



Transactions on Data Analysis in Social Science

Journal Homepage: <https://transoscience.ir>



Machine Minds vs. Statistical Traditions: A Comparative Study in Humanities Data Analysis

M. Ghaem Tajgardoony^{1,*}, Z. Barzegar²

¹ Ministry of Science, Research and Technology, Tehran, Iran

² Advanced Technologies in Medicines Dept., Iran University of Medical Science

ARTICLE INFO	ABSTRACT
Article History: Received 10 January 2024 Received in revised form 29 March 2025 Accepted 17 May 2025 Available online 11 June 2025 Keywords: Big Data Analysis, Machine Learning, Traditional Statistics, Humanities Researches.	The rapid growth in the volume, variety and complexity of digital data has changed how research in the humanities is designed and analyzed. Much of these data are unstructured and heterogeneous, which challenges traditional statistical tools and has encouraged the adoption of machine learning (ML). This article critically reviews and compares the two paradigms as applied in humanities research. ML can identify patterns in large and unstructured datasets without predefined models, but its limited interpretability raises concerns about the transparency, reliability and accountability of findings, which are central values in the humanities. Traditional statistical methods, grounded in hypothesis testing and statistical inference, offer transparency and well-established criteria for evaluating evidence, but they are less suited to complex, high-dimensional and unstructured data. The review compares the two approaches with respect to data characteristics, research aims, interpretability and the ethical implications of analytical choices. It argues that the choice between them should be driven by the research question and the nature of the data rather than by methodological trends, and that combined strategies, such as interpretable ML or ML-based exploration followed by statistical inference, can draw on the strengths of both. The article offers practical guidance for humanities researchers in selecting analytical methods.

1. INTRODUCTION

The landscape of humanities research has undergone a profound transformation in recent years, marked by an escalating influx of digital data and the consequential need for advanced analytical methods [1-4]. As traditional approaches grapple with the vastness and complexity of contemporary datasets, the integration of data analysis techniques has become imperative for researchers seeking a nuanced understanding of human culture, history, and expression [5,6].

Historically rooted in qualitative methodologies, humanities research is now experiencing a paradigm shift, with an increasing recognition of the value brought by quantitative analysis [7-9]. The explosion of digital content, archives, and cultural artifacts has necessitated a departure from traditional manual examination toward more

* Corresponding Author: m_tajgardoony@msrt.ir
Ministry of Science, Research and Technology, Tehran, Iran



efficient, scalable, and insightful approaches [10-14]. The "quantitative turn" in humanities marks a departure from exclusive reliance on qualitative methods. Scholars are recognizing the potential of quantitative data analysis to complement and enhance traditional interpretative approaches [4,7]. Collaborations between humanities scholars and experts in data science, statistics, and machine learning are fostering innovative research methodologies. Interdisciplinary projects are harnessing the power of diverse expertise to tackle complex questions at the intersection of humanities and data analysis. Advances in text mining and natural language processing have enabled scholars to analyze large volumes of textual data efficiently [11,12]. This evolution has facilitated the exploration of linguistic patterns, sentiment analysis, and thematic trends within literary and historical texts. Geographic Information Systems (GIS) and temporal analysis tools empower researchers to explore spatial and temporal dimensions in humanities datasets, unveiling patterns that were previously challenging to discern [13-16].

In response to evolving trends and the dynamic shifts in the utilization of methodologies for data analysis, driven by structural changes in humanities data, this article aims to conduct a comparative analysis. We will contrast traditional statistical approaches, rooted in foundational statistical data analysis, with contemporary methods grounded in artificial intelligence, such as machine learning. The intended audience for this article comprises researchers specializing in humanities data analysis. The insights provided in this article aim to assist them in making informed choices between statistical and artificial intelligence methods and techniques, aligning with the specific nature of their research problems.

2. BACKGROUND

2.1. The digital shift in the humanities

The advent of digital technologies has revolutionized the way scholars in the humanities access, preserve, and analyze information [1,5,9]. Libraries and cultural institutions have digitized vast collections of texts, manuscripts, artworks, and historical records, making them accessible to a global audience. This digital archive democratizes access to knowledge, breaking down geographical barriers and facilitating collaborative research on an unprecedented scale [17]. Moreover, the digital transformation has prompted a reevaluation of traditional research methodologies in the humanities. Scholars are incorporating computational tools, data analytics, and machine learning techniques to extract patterns, trends, and connections from large datasets [18,19]. This interdisciplinary approach, known as digital humanities, harnesses the power of technology to uncover hidden narratives, reveal contextual complexities, and offer fresh perspectives on cultural phenomena [20]. The digital transformation in the humanities has notably influenced the scale and complexity of accessible data for analysis, resulting in several significant changes as outlined below [21].

