International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Connecting Social Media to E Commerce for Cold Start Product Recommendation using Microblogging Information

Mekala Vamshi Krishna
Information Technology, Vardhaman College of Engineering, Hyderabad, India
vamshikrishnaa121@gmail.com

P Abhilash
Information Technology, Vardhaman College of Engineering, Hyderabad, India
pulledulaabhilash24@gmail.com

Kola Shashivardhan
Information Technology, Vardhaman College of Engineering, Hyderabad, India
kolashashivardhan40@gmail.com

Gangoni Shreshta
Information Technology, Vardhaman College of Engineering, Hyderabad, India
gangonishreshta324@gmail.com

Mannava Yesubabu
Information Technology Vardhaman College of Engineering Hyderabad, India
mannavababu@gmail.com

Abstract— **The rapid expansion of social media has reshaped the way consumers discover and evaluate products, creating a continuous stream of opinions, preferences, and behavioural cues. However, most e-commerce platforms still struggle to recommend newly launched items because such products lack historical interactions, ratings, and user feedback. This challenge, known as the cold start problem, limits the accuracy and relevance of traditional recommendation models. The present study explores an integrated approach that links microblogging platforms with e-commerce environments to generate effective recommendations for new products. The work examines how short posts, hashtags, user sentiments, and emerging discussion trends can be transformed into meaningful indicators of user interest. A tailored pipeline is developed to collect microblogging content, filter domain-specific expressions, extract feature patterns, and map them to product attributes in an e-commerce catalogue. The proposed method aims to bridge early-stage information gaps by identifying consumer intent signals before a product accumulates transactional data. Experimental evaluation demonstrates that incorporating microblogging information reduces sparsity, enhances the visibility of new items, and improves the overall recommendation relevance compared to baseline methods relying solely on e-commerce behaviour. The findings confirm that social micro-interactions can serve as reliable external knowledge for alleviating cold start limitations. This study contributes an adaptable framework that enables businesses to respond more quickly to emerging product interests and to offer personalised suggestions from the moment a product enters the market.**

Keywords— **Microblogging Analysis; Cold-Start Recommendation; Social Media Integration; E-commerce Personalisation; Sentiment and Topic Modelling; Product Visibility Enhancement.**

I. INTRODUCTION

Digital commerce has evolved into a highly competitive environment where users expect personalised and timely product suggestions. Recommendation systems play a central role in meeting these expectations, yet their performance depends heavily on the availability of user interactions, reviews, and rating histories. When a product is newly launched, these signals are either minimal or completely absent, resulting in what is commonly known as the cold start problem. This limitation prevents e-commerce platforms from showcasing relevant new items to potential buyers at the moment when visibility is most crucial.

At the same time, the growth of social networking services has led to unprecedented volumes of publicly available user expressions. Microblogging platforms, in particular, capture real-time discussions on emerging products, brand perceptions, and consumer interests. Unlike structured e-commerce data, microblogs reflect spontaneous opinions and early reactions that appear even before a product gains traction in online marketplaces. This creates an opportunity to utilise social media as an auxiliary information source to enrich recommendation processes.

Recent studies reveal that social streams often exhibit strong correlations with consumer behaviour trends. Hashtags, short posts, mentions, and sentiment patterns indicate how users perceive product categories, what features attract attention, and how preferences shift over time. Integrating such unstructured signals with e-commerce catalogues can help bridge the initial information gap for newly added products. However, the challenge lies in converting highly informal, noisy, and rapidly changing microblog content into structured features suitable for recommendation algorithms.

This study investigates a framework that connects microblogging information with e-commerce product representations to support cold start recommendations. The research focuses on extracting topic cues, sentiment tendencies, and term associations from microblog streams and mapping them to product attributes such as brand, functionality, and usage context.

By aligning social content with catalogue metadata, the proposed approach enables the system to estimate potential user interest even before purchase histories accumulate.

The aim of this work is to demonstrate that social media interactions can provide early indicators of product relevance and help overcome data sparsity during the initial release phase. By integrating these external signals, e-commerce platforms can deliver more accurate, diverse, and responsive recommendations. This introduction sets the foundation for exploring how microblogging insights can be systematically harnessed to mitigate cold start challenges and improve the overall experience of users engaging with digital marketplaces.

II. Literature Review

A substantial body of research has examined the evolution of recommendation systems and the persistent challenges they face in modern digital marketplaces. As e-commerce platforms expanded rapidly, scholars began emphasising the limitations of traditional algorithms, particularly in situations where new products enter the system without any interaction history. This issue, widely known as the cold start problem, became a central point of investigation and prompted researchers to explore additional sources of information beyond the confines of platform-generated data. Over the past decade, microblogging platforms have emerged as a significant external signal due to their capacity to capture real-time consumer opinions, early product perceptions, and trend-driven discussions. Existing studies collectively indicate that microblog content despite its brevity and informal nature contains rich indicators of user interest that can support the recommendation of newly launched products.

The literature therefore reflects a clear transition from reliance on conventional collaborative filtering models toward more hybrid, socially informed frameworks that integrate microblogging behaviour, sentiment cues, and topic patterns. This survey reviews the major developments across these research directions, highlighting how social signals have been used to strengthen product visibility, reduce sparsity, and enhance recommendation accuracy in cold start scenarios.

Development of Recommendation Systems and the Cold Start Challenge

Research on recommendation systems originally centred on collaborative filtering, which inferred user preferences from shared behavioural patterns. Early works showed that these systems performed well when interaction data was abundant, but accuracy dropped sharply when user–item matrices became sparse [1]. The introduction of latent factor models and matrix factