International Journal of Bio-resource and Stress Management https://ojs.pphouse.org/index.php/IJBSM

Crossref

IJBSM June 2026, 17(6): 01-10

Article AR7025

Research Article

Natural Resource Management

DOI: HTTPS://DOI.ORG/10.23910/1.2026.7025

Adoption of Farm Technologies Across the Terai Region of West Bengal: A Comparative Topic Modelling Study of Empirical Literature Review

Debraj Saha[1], Soumyadeep Thakur[2] and Kalyan Kanti Das[1]

[1]Dept. of Agricultural Economics, Uttar Banga Krishi Viswavidyalaya, Pundibari, CoochBehar, West Bengal (736 165), India
[2]Dept. of Agricultural Extension Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal (741 252), India

Open Access

Corresponding soumyadeepthakur519@gmail.com

0009-0008-3637-6380

ABSTRACT

This study was executed from June, 2025 to January, 2026 in the Terai region of West Bengal mainly in Jalpaiguri and Cooch Behar district. The objectives were to identify technologies introduced in KVK villages over the past years, compare adoption gaps and performance between intervened and non-intervened farmers, analyse socio-technical and economic determinants, and design a feasible extension roadmap. To achieve these objectives, a systematic qualitative comparative analysis using literature review and advanced topic modelling was undertaken. A curated corpus of forty-five high-quality researches was constructed through rigorous inclusion criteria and quality assessment, followed by preprocessing and implementation of five computational approaches: Non-negative Matrix Factorization (NMF), Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and optimized variants. Multi-metric evaluation integrated reconstruction error, perplexity, topic diversity, interpretability, and computational efficiency. Results revealed consistent thematic clusters across models, including the central role of KVK as institutional hubs, gendered pathways of adoption through SHGs and training, and regional specificity, with Jalpaiguri and Cooch Behar emerging as technology adoption hotspots. Technology-specific themes such as solar irrigation under PM-KUSUM, mechanization via power tillers, direct-seeded rice, and integrated pest management appeared across findings. Divergence patterns highlighted methodological trade-offs like NMF with count data produced granular differentiation, while probabilistic LDA yielded more coherent narrative themes. The findings underscored both robust thematic structures and methodology-dependent artifacts, positioning topic modelling as a valuable but interpretive tool. This study provided a data-driven foundation for inclusive agricultural strategies tailored to smallholder realities in South Asia.

KEYWORDS: Technology adoption, literature review, topic modelling

Citation* (VANCOUVER)*: Saha et al., Adoption of Farm Technologies Across the Terai Region of West Bengal: A Comparative Topic Modelling Study of Empirical Literature Review. *International Journal of Bio-resource and Stress Management,* 2026; 17(6), 01-10. HTTPS://DOI.ORG/10.23910/1.2026.7025.

Data Availability Statement: Legal restrictions are imposed on the public sharing of raw data. However, author have full right to transfer or share the data in raw form upon request subject to either meeting the conditions of the original consents and the original research study. Further, access of data needs to meet whether the user complies with the ethical and legal obligations as data controllers to allow for secondary use of the data outside of the original study.

Conflict of interests: The author has declared that no conflict of interest exists.

RECEIVED on 02nd March 2026 RECEIVED in revised form on 12th May 2026 ACCEPTED in final form on 27th May 2026 PUBLISHED on 01st June 2026

1. INTRODUCTION

Agriculture in India is undergoing a profound transformation driven by technological advancements, shifting markets, and climatic uncertainties (Anonymous, 2023). This transition is particularly complex in Northern West Bengal, spanning high-altitude hills to fertile Terai plains, creating diverse opportunities and challenges for innovation diffusion (Anonymous, 2022). While some areas have adopted mechanization and modern practices, others rely on traditional or low-cost organic inputs due to terrain constraints (Anonymous, 2021). Consequently, technology adoption is not merely an individual decision but a socio-technical process shaped by ecological risks, institutional support, and local knowledge systems. Recent studies show that smallholder farmers in the Gangetic plains, including parts of West Bengal, face significant barriers to adopting climate-smart agriculture (CSA), particularly due to groundwater over-exploitation, land fragmentation, and limited extension services (Sengupta et al., 2025). In the Indo-Gangetic Plains, adoption of conservation agriculture and mechanization remains uneven, shaped by factors such as cooperative membership, extension access, finance, education, age, gender, and risk perception (Kumar et al., 2023). Persistent gender disparities further restrict women's access to technologies and services, increasing vulnerability among marginal farmers (Anonymous, 2023). Empirical evidence from West Bengal shows that sericulture farmers' technology adoption is influenced by education, training, and management practices, while uptake of digital and precision tools remains limited due to awareness gaps and cost. In hazard-prone areas, CSA adoption can enhance food security, with economic incentives and peer influence playing key roles (Geda and Kuhl, 2025). Mechanization adoption is further constrained by plot size, tenancy, and environmental factors (Paudel et al., 2020). Despite decades of research highlighting landholding size, education, and credit access as key drivers of change (Feder et al., 1985), adoption remains uneven across agro-climatic zones (Anonymous, 2024; Ranganathan et al., 2018). Critical issues, including gender disparities in extension services and the needs of marginal smallholders, are often overlooked in regional analyses (Anonymous, 2023; Sapbamrer and Thammachai, 2021). Although tools like topic modelling help synthesize large bodies of literature, their effectiveness depends on the underlying algorithms (Dissanayake et al., 2022), with limited rigorous evaluation in agricultural development contexts (Anonymous, 2022; Zeweld et al., 2017). A growing body of work employs integrated models (e.g., TAM-TPB) to explain adoption intentions for ecological technologies, while meta-analyses and systematic reviews underscore the role of perceived benefits, economic incentives, and social influences (Ruzzante et al., 2021). Theories such as the Theory of Planned Behaviour, Diffusion of Innovation, and Unified Theory of Acceptance and Use of Technology have been widely applied to understand adoption in developing contexts. This study addresses these gaps by conducting a systematic, comparative analysis of multiple topic modelling techniques applied to a curated corpus of research on farm technology adoption in Northern West Bengal. By moving beyond a simple narrative review, the research seeks to identify the core socio-economic and institutional factors that consistently influence how farmers adopt new tools (Spiegel et al., 2021). The study intends to map regional disparities, highlight persistent barriers for vulnerable groups, and evaluate which computational methods provide the most accurate and interpretable insights for policymakers (Anonymous, 2025). Ultimately, this work provides a data-driven foundation for more inclusive agricultural strategies that are better aligned with the diverse realities of farmers in South Asia (Anonymous, 2025).