2.1.1. Exponential Increase in Data Volume

The digitization of content, archives, and cultural artifacts has led to an exponential increase in the volume of available data. Digital formats allow for the storage and dissemination of vast amounts of information, including texts, images, audio, and video files. As a result, researchers now have access to extensive datasets that were previously challenging to assemble and navigate [9,15].

2.1.2. Diverse Data Types and Formats

Digital transformation has introduced diverse data types and formats into the realm of humanities research. Beyond traditional textual materials, researchers now deal with multimedia content, interactive datasets, and complex metadata. This diversity adds richness to the analytical landscape but also requires scholars to adapt their methodologies to handle varied data formats [17, 19].

2.1.3. Global Accessibility and Collaboration

Digital archives and repositories enable global accessibility to cultural materials, fostering collaboration among researchers worldwide. The scale of data available for analysis is no longer confined by geographical limitations, encouraging interdisciplinary collaborations and the exploration of cross-cultural perspectives. Scholars can engage with and analyze data without the constraints of physical proximity [19,20].

2.1.4. Temporal and Spatial Dimensions

Digital data often come with embedded temporal and spatial dimensions, providing researchers with new opportunities for analysis. Temporal datasets, such as historical records or timelines, allow for the examination of changes over time, while spatial data, such as maps or geographical information, offer insights into the spatial distribution of cultural phenomena. This added dimensionality enhances the complexity and depth of research inquiries [17,21].

2.1.5. Challenges in Data Management and Processing

While the digital shift has expanded the scope of available data, it has also introduced challenges in data management and processing. The sheer volume and complexity of digital datasets require advanced tools and methodologies for effective analysis. Scholars must address issues related to data storage, retrieval, and the development of analytical techniques capable of handling large and diverse datasets [18,19].

2.1.6. Emergence of Big Data in Humanities

The accumulation of extensive and complex datasets in the digital realm has contributed to the emergence of big data in humanities research. Big data characteristics, such as high volume, velocity, and variety, pose both opportunities and challenges. Researchers are exploring ways to harness big data analytics to extract meaningful patterns and insights from these expansive datasets [20].

In essence, the digital shift has reshaped the landscape of data available to humanities researchers, presenting both unprecedented opportunities and new complexities. Navigating this digital wealth requires a nuanced understanding of data types, advanced analytical tools, and a willingness to adapt research methodologies to the evolving nature of digital scholarship.

2.2. Big Data in Humanities

The emergence of big data in humanities research represents a significant paradigm shift, bringing about a substantial increase in the volume, velocity, and variety of data. The volume of data in humanities research has grown exponentially with the digitization of vast cultural repositories, archives, and scholarly materials. Texts, images, audio recordings, and other cultural artifacts are being digitized at an unprecedented scale, resulting in datasets of immense size. The velocity of data refers to the speed at which it is generated, collected, and processed. In the context of humanities research, the velocity has increased with the digitization of real-time data sources and dynamic content. Social media, online forums, and other digital platforms generate a continuous flow of data reflecting contemporary cultural expressions and interactions [17].

The variety of data in humanities research has expanded significantly, encompassing diverse formats, sources, and types of information. Digital archives include not only traditional textual documents but also multimedia content, geographical data, and interactive datasets. The integration of various data types allows for a more comprehensive understanding of cultural expressions and historical contexts [20].

Big data in humanities research necessitates the use of advanced computational methods and machine learning techniques to derive meaningful insights. Traditional qualitative methods are complemented by quantitative approaches, enabling researchers to uncover patterns, correlations, and trends within massive datasets [21]. Machine

learning algorithms contribute to the identification of complex relationships and the development of predictive models.

The increasing volume, velocity, and variety of data in humanities research also raise ethical considerations. Scholars must address issues related to data privacy, informed consent, and the responsible use of sensitive information. The ethical implications of big data research in the humanities underscore the importance of establishing guidelines and frameworks to ensure ethical research practices [16].

In conclusion, the emergence of big data in humanities research signifies a profound shift in the way scholars approach cultural analysis. The substantial increase in data volume, velocity, and variety opens new avenues for exploration while requiring researchers to develop innovative methodologies and address ethical considerations in this evolving digital landscape.

2.3. Challenges and Impacts of Big Data on Traditional qualitative approaches

Traditional qualitative approaches, which have long been the cornerstone of humanities research, are facing challenges in handling the vastness of contemporary datasets. Several factors contribute to the constraints of traditional methods when handling extensive and intricate datasets, as detailed in Table 1.

