the scikit-learn project*. Retrieved from Arxiv.org is a free online repository for scientific research manuscripts: https://arxiv.org/abs/1309.0238
8. Mayer M (2019) Big Data and Artificial Intelligence in Management. Disruptive Technologies as a success factor for decision-making. GRIN, The German National Library
9. Nothof C (2019) How will Big Data change the way managers make decisions? Artificial intelligence and its impact on managerial decision making. Open Publishing GmbH, München,, Deutschen Nationalbibliothek