2. MATERIALS AND METHODS

The study was carried out from June, 2025 to January, 2026, specifically to identify technologies introduced in KVK intervened villages over the past years, compare adoption gaps and performance between intervened and non-intervened farmers, analyse technical, social, and economic determinants of adoption, and design a feasible extension roadmap for effective scaling. Aligning with the objectives the study followed a multi-stage methodological framework (Page et al., 2021) integrating systematic literature review protocols with advanced computational text mining (Egger and Yu, 2022) and unsupervised machine learning techniques (Grootendorst, 2022). The research design was structured around three interconnected phases of corpus construction, preprocessing (Bird et al., 2009), comparative topic modelling implementation (Egger and Yu, 2022), and multi-criteria performance evaluation (Röder et al., 2015). Following an adapted PRISMA framework (Page et al., 2021), comprehensive bibliometric searches (Donthu et al., 2021) across multiple academic databases, including Scopus, Web of Science, and Google Scholar, were carried out, covering the temporal window from 2000 to 2024. Aligned with this methodological rigor, The searching strategy employed Boolean operators and nested query structures (Donthu et al., 2021), combining domain-specific keywords with geographical identifiers to capture variations in terminology related to agricultural technology adoption, farm mechanization, irrigation systems, institutional extension services, and gender-inclusive agricultural development within the West Bengal context.

The initial retrieval phase (Page et al., 2021) yielded 312 potentially relevant documents across peer-reviewed journals, working papers, and institutional reports from Krishi Vigyan

Kendras, Agricultural Technology Management Agencies, and state-level agricultural departments. Duplicate word removal was performed using hash-based deduplication algorithms that identified exact and near-duplicate records based on metadata fields including title, author list, and publication year. Following deduplication, seventy-four documents were retained for full-text assessment through a structured decision tree incorporating multiple inclusion and exclusion filters. Inclusion criteria (Page et al., 2021) required that studies report primary or secondary empirical data on technology adoption with explicit focus on northern districts of West Bengal, examine determinants of adoption through quantitative or qualitative methods (Feder et al., 1985), and document measurable outcomes related to productivity, economic returns, resource use efficiency, or social impacts. Quality assessment was carried out using a composite scoring system that considered three key dimensions. The first dimension was methodological rigor, which examined whether the sample size was adequate, the sampling strategy was appropriate, and the statistical approach was sufficiently sophisticated. The second dimension was evidential clarity, focusing on the transparency of reported findings and the inclusion of supporting data visualizations. The third dimension was contextual relevance, which measured how directly the study applied to the agro-climatic conditions of Northern West Bengal. Each of these dimensions was rated on a five-point Likert scale (Likert, 1932), and studies with aggregate scores below three were excluded. This process resulted in a final selection of 45 high-quality empirical studies.

The selected documents underwent extensive text preprocessing (Thakur and Basu, 2025) to transform unstructured narrative content into structured numerical representations suitable for machine learning algorithms. Document digitization was performed using optical character recognition (OCR) (Smith, 2007) for printed materials and direct text extraction for digital formats, followed by a multi-stage cleaning pipeline beginning with case normalization, punctuation removal, and special character filtering. Tokenization (Bird et al., 2009) decomposed continuous text streams into discrete lexical units using whitespace and punctuation boundaries as delimiters. Stop word removal (Schofield et al., 2017) eliminated high-frequency, low-information words using a custom list combining standard English stop words with domain-specific exclusions identified through iterative frequency analysis. Lemmatization (Egger and Yu, 2022) reduced inflectional and derivational word forms to their base forms using part-of-speech-aware algorithms that considered grammatical context to disambiguate homographs. Following lexical preprocessing, two alternative document-term matrix representations were constructed: raw term frequency counts creating a sparse matrix with integer count values, and Term Frequency-Inverse Document Frequency (TF-IDF) transformation (Sparck Jones, 1972), computing the product of normalized term frequency and logarithmically scaled inverse document frequency. Both matrix representations underwent vocabulary pruning that removed terms appearing in fewer than two documents or in more than 95% of documents, eliminating idiosyncratic noise and universally common terms.