Table 1. Challenges and Impacts of Traditional Qualitative Methods with Large and Complex Datasets

Factor	Challenge	Impact
Volume Overwhelm	The sheer volume of contemporary datasets, often in the realm of big data, exceeds the capacity of traditional qualitative methods	Researchers find it impractical and time-consuming to manually analyze and interpret massive amounts of textual, visual, or multimedia data
Limited Scalability	Traditional qualitative methods are often labor-intensive and may lack scalability when applied to extensive datasets	As datasets grow, the scalability of traditional methods becomes a bottleneck, hindering the ability to maintain depth and detail in analysis
Inability to Uncover Patterns	Qualitative methods may struggle to identify patterns, trends, or correlations within large datasets due to their manual and subjective nature	Researchers risk missing valuable insights and may overlook subtle connections present in extensive and diverse datasets
Time-Consuming Analysis	Qualitative analysis typically involves in-depth examination and interpretation, making it time-consuming	With large datasets, the time required for thorough qualitative analysis becomes prohibitive, delaying research outcomes and limiting the timely exploration of dynamic cultural phenomena
Subjectivity and Bias	Qualitative methods can be influenced by researcher subjectivity and biases, especially when dealing with extensive datasets	The potential for subjective interpretations increases with the vastness of the data, introducing challenges in maintaining consistency and reliability across the analysis
Need for Interdisciplinary Insights	Complex datasets often require interdisciplinary insights, involving skills from data science, computer science, and quantitative methods	Traditional qualitative researchers may face difficulties in incorporating such interdisciplinary approaches, limiting their ability to extract meaningful patterns from diverse datasets
Lack of Automation	Traditional qualitative analysis is manual and does not leverage the advantages of automated tools and algorithms	Automation is crucial for handling the scale of contemporary datasets efficiently; the absence of automated processes hampers the speed and efficiency of analysis
Risk of Overlooking Context	With vast datasets, there is a risk of overlooking the rich contextual nuances that traditional qualitative approaches often emphasize	Research findings may lack the depth and context necessary for a comprehensive understanding of cultural phenomena

2.4. Role of Technology in Data Analysis

Technology plays a crucial role in enabling the exploration of patterns, trends, and relationships within datasets, offering insights that were previously challenging to unveil. The integration of computational tools, machine learning algorithms, and advanced data analysis techniques has brought about a profound transformation in the landscape of humanities research. This discussion delves into how technology facilitates the exploration of these aspects, providing detailed insights presented in Table 2.

Table 2. Challenges and Role of Technology on humanities data analysis

Key highlight	Challenge	Technology role
Handling Large and Complex Datasets	Traditional methods struggled to analyze large and complex datasets prevalent in the humanities	Computational tools and algorithms can efficiently handle vast volumes of data, enabling researchers to explore patterns across extensive corpora of texts, images, and multimedia content
Automated Pattern Recognition	Identifying subtle patterns manually in extensive datasets is time-consuming and prone to oversight	Machine learning algorithms excel in automated pattern recognition, allowing researchers to identify intricate and nuanced patterns within data. This includes textual patterns, visual motifs, and trends that may not be immediately apparent through manual analysis
Temporal Analysis	Traditional methods may struggle to analyze temporal changes and trends over extended periods	Computational tools enable temporal analysis by processing and visualizing data changes over time. Researchers can explore historical trends, cultural shifts, and evolving patterns, enhancing the understanding of temporal dynamics
Multidimensional Exploration	Traditional methods might be limited in exploring data across multiple dimensions simultaneously	Advanced analytics tools allow researchers to explore datasets in multidimensional space. This capability is crucial for understanding the complex interplay of factors and relationships across diverse aspects of humanities data
Interactive Data Visualization	Traditional static representations may hinder the exploration of dynamic relationships in datasets	Interactive data visualization tools empower researchers to dynamically explore relationships, trends, and patterns. Visualizations can be manipulated in real-time, providing a more interactive and intuitive understanding of the data
Uncovering Hidden Connections	Traditional methods may not efficiently uncover hidden connections or latent relationships within datasets	Machine learning algorithms, particularly unsupervised learning techniques, can reveal hidden connections and associations within data. This aids researchers in discovering unexpected relationships and gaining deeper insights into the underlying structures
Predictive Modeling	Traditional methods may not predict future trends or behaviors based on historical data	Machine learning models enable predictive modeling, allowing researchers to forecast trends and outcomes within humanities datasets. This forward-looking capability is valuable for understanding potential future developments
Network Analysis	Traditional network analysis may be limited in handling large-scale interconnected datasets	Computational tools facilitate network analysis on a broader scale, enabling researchers to explore complex networks of relationships, influence, and collaboration within humanities data
Semantic Analysis	Extracting nuanced semantic meanings from textual data manually is challenging	Natural Language Processing (NLP) tools assist in semantic analysis, allowing researchers to delve into the layers of meaning within texts. This capability enhances the exploration of themes, sentiments, and cultural nuances
Efficient Exploratory Data Analysis (EDA)	Traditional EDA might be limited in its ability to efficiently explore large datasets	Computational tools streamline EDA by providing efficient ways to explore data distributions, outliers, and relationships. Researchers can iteratively refine their hypotheses based on real-time insights