The core analytical component involved parallel implementation of five distinct topic modelling techniques representing different mathematical paradigms for latent semantic structure discovery. Non-negative Matrix Factorization (NMF) with TF-IDF input (Lee and Seung, 1999) decomposed the document-term matrix into two lower-dimensional non-negative matrices through constrained optimization using multiplicative update rules with alternating least squares minimization, initialized using non-negative double singular value decomposition and converging after four iterations. Standard Latent Dirichlet Allocation (LDA) (Blei et al., 2003) modelled documents as finite mixtures over latent topics with Dirichlet priors, performing inference through variational Bayes approximation using variational expectation-maximization algorithm across 50 iterations with symmetric Dirichlet hyperparameters and perplexity monitoring on held-out data. Latent semantic analysis (LSA) via Singular value decomposition (SVD) (Deerwester et al., 1990) applied truncated SVD to decompose the document-term matrix into left singular vectors, singular values, and right singular vectors, retaining the top eight components corresponding to the largest singular values across hundred iterations. Non-negative Matrix Factorization with count-based input employed Frobenius norm minimization with L2 regularization on raw term frequency data, converging after forty-two iterations. Optimized Latent Dirichlet Allocation incorporated asymmetric Dirichlet priors learned through empirical Bayes estimation with online learning adaptations, processing documents in mini-batches across hundred iterations. Each implementation extracted eight latent topics based on preliminary exploratory analysis and domain knowledge regarding expected thematic diversity.

Model performance was evaluated using a comprehensive multi-criteria framework (Egger and Yu, 2022) integrating quantitative metrics, qualitative interpretability assessment, and computational efficiency profiling. Topic diversity (Nan et al., 2019) was computed by analysing word overlap across topic-word distributions, calculating the proportion of unique top-weighted terms appearing in only one topic versus being shared across multiple topics, and examining the top twenty terms per topic to construct a uniqueness index ranging from zero to one. For probabilistic models, perplexity (Blei et al., 2003) measured generalization performance through the exponential of negative log-likelihood per word, computed using five-fold cross-validation with corpus partitioning into training and test sets. For matrix factorization approaches, reconstruction error quantifies low-rank factorization approximation quality through the Frobenius norm (Lee and Seung, 1999) of the difference between original and reconstructed matrices. Interpretability (Chang et al., 2009) was assessed through qualitative expert review by domain specialists in agricultural economics and rural development, examining top-weighted terms for semantic coherence, relevance to known adoption themes, and actionability for research applications.

3. RESULTS AND DISCUSSION

The Results and Discussion section was organized into two parts, reflecting the dual focus of the study. The first part highlighted the thematic structures identified through comparative topic modelling (Table 1), emphasizing convergence and divergence across different computational techniques. The second part presented the multi-metric evaluation of these models (Table 2), analysing their performance in terms of diversity, interpretability, and efficiency. Together, these results provided both substantive insights into technology adoption and methodological reflections on topic modelling approaches.

The comparative topic modelling exercise revealed complex thematic architecture underlying farm technology adoption research in Northern West Bengal, with substantive convergence and divergence patterns across the five computational approaches in Table 1 (Feder et al., 1985; Ruzzante et al., 2021). Although the methods relied on different mathematical foundations like matrix factorization, probabilistic generative modelling, and dimensionality reduction, they all consistently revealed core thematic clusters that formed the pillars of adoption discourse in the region. (Feder and Umali, 1993). The institutional centrality of Krishi Vigyan Kendras emerged as a universal theme across all models, appearing prominently in topics extracted by NMF-TF-IDF, standard LDA, LSA-SVD, NMF-Count, and optimized LDA. This cross-model consistency validated the empirical reality that KVKs functioned as primary knowledge brokers and technology dissemination hubs in Northern West Bengal's agricultural innovation system. Similarly, gender dimensions of technology adoption surfaced repeatedly, with women's participation, training access, income generation, and involvement in self-help groups constituting distinct thematic clusters across multiple models (Aryal et al., 2020; Sapbamrer and Thammachai, 2021). This persistent emergence of gender-related themes reflected both the substantial empirical attention devoted to women's agricultural roles in the reviewed literature and the genuine significance of gender as a structural determinant of adoption patterns in the region.

Regional specificity represented another area of remarkable thematic convergence, with Jalpaiguri, Cooch Behar, Kalimpong, Malda, and other districts appearing as

Table 1: Comparative topic modelling

Topic ID	NMF_TF-IDF	LDA_Standard	LSA_SVD	NMF_Count	LDA_Optimized
1	farmers, adoption, use, jalpaiguri, kvk, cooch, cooch behar, behar, cost, yield	women, months, irrigation, labour, average, income, input, kusum, pm kusum, solar	farmers, adoption, use, kvk, jalpaiguri, women, cooch behar, behar, cooch, units	ipm, hyv, seeds, hyv seeds, pest, farmers, management, adoption, pest management, practices	labour, technologies, input, irrigation, mechanization, women, northern, northern west, west bengal, bengal
2	units, women, kvk, income, training, average, farmers, adoption, shg, kalimpong	jalpaiguri, use, increase, water, seasons, sustained, farmers, kvk, dsr, water savings	units, women, income, shg, training, monthly, average, kalimpong, farmers, mushroom	kvk, yield, water, use, jalpaiguri, sustained, farmers, farms, increase, seasons	farmers, months, marginal farmers, marginal, revealed, survey, tillers, power, power tillers, chcs
3	adoption, farmers, west, west bengal, bengal, northern west, northern, months, higher, women	survey, showed, adoption, hyv, ipm, hectares, kvk, pest, cooch, cooch behar	west bengal, months, northern, northern west, use, agricultural, jalpaiguri, technology	adoption, west, bengal, west bengal, northern, northern west, agricultural, credit, extension, higher	use, farmers, jalpaiguri, flood, sustained use, survey, hectares, showed, timing, meghdoot

Table 1: Continue...

Topic ID	NMF_TF-IDF	LDA_Standard	LSA_SVD	NMF_Count	LDA_Optimized
4	farmers, adoption, women, use, kvk, jalpaiguri, units, behar, cooch behar, cooch	adoption, farmers, flood, post, training, kvk, women, improved, units, prone	women, marginal farmers, marginal, water, farms, power, tillers, power tillers, months, yield	farmers, marginal farmers, marginal, use, chcs, tillers, power tillers, power, small, cooch	women, units, income, training, shg, mushroom, kvk, monthly, cultivation, led
5	farmers, adoption, use, cost, kvk, jalpaiguri, units, cooch, cooch behar, behar	farmers, marginal, marginal farmers, use, cooch, cooch behar, behar, power tillers, chcs, power	solar, kusum, pm kusum, pm, cost, months, ipm, dinajpur, issues, subsidy	women, units, income, training, shg, mushroom, kalimpong, led, average, monthly	average, increase, season, yield, jalpaiguri, farmers, supported, unit, seasons, established
6	farmers, adoption, use, kvk, jalpaiguri, women, cooch behar, behar, cooch, cost	units, adoption, pest management, cooch behar, management, behar, pest, kvks, cooch, farmers	turnaround, cropping intensity, fields, intensity, cropping, rice, malda, mechanization, fpo, mechanized	rice, adoption, behar, cooch, cooch behar, cost, dsr, seeded, direct, direct seeded	adoption, ipm, seeds, kvks, management, hyv, pest management, behar, cooch behar, cooch
7	farmers, adoption, use, kvk, jalpaiguri, cooch, cooch behar, behar, cost, women	zones, adoption, systems, access, cropping, farmers, icar, bengal, unit, climate	hectares, turnaround, intensity, fields, cropping intensity, cropping, seeds, fpo, malda, water	months, solar, pm kusum, cost, irrigation, average, subsidy, adoption	systems, water, drought, millet, adoption, yield, based, technology, zones, level
8	farmers, adoption, use, kvk, jalpaiguri, women, units, behar, cooch behar, cooch	rice, direct seeded, seeded, rice dsr, seeded rice, direct, adoption, dsr, yield, kalimpong	systems, based, supported, polyhouse, icar, unit, season, kalimpong, alipurduar, years	cropping, malda, turnaround, fields, intensity, cropping intensity, rice, hectares, mechanization, mechanized	adoption, farms, farmers, use, jalpaiguri, reduced, trials, rice, kvk, darjeeling

Source- Authors' analysis

distinctive geographical indicators across topic models (Ranganathan et al., 2018; Sengupta et al., 2025). The Terai belt districts of Jalpaiguri and Cooch Behar dominated adoption narratives across all techniques, appearing in multiple topics within each model configuration. This spatial concentration suggested these districts functioned as technology adoption hotspots, likely due to their favourable agro-climatic conditions, better infrastructure connectivity, higher institutional density of extension services, and greater market integration compared to the more geographically isolated hill districts. Technology-specific themes also exhibited cross-model stability, with irrigation technologies, particularly solar pumps under the PM-KUSUM scheme, mechanization through power tillers and custom hiring centres, direct-seeded rice cultivation, integrated pest management, and high-yielding variety seeds forming coherent thematic clusters (Kumar et al., 2023). The recurrence of these technological categories across diverse modelling approaches indicated they represented genuine focal points of adoption research rather than methodological artifacts. However, the degree of thematic specificity and separation varied substantially across techniques. NMF with count-based input produced the most granular thematic differentiation, generating topics with minimal word overlap such that mechanization, irrigation, pest management, climate adaptation, and gender empowerment emerged as clearly delineated themes. In contrast, NMF with TF-IDF weighting yielded highly overlapping topics where core