In essence, technology's role in humanities research extends beyond data processing; it empowers researchers to explore datasets in ways that were previously impractical or challenging. The ability to uncover hidden patterns, trends, and relationships enhances the depth and breadth of insights gained from humanities data, contributing to a more nuanced understanding of human culture and history.

3. QUANTITATIVE EVOLUTION IN HUMANITIES: ML VS TRADITIONAL STATS

The juxtaposition of machine learning and traditional statistical methods in humanities research represents a dynamic interplay between established quantitative approaches and innovative, data-driven techniques. This juxtaposition reflects the evolving landscape of research methodologies in the humanities, where scholars are increasingly integrating computational tools to complement or challenge traditional statistical traditions. In this section, we delve into the nuanced interplay between machine learning and traditional statistical methods in humanities research, capturing the shifting landscape of research methodologies. Scholars are progressively weaving computational tools into their analytical frameworks, either to enhance or question established statistical traditions, embodying a vibrant fusion of established quantitative approaches with innovative, data-driven techniques.

3.1. Traditional Statistic

Traditional statistical methods in humanities research have long been employed to analyze data, draw inferences, and test hypotheses. These methods outlined in Table 3.

Table 3. Traditional Statistical Methods for Humanities Data Analysis

Method	Description
Descriptive Statistics	Techniques such as mean, median, and standard deviation provide a summary of data distributions
Inferential Statistics	Methods like t-tests and ANOVA help researchers make inferences about populations based on sample data
Regression Analysis	Used to explore relationships between variables and make predictions
Content Analysis	Qualitative data coded into quantitative categories for analysis
Interactive Data Visualization	Traditional static representations may hinder the exploration of dynamic relationships in datasets
Uncovering Hidden Connections	Traditional methods may not efficiently uncover hidden connections or latent relationships within datasets
Predictive Modeling	Traditional methods may not predict future trends or behaviors based on historical data
Network Analysis	Traditional network analysis may be limited in handling large-scale interconnected datasets
Semantic Analysis	Extracting nuanced semantic meanings from textual data manually is challenging
Efficient Exploratory Data Analysis (EDA)	Traditional EDA might be limited in its ability to efficiently explore large datasets

Traditional statistical methods have proven valuable in exploring patterns and relationships in humanities data, providing a structured and interpretable framework for researchers.

3.2. Machine Learning

Machine learning (ML) stands as a contemporary paradigm, highlighting computational models designed to discern patterns from data autonomously, eliminating the need for explicit programming. This innovative approach marks a departure from traditional methodologies, as ML systems can iteratively refine their performance based on the patterns they identify in diverse datasets. Table 4 delineates the ML techniques utilized in humanities research.

Table 4. Machine Learning Methods for Humanities Data Analysis

Method	Description
Clustering Algorithms	Grouping data points based on inherent similarities
Natural Language Processing (NLP)	Processing and understanding human language, enabling analysis of large textual corpora
Topic Modeling	Identifying latent topics within a collection of documents
Image and Speech Recognition	Analyzing visual or auditory data for patterns and meanings
Regression Analysis	Predicting numerical outcomes by examining relationships among variables
Classification Algorithms	Categorizing data into predefined classes or groups
Sentiment Analysis	Evaluating and determining the sentiment expressed in textual data
Dimensionality Reduction	Reducing the number of input variables while retaining essential information
Anomaly Detection	Identifying unusual patterns or outliers in the data
Time Series Analysis	Analyzing data points collected over time to identify temporal patterns
Association Rule Mining	Discovering interesting relationships or patterns among variables
Network Analysis	Examining relationships and connections among entities in a network
Latent Semantic Analysis (LSA)	Extracting the underlying meaning of words in textual data
Collaborative Filtering	Recommending items or content based on user preferences and behaviors
Deep Learning	Utilizing neural networks for complex pattern recognition and feature learning
Sequence Analysis	Analyzing sequential data, such as time-ordered events or narratives
Clonal Selection Algorithms	Simulating biological evolution for optimization and pattern recognition
Evolutionary Algorithms	Applying principles of natural selection for optimization tasks
Survival Analysis	Examining the time until an event of interest occurs
Graph Embedding	Transforming graph-structured data into low-dimensional vector spaces
Transfer Learning	Leveraging knowledge from one domain to improve performance in another

Machine learning excels in handling large and complex datasets, uncovering intricate patterns, and making predictions based on learned associations.