Table 2: Multi-metric Analysis of Topic Modelling Techniques

Technique	Approach	Input type	Reconstruction error	Perplexity	Iterations	Topic diversity	Diversity rank	Interpretability	Speed	Interpretation
NMF (TF-IDF)	Matrix Factorization	TF-IDF	7.874	–	4	0.3	5	High	Fast	Low diversity–topics overlap significantly
LDA (Standard)	Probabilistic	Count	–	327.2565	50	0.75	4	Very high	Medium	Good diversity–Balanced topic separation
LSA (SVD)	Dimensionality Reduction	TF-IDF	–	100	100	0.7875	3	Medium	Very fast	Good diversity–Distinct semantic clusters
NMF (Count)	Matrix Factorization	Count	23.4731	–	42	0.875	1	High	Fast	Highest diversity – Most unique words per topic
LDA (Optimized)	Probabilistic	Count	–	482.2117	100	0.85	2	Very high	Slow	High diversity – Well-separated topics

Source: Authors' Analysis

terms like farmers, adoption, use, KVK, and district names appeared repeatedly across multiple topics, suggesting this configuration struggled to identify distinctive semantic structures within the corpus (Dissanayake et al., 2022).

The probabilistic LDA-based approaches demonstrated superior capacity for generating interpretable narrative themes that aligned with established conceptual frameworks in agricultural economics and rural development (Dissanayake et al., 2022). Standard LDA extracted coherent topics distinguishing between input-oriented themes encompassing women, labour, irrigation technologies, and income effects; institutional themes centred on KVKs, extension systems, and training programs; farmer typology themes differentiating marginal and small farmers; and technology-specific themes focusing on flood-tolerant varieties, direct-seeded rice, and pest management practices. The optimized LDA variant further refined these distinctions, producing topics that cleanly separated mechanization technologies from climate adaptation strategies, institutional actors from farmer characteristics, and economic outcomes from social participation dimensions. This enhanced thematic clarity stemmed from the asymmetric Dirichlet priors and extended inference iterations that allowed the model to identify hierarchical topic structures where certain themes exhibited corpus-wide relevance while others remained specialized to particular sub-literatures. LSA, through singular value decomposition, offered a middle ground, extracting latent semantic dimensions that captured major thematic axes but sometimes blended conceptually distinct elements within single components. For instance, LSA Topic 6 combined cropping intensity, rice cultivation systems, mechanization, and farmer-producer organizations, elements that represented distinct analytical dimensions in agricultural adoption research but co-occurred frequently enough in documents to form a unified latent semantic dimension (Spiegel et al., 2021; Anonymous, 2022).

Beyond these macro-thematic patterns, the topic modelling outputs revealed several substantive insights with implications for understanding adoption dynamics in Northern West Bengal. The prominence of marginal and small farmer categories across multiple topics and models underscored that smallholder agriculture dominated the regional farming landscape, and adoption patterns could not be understood through frameworks developed for large-scale commercial agriculture (Feder et al., 1985; Paudel et al., 2020). The frequent co-occurrence of training, income, and self-help groups in topics related to women's participation suggested that gender-inclusive technology adoption operated through specific institutional pathways, particularly group-based approaches that combined skill development with collective economic organization (Aryal et al., 2020; Sapbamrer and Thammachai, 2021). The emergence of climate adaptation as a distinct theme encompassing flood-tolerant varieties, drought-resistant crops like millet, water management strategies, and agro-advisory services like Meghdoot reflected the growing salience of climatic uncertainty as a driver of technology choice in the region. Interestingly, topics related to climate adaptation exhibited a stronger association with specific districts experiencing particular climate stresses, such as flood-prone areas of the Terai belt adopting submergence-tolerant rice varieties and drought-affected zones experimenting with alternative cropping systems (Mitra et al., 2024; Sengupta et al., 2025).

The divergence patterns across models proved equally informative, revealing how methodological choices shaped the analytical lens through which adoption themes were identified and interpreted. Techniques prioritizing topic diversity, particularly NMF with count data, generated specialized topics that isolated specific technological domains, institutional mechanisms, and farmer segments. This granular decomposition facilitated targeted analysis of particular adoption pathways but might have fragmented interconnected processes that operated holistically in farmers' decision-making contexts. Conversely, techniques producing lower diversity scores, especially NMF with TF-IDF, created broader, more integrative topics that captured the multi-dimensional nature of adoption but sacrificed analytical precision regarding specific determinants or outcomes. The differential treatment of economic dimensions illustrated this trade-off: high-diversity models separated cost-related themes from yield and income themes, enabling distinct analysis of adoption barriers versus adoption benefits, while lower-diversity models combined these elements into unified economic outcome topics that better reflected farmers' holistic economic calculus but complicated isolation of specific causal mechanisms. Similarly, institutional themes showed varying degrees of differentiation across models, with some configurations distinguishing KVK roles from credit access from extension services, while others amalgamated institutional support into broader enabling environment topics (Zeweld et al., 2017).