3.3. Juxtaposition of ML vs Traditional Stats in Humanities Research

The juxtaposition of these approaches in humanities research reflects a recognition of the strengths and limitations of both. Traditional statistical methods offer interpretability and a well-established theoretical foundation. Machine learning, on the other hand, excels in handling unstructured data and discovering complex patterns. Combining both approaches allows researchers to leverage the strengths of each, providing a more comprehensive analysis.

Traditional statistics often aim for clear explanations of relationships, while machine learning models, particularly complex ones, may prioritize predictive accuracy over interpretability. The juxtaposition allows researchers to choose methods based on the goals of their study whether it's exploration, explanation, or prediction.

The juxtaposition encourages collaboration between statisticians and computer scientists. Traditional statisticians may bring expertise in experimental design and hypothesis testing, while computer scientists contribute expertise in algorithm development and handling large-scale data. Traditional statistical methods may struggle with high-dimensional or non-linear data, where machine learning algorithms can shine. By juxtaposing these methods, researchers can adapt to the specific characteristics of their data.

The evolving landscape of humanities research reflects an ongoing dialogue between traditional statistical methods and machine learning techniques. This juxtaposition not only broadens the methodological toolkit available

to researchers but also encourages a thoughtful and adaptable approach to addressing complex questions in the study of human culture and expression.

3.4. Comparative Analysis of ML vs Traditional Stats in Analyzing Large-Scale Humanities Datasets

In this section, a comparative study between "Machine Minds" and "Statistical Traditions" in humanities data analysis is provided. This involves gauging their respective capabilities in uncovering patterns, trends, and meaningful insights within diverse and expansive cultural, historical, and artistic data. Figure 1 provides an in-depth exploration of the strengths and limitations associated with both approaches Machine Minds and Statistical Traditions employed in humanities data analysis. This comparative analysis offers a comprehensive examination of the merits and constraints inherent in each methodology, facilitating a nuanced understanding of their respective contributions and potential challenges within the context of humanities research.

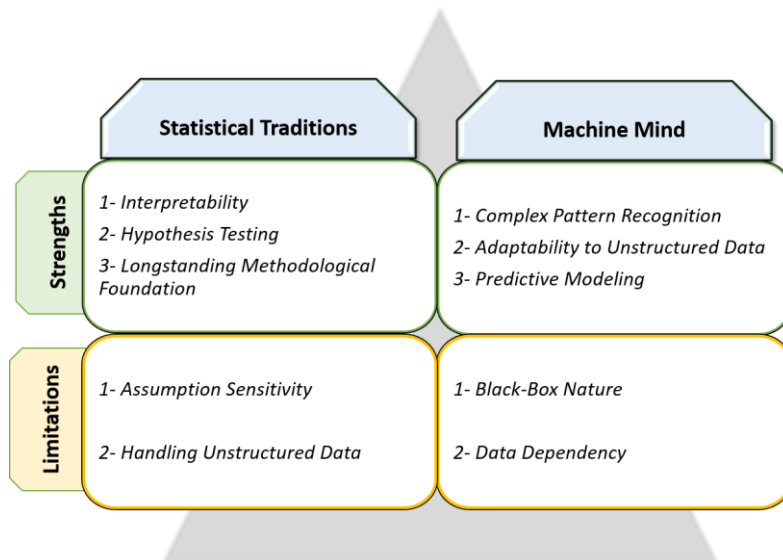


Fig. 1. Analyzing Large-Scale Humanities Datasets Machine Minds vs Statistical Traditions

Traditional statistical methods offer a clear and interpretable framework for analyzing data. Researchers can easily understand and explain the relationships between variables, enhancing the transparency of the analytical process. Statistical traditions provide robust tools for hypothesis testing, enabling researchers to make inferences about populations and draw conclusions based on statistical significance. They have a well-established methodological foundation with a history of successful application in various fields. This legacy contributes to the reliability and acceptance of results within the academic community.

Traditional statistics are sensitive to assumptions and may face challenges in handling unstructured or high-dimensional data common in humanities research, limiting their applicability to certain types of datasets.