The methodological implications extended beyond technique selection to broader questions about the epistemology of computational text analysis in agricultural research. Topic modelling did not discover pre-existing, objective thematic structures waiting to be revealed but rather constructed interpretive frameworks shaped by algorithmic assumptions, input representations, and parameterization choices (Ruzzante et al., 2021; Dissanayake et al., 2022). The extracted eight topics represented one among many possible decompositions of the semantic space, and alternative topic numbers yielded different thematic resolutions. Fewer topics produced coarser distinctions, collapsing the gender, mechanization, and climate themes into broader technology adoption categories, while more topics fragmented current themes into finer-grained sub-topics, distinguishing specific crops, technologies, or institutional actors. This inherent subjectivity positioned it as a structured approach to generate interpretive hypotheses that had to be validated against substantive domain knowledge and triangulated with other evidence sources. The remarkable consistency of certain themes across all five techniques, namely KVK importance, gender dimensions, regional specificity, and core technology categories, suggested these represented robust features of the adoption literature that transcended methodological variation. In contrast, themes that appeared in only one or two models, or that exhibited substantial compositional differences across techniques, had to be interpreted more cautiously as potentially methodology-dependent patterns requiring further empirical validation. This methodological pluralism, comparing multiple topic modelling approaches rather than relying on a single technique, thus served as an internal validity check that distinguished robust thematic structures from algorithmic artifacts, enhancing confidence in substantive interpretations while maintaining appropriate epistemic humility about the constructed nature of computational text analysis (Anonymous, 2022; Ranganathan et al., 2018).

The multi-metric comparison presented in Table 2 provided critical methodological insight into how different topic modelling techniques performed when applied to the curated corpus of agricultural technology adoption literature from Northern West Bengal (Botero-Valencia et al., 2025). One of the most striking findings was the substantial variation in topic diversity across modelling approaches, which directly influenced the analytical resolution of adoption themes. NMF with TF-IDF input recorded the lowest diversity score (0.30) and ranked last, indicating significant overlap among extracted topics. This suggested that TF-IDF weighting, while useful for emphasizing rare but important terms, may have suppressed corpus-wide structural differentiation in this dataset, where core terms such as "farmers," "adoption," and district names appeared consistently across studies (Botero-Valencia et al., 2025). Consequently, this configuration produced broader but less distinct thematic clusters, limiting its usefulness for isolating specific adoption pathways such as mechanization, irrigation, or gender-based participation. In contrast, NMF using count-based input achieved the highest diversity score (0.875), demonstrating superior capacity to separate specialized adoption themes with minimal lexical overlap (Rejeb et al., 2025). This confirmed that raw count-based representations preserved the original distributional structure of the literature more effectively, allowing clearer identification of discrete domains such as integrated pest management, solar irrigation under PM-KUSUM, and mechanization through power tillers and custom hiring centres. LSA (SVD) and standard LDA occupied intermediate positions, with diversity scores of 0.7875 and 0.75, respectively, reflecting their ability to capture meaningful semantic differentiation while still maintaining conceptual cohesion across related topics (Li et al., 2024).

Equally important were the trade-offs between interpretability, computational efficiency, and probabilistic coherence revealed in the table, which had direct implications for methodological selection in agricultural literature synthesis. The probabilistic LDA models, particularly the optimized LDA variant, demonstrated very high interpretability and strong topic separation, highlighting their effectiveness in producing conceptually coherent themes aligned with known socio-economic and institutional determinants of technology adoption (Botero-Valencia et al., 2025). Although the optimized LDA exhibited the highest perplexity (482.21) and slowest processing speed due to extended iterations and asymmetric prior optimization, this reflected the model's deeper exploration of latent semantic structure rather than poorer performance. Its high diversity score (0.85) combined with very high interpretability, indicated that it was especially effective in distinguishing complex adoption drivers such as institutional support, climate adaptation strategies, and gender-mediated adoption pathways. Standard LDA offered a balanced compromise, with moderate computational cost, strong interpretability, and good diversity, making it suitable for generating policy-relevant thematic insights without excessive computational burden (Rejeb et al., 2025). Meanwhile, LSA demonstrated exceptional computational efficiency as the fastest method, but its medium interpretability suggested that dimensionality reduction techniques may have blended related but analytically distinct themes, reducing conceptual clarity (Li et al., 2024). Taken together, these results reinforced a key conclusion of the study: methodological choice significantly shaped the analytical lens through which agricultural technology adoption patterns were interpreted. High-diversity methods such as count-based NMF and optimized LDA were particularly valuable for identifying distinct adoption domains, while standard LDA provided a robust balance between interpretability and efficiency. This comparative evaluation strengthened the study's methodological rigor by demonstrating that the identification of core adoption themes such as institutional extension systems, gender inclusion, mechanization, and climate-adaptive technologies was not an artifact of a single modelling approach but emerged consistently across multiple computational frameworks (Botero-Valencia et al., 2025).

4. CONCLUSION

The analysis showed that technology adoption in Northern West Bengal was influenced by institutional networks, gendered access to resources, district-level variations, and climate pressures. Across all topic modelling approaches, recurring themes included the pivotal role of KVK, women's engagement through self-help groups, and localized adoption hotspots. By comparing diverse computational techniques, the study uncovered both consistent and model-specific patterns. These findings offered research scholars and policymakers a richer, evidence-based understanding to design inclusive and region-sensitive agricultural strategies.