Machine Minds, powered by advanced machine learning algorithms, excel in recognizing intricate and nonlinear patterns within large and complex datasets. This capability is particularly advantageous in capturing nuanced relationships within humanities data. Machine learning approaches are well-suited for handling unstructured data types prevalent in humanities research, such as textual content, images, and audio. Natural Language Processing (NLP) and image recognition algorithms contribute to a holistic understanding of cultural expressions. They facilitate predictive modeling, allowing for the forecasting of cultural trends, identifying potential themes, and anticipating shifts in historical or artistic movements based on learned associations.

The inherent complexity of some machine learning models can result in a "black-box" effect, where the decision-making process becomes less interpretable. This challenges the transparency and interpretability of the analytical

outcomes. Machine learning models heavily depend on the quality and representativeness of the training data. In cases of biased or unrepresentative datasets, the model may produce skewed results

3.5. Examination of Methodological Differences between ML vs Traditional Stats in Humanities Data Analysis

The methodological differences between "Machine Minds" and "Statistical Traditions" reflect distinct philosophies in approaching humanities data. While "Machine Minds" thrive in the exploration of unstructured and nonlinear patterns, statistical traditions provide a structured and hypothesis-driven framework for making inferential conclusions. The choice between these approaches hinges on the nature of the research question, the characteristics of the dataset, and the specific goals of the humanities analysis. A nuanced understanding of these methodological differences allows researchers to select the most suitable approach for their analytical needs. To understand the distinctions between "Machine Minds" and "Statistical Traditions" in handling humanities data, we need to delve into their underlying principles, assumptions, and approaches. This exploration sheds light on their contrasting philosophies, providing insights into the methodological differences that define each approach. Figure 2 meticulously outlines the major underlying principles, assumptions and approaches that delineate these two methodological frameworks, providing a detailed reference for researchers and practitioners in the field of humanities data analysis.

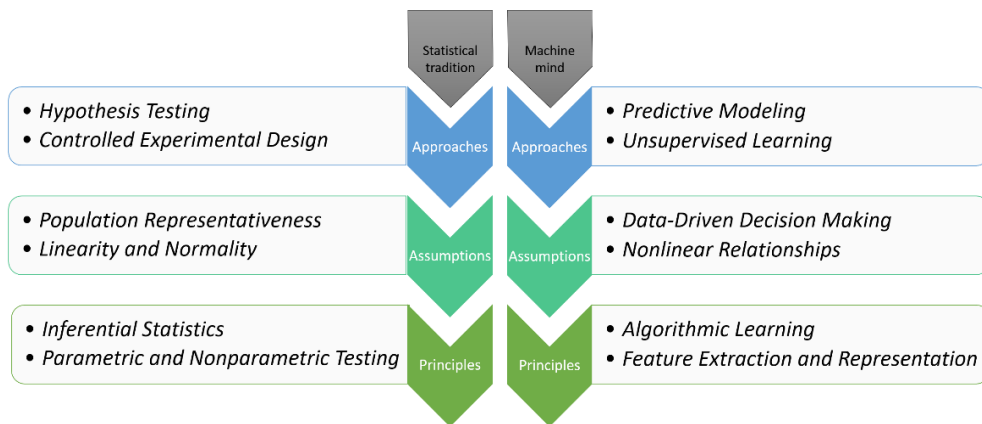


Fig. 2. Methodological Principles and Assumptions: Machine Minds vs Statistical Traditions

Statistical traditions heavily involve hypothesis testing, where researchers formulate hypotheses about relationships between variables and use statistical tests to assess the significance of these relationships. Experimental designs, including control groups and randomized sampling, are common. Traditional statistical methods often rely on controlled experiments to isolate specific variables and draw causal inferences.

Statistical traditions assume that insights gained from a representative sample can be generalized to the broader population. This is based on the idea that the analyzed subset is reflective of the characteristics of the entire humanities dataset. Assumptions regarding linearity and normality underlie many traditional statistical methods. Linear regression, for instance, assumes a linear relationship between variables, and some tests assume a normal distribution of data.

Statistical traditions are rooted in inferential statistics, aiming to draw conclusions about populations based on sample data. This involves making generalizations and inferences about broader cultural trends from a representative subset of the data. Traditional statistical methods employ parametric and nonparametric tests to analyze relationships between variables. These tests often rely on assumptions about the distribution of data and aim to provide interpretable results.

Machine Minds often prioritize predictive modeling, aiming to forecast trends, identify hidden patterns, and make informed predictions about future cultural developments based on learned associations. Unsupervised learning

approaches, such as clustering and topic modeling, are common. These methods allow the algorithm to identify patterns and structures within data without predefined labels, facilitating exploratory analysis.

The central assumption is that patterns and insights within humanities data can be effectively discovered through data-driven decision-making processes. This approach assumes that the data itself contains meaningful information that can be identified through algorithmic analysis. Machine learning methods are well-suited for capturing nonlinear relationships within data. The assumption is that cultural, historical, and artistic phenomena often exhibit complex and nonlinear patterns that can be better understood through advanced modeling techniques

Machine Minds are characterized by algorithmic learning, where models autonomously learn patterns and associations from data without explicit programming. This enables them to adapt and evolve in response to the complexities inherent in humanities datasets. Machine learning algorithms often emphasize the automated extraction and representation of features from raw data. This is particularly advantageous for unstructured humanities data, such as text, images, and audio, where patterns may not be readily apparent.

The choice between "Machine Minds" and "Statistical Traditions" depends on the specific goals and characteristics of the humanities research in question. Machine learning excels in uncovering complex patterns and adapting to diverse data types, making it particularly valuable for exploratory analyses. However, the interpretability challenges and data dependency should be carefully considered. On the other hand, statistical traditions provide a transparent and well-established framework for hypothesis testing, offering reliability but with potential limitations in handling certain types of data.

A balanced approach may involve leveraging the strengths of both methodologies, utilizing "Machine Minds" for exploratory insights and "Statistical Traditions" for confirmatory analyses. This integrative strategy can lead to a more comprehensive understanding of large-scale humanities datasets, combining the adaptability of machine learning with the interpretability of traditional statistical methods.

4. CONCLUSION

The choice between "Machine Minds" and "Statistical Traditions" in humanities research is contingent upon the unique goals and characteristics of each study. Machine learning, with its proficiency in uncovering complex patterns across diverse data types, proves valuable for exploratory analyses. However, considerations of challenges related to interpretability and data dependency must be meticulous. In contrast, statistical traditions provide a transparent framework for hypothesis testing, ensuring reliability while acknowledging potential limitations with certain data types.

A well-balanced strategy entails leveraging the strengths of both methodologies utilizing "Machine Minds" for exploratory insights and resorting to "Statistical Traditions" for confirmatory analyses. This integrated approach fosters a more comprehensive understanding of large-scale humanities datasets by amalgamating the adaptability of machine learning with the interpretability inherent in traditional statistical methods.

Methodological distinctions between "Machine Minds" and "Statistical Traditions" mirror differing philosophies in humanities data analysis. While "Machine Minds" excel in exploring unstructured patterns, statistical traditions offer a structured framework for drawing inferential conclusions. The selection between these approaches is contingent on factors such as the research question, dataset characteristics, and specific goals, necessitating a nuanced understanding for researchers to choose the most suitable approach for their analytical needs.

The objective evaluation of the analytical performance of "Machine Minds" and "Statistical Traditions" through predefined metrics enables a quantitative assessment. Researchers can tailor their approach based on performance criteria and the nature of the data, taking ethical considerations into account. Proactive measures in ethical considerations ensure fairness, transparency, and responsible data use in both approaches. The ethical framework should adapt alongside technological advancements, fostering a culture of responsible research practices in

humanities analysis. Ultimately, the choice between these methodologies should align with the research question, data characteristics, transparency requirements, and ethical considerations, allowing for adaptability as research progresses.

Authors' Contributions

All authors read and approved the final manuscript.

Declaration of Generative AI Use

During the preparation of this work, the authors used AI to improve the language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project

Declaration of Competing Interests

The authors declare no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Considerations

This article is a review based on previously published literature. It did not involve human participants or the collection of new data; therefore, ethics approval and informed consent were not required.

REFERENCES

- [1] Tollenaar, N., & van der Heijden, P. G. (2013). Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 176(2), 565-584. <https://doi.org/10.1111/j.1467-985X.2012.01056.x>
- [2] Garg, A., & Tai, K. (2013). Comparison of statistical and machine learning methods in modelling of data with multicollinearity. *International Journal of Modelling, Identification and Control*, 18(4), 295-312. <https://doi.org/10.1504/IJMIC.2013.053535>
- [3] Boulesteix, A. L., & Schmid, M. (2014). Machine learning versus statistical modeling. *Biometrical Journal*, 56(4), 588-593. <https://doi.org/10.1002/bimj.201300226>
- [4] Yoo, C., Ramirez, L., & Liuzzi, J. (2014). Big data analysis using modern statistical and machine learning methods in medicine. *International neurourology journal*, 18(2), 50. <https://doi.org/10.5213/inj.2014.18.2.50>
- [5] Malhotra, R. (2014). Comparative analysis of statistical and machine learning methods for predicting faulty modules. *Applied Soft Computing*, 21, 286-297. <https://doi.org/10.1016/j.asoc.2014.03.032>
- [6] Iniesta, R., Stahl, D., & McGuffin, P. (2016). Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological medicine*, 46(12), 2455-2465. <https://doi.org/10.1017/S0033291716001367>