5. ACKNOWLEDGEMENT

The authors gratefully acknowledge the academic support provided by the Uttar Banga Krishi Viswavidyalaya and Bidhan Chandra Krishi Viswavidyalaya during the progression of this research.

6. REFERENCE

Anonymous, 2021. Cereal Systems Initiative for South Asia (CSISA) annual progress report. International Maize and Wheat Improvement Center (CIMMYT). Available from: https://csisa.org/wp-content/uploads/sites/2/2021/07/210705-CSISA-PIII-Semiannual-Report-Oct20-to-Mar21_web.pdf. Accessed Date:06.08.2025.

Anonymous, 2022. Diffusion of agricultural innovations in the Eastern Gangetic Plains. International Maize and Wheat Improvement Center (CIMMYT). Available from: https://csisa.org/wp-content/uploads/sites/2/2023/01/CSISA-III-Annual-Report-Oct21-to-Sep22_web.pdf. Accessed Date: 15.09.2025.

Anonymous, 2023. The state of food and agriculture 2023: Revealing the true cost of food to transform agrifood systems. Food and Agriculture Organization of the United Nations. Available from: https://doi.org/10.4060/cc7724en. Accessed Date: 22.09.2025.

Anonymous, 2024. Technology adoption and climate-resilient practices in Northern India: Annual technical insights. International Maize and Wheat Improvement Center (CIMMYT). Available from: https://repository.cimmyt.org/handle/10883/24910. Accessed Date: 24.10.2025.

Anonymous, 2025. Data-driven policy insights for sustainable agriculture in South Asia. International Food Policy Research Institute. From https://repository.ifpri.org/bitstream/handle/20.500.12348/5626/IFPRI%20Annual%20Report%202022.pdf?sequence=1&isAllowed=y. Accessed Date: 16.09.2025.

Aryal, J.P., Farnworth, C.R., Khurana, R., Ray, S., Sapkota, T.B., Rahut, D.B., 2020. Does women's participation in agricultural technology adoption decisions affect the adoption of climate-smart agriculture? Insights from the Indo-Gangetic Plains of India. Review of Development Economics, 24(3), 973–990. Available from: https://doi.org/10.1111/rode.12670.

Aryal, J.P., Rahut, D.B., Maharjan, S., Erenstein, O., 2018. Factors affecting the adoption of multiple climate-smart agricultural practices in the Indo-Gangetic plains of India. Natural Resources Forum 42(3), 141–158. Available from: https://doi.org/10.1111/1477-8947.12152.

Aryal, J.P., Sapkota, T.B., Khurana, R., Khatri-Chhetri, A., Rahut, D.B., Jat, M.L., 2020. Climate change and agriculture in South Asia: Adaptation options in smallholder production systems. Environment, Development and Sustainability 22(6), 5045–5075. Available from: https://doi.org/10.1007/s10668-019-00414-4.

Bird, S., Klein, E., Loper, E., 2009. Natural language processing with Python. O'Reilly Media. Available from: https://www.nltk.org/book/.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research 3, 993–1022. Available from: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Botero-Valencia, J., Hernandez Hernandez, G.C., Franco, M., 2025. Machine learning in sustainable agriculture: Systematic review and research perspectives. Agriculture 15(4), 377. Available from: https://doi.org/10.3390/agriculture15040377.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J.L., Blei, D.M., 2009. Reading tea leaves: How humans interpret topic models. Advances in Neural Information Processing Systems 22, 288–296. Available from: https://proceedings.neurips.cc/paper_files/paper/2009/file/f92586a25bb3145facd5ab5310b9b6b1-Paper.pdf.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R., 1990. Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(6), 391–407. DOI: https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9.

Dissanayake, D., Jayasinghe, S., Perera, K., 2022. Application of topic modelling techniques in agricultural and environmental research: A systematic review. Agricultural Systems 195, 103278. From https://doi.org/10.1016/j.agsy.2021.103278.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., Lim, W.M., 2021. How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research 133, 285–296. Available from: https://doi.org/10.1016/j.jbusres.2021.04.070.

Egger, R., Yu, J., 2022. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. Frontiers in Sociology 7, Article 886498. https://doi.org/10.3389/fsoc.2022.886498.

Feder, G., Umali, D.L., 1993. The adoption of agricultural innovations: A review. Technological Forecasting and Social Change 43(3–4), 215–239. https://doi.org/10.1016/0040-1625(93)90053-A.

Feder, G., Just, R.E., Zilberman, D., 1985. Adoption of agricultural innovations in developing countries: A survey. Economic Development and Cultural Change 33(2), 255–298. https://doi.org/10.1086/451461.

Geda, A., Kuhl, R., 2025. Climate-smart technology adoption in hazard-prone areas: Impact on food security. Humanities and Social Sciences Communications 12, 670. https://doi.org/10.1057/s41599-025-04954-1.

Grootendorst, M., 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794. Available from: https://arxiv.org/abs/2203.05794.

Kumar, R., Singh, K.M., Meena, M.S., 2023. Mechanization, custom hiring centres and technology adoption among smallholders in eastern India. Indian Journal of Agricultural Economics 78(1), 1–14. Available from: https://www.isaeindia.org/pdf/Indian-Journal-of-Agricultural-Economics-Vol-78-No-1-Jan-Mar-2023.pdf.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401(6755), 788–791. https://doi.org/10.1038/44565.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of Psychology 140, 1–55. Available from: https://psycnet.apa.org/record/1933-01885-001.

Li, H., Yang, S., Yan, J., Gao, W., Cui, J., Chen, Y., 2024. From conventional to organic agriculture: influencing factors and reasons for tea farmers' adoption of organic farming in Pu'er City. Sustainability 16(22), 10035. DOI: https://doi.org/10.3390/su162210035.

Mitra, R., Roy, M.N., Das, S., Mukherjee, D., 2024. Groundwater and climate-smart agriculture: A reflection from West Bengal, India. Mitigation and Adaptation Strategies for Global Change 29(5), Article 45. DOI: https://doi.org/10.1007/s11027-024-10137-1.

Nan, F., Ding, R., Nallapati, R., Xiang, B., 2019. Topic modeling with Wasserstein autoencoders. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 6345–6381. DOI: https://doi.org/10.18653/v1/P19-1640.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Moher, D., 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, Article n71. Available From: https://doi.org/10.1136/bmj.n71.

Paudel, G.P., Ghimire, R., Rahut, D.B., 2020. Adoption of climate-smart agricultural practices in South Asia. Land Use Policy 95, 104614. DOI: https://doi.org/10.1016/j.landusepol.2020.104614.

Paudel, G.P., Dilli Bahadur, K.C., Rahut, D.B., Khanal, N.P., Justice, S.E., McDonald, A.J., 2020. Smallholder farmers' willingness to pay for scale-appropriate farm mechanization: Evidence from the mid-hills of Nepal. Technology in Society 59, 101196. DOI: https://doi.org/10.1016/j.techsoc.2019.101196.

Ranganathan, T., Singh, S., Pal, S., 2018. Regional disparities in agricultural technology adoption in eastern India. Agricultural Economics Research Review 31(1), 45–56. DOI: https://doi.org/10.5958/0974-0279.2018.00005.5.

Rejeb, A., Appolloni, A., Rejeb, K., 2025. The impact of machine learning applications in the agricultural supply chain: A topic modeling-based review. Discover Food 4(4), 419. DOI: https://doi.org/10.1007/s44187-025-00419-1.

Roder, M., Both, A., Hinneburg, A., 2015. Exploring the space of topic coherence measures. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 399–408. DOI: https://doi.org/10.1145/2684822.2685324.

Roy, M.N., Das, S., 2024. Climate-resilient and smart practices for smallholders in West Bengal. Environment, Development and Sustainability. DOI: https://doi.org/10.1007/s10668-023-03324-3.

Ruzzante, S., Labarta, R., Bilton, A., 2021. Adoption of agricultural technology in the developing world: A meta-analysis of the empirical literature. World Development 146, 105599. DOI: https://doi.org/10.1016/j.worlddev.2021.105599.

Saha, S., Bose, M., Hossain, S., 2025. Precision agriculture for improving crop yield predictions: A literature review. Frontiers in Agronomy 7, 1566201. DOI: https://doi.org/10.3389/fagro.2025.1566201.

Sapbamrer, R., Thammachai, A., 2021. Factors affecting women farmers' adoption of agricultural innovations: A systematic review. Sustainability 13(7), 3842. DOI: https://doi.org/10.3390/su13073842.

Schofield, A., Magnusson, M., Mimno, D., 2017. Pulling out the stops: Rethinking stopword removal for topic models. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers: 432–436. DOI: https://doi.org/10.18653/v1/E17-2069.

Sengupta, M., Mallick, S., Biswas, S., 2025. Identification of socio-management issues associated with technology adoption tendency among sericulture farmers of West Bengal. Asian Journal of Agricultural Extension, Economics & Sociology 2025 43(11), 131–138. Available from: https://ageconsearch.umn.edu/record/389120/files/Mallick43112025AJAEES146454.pdf.

Spiegel, A., Britz, W., Djanibekov, U., Finger, R., 2021. Risk, risk aversion, and agricultural technology adoption-A novel valuation method based on real options and inverse stochastic dominance. Q Open 1(2), qoab016. DOI: https://doi.org/10.1093/qopen/qoab016 https://academic.oup.com/qopen/article/1/2/qoab016/6325550.

Smith, R., 2007. An overview of the tesseract OCR engine. Proceedings of the Ninth International Conference on Document Analysis and Recognition, 629–633. DOI: https://doi.org/10.1109/ICDAR.2007.4376991.

Thakur, S., Basu, D., 2025. Harnessing advanced NLP for multidimensional CEO leadership analysis: A qualitative lexical network and literal approach. International Journal of Agriculture Extension and Social Development 8(6), 612–622. DOI: https://doi.org/10.33545/26180723.2025.v8.i6h.2086.

Zeweld, W., Huylenbroeck, V.G., Tesfay, G., Speelman, S., 2017. Smallholder farmers' behavioural intentions towards sustainable agricultural practices. Journal of Environmental Management 187, 71–82. DOI: https://doi.org/10.1016/j.jenvman.2016.11.014.