- [7] Shirzadi, A., Shahabi, H., Chapi, K., Bui, D. T., Pham, B. T., Shahedi, K., & Ahmad, B. B. (2017). A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena*, 157, 213-226. <https://doi.org/10.1016/j.catena.2017.05.016>
- [8] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3), e0194889. <https://doi.org/10.1371/journal.pone.0194889>
- [9] Berrocal, V. J., Guan, Y., Muyskens, A., Wang, H., Reich, B. J., Mulholland, J. A., & Chang, H. H. (2020). A comparison of statistical and machine learning methods for creating national daily maps of ambient PM_{2.5} concentration. *Atmospheric Environment*, 222, 117130. <https://doi.org/10.1016/j.atmosenv.2019.117130>
- [10] Jamthikar, A., Gupta, D., Saba, L., Khanna, N. N., Araki, T., Viskovic, K., ... & Suri, J. S. (2020). Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models. *Cardiovascular diagnosis and therapy*, 10(4), 919. <https://doi.org/10.21037/cdt.2020.01.07>
- [11] Castilho, C. M. (2022). Time Series Forecasting With Exogenous Factors: Statistical vs. Machine Learning Approaches (Doctoral dissertation, Universidade do Porto (Portugal)).
- [12] Spiliotis, E., Makridakis, S., Semenoglou, A. A., & Assimakopoulos, V. (2022). Comparison of statistical and machine learning methods for daily SKU demand forecasting. *Operational Research*, 22(3), 3037-3061. <https://doi.org/10.1007/s12351-020-00605-2>
- [13] Huang, F., Cao, Z., Guo, J., Jiang, S. H., Li, S., & Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena*, 191, 104580. <https://doi.org/10.1016/j.catena.2020.104580>
- [14] Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., & Fanos, V. (2020). Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina*, 56(9), 455. <https://doi.org/10.3390/medicina56090455>
- [15] Yarotskaya, L. V., & Aleinikova, D. V. (2023). Reviewing learning and teaching content in the scope of artificial intelligence: For humanities and social sciences majors. *RUDN Journal of Psychology and Pedagogics*, 20(1), 145-162. <https://doi.org/10.22363/2313-1683-2023-20-1-145-162>
- [16] Richards, C. E., Cernev, T., Tzachor, A., Zilgalvis, G., & Kaleagasi, B. (2023). Safely advancing a spacefaring humanity with artificial intelligence. *Frontiers in Space Technologies*, 4, 1199547. <https://doi.org/10.3389/frspt.2023.1199547>
- [17] El-Hajj, H., Eberle, O., Merklein, A., Siebold, A., Shlomi, N., Büttner, J., ... & Valleriani, M. (2023). Explainability and transparency in the realm of digital humanities: toward a historian XAI. *International Journal of Digital Humanities*, 5(2), 299-331. <https://doi.org/10.1007/s42803-023-00070-1>
- [18] Bharadiya, J. P. (2023). A comparative study of business intelligence and artificial intelligence with big data analytics. *American Journal of Artificial Intelligence*, 7(1), 2024.
- [19] Peters, M. A., Jackson, L., Papastephanou, M., Jandrić, P., Lazaroiu, G., Evers, C. W., ... & Fuller, S. (2023). AI and the future of humanity: ChatGPT-4, philosophy and education-Critical responses. *Educational Philosophy and Theory*, 1-35. <https://doi.org/10.1080/00131857.2020.1867108>
- [20] Hazzan, O., & Mike, K. (2023). Data Science for Social Science and Digital Humanities Research. In *Guide to Teaching Data Science: An Interdisciplinary Approach* (pp. 283-301). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-24758-3_19

- [21] Hajkowicz, S., Sanderson, C., Karimi, S., Bratanova, A., & Naughtin, C. (2023). Artificial intelligence adoption in the physical sciences, natural sciences, life sciences, social sciences and the arts and humanities: A bibliometric analysis of research publications from 1960-2021. *Technology in Society*, 74, 102260. <https://doi.org/10.1016/j.techsoc.2023.102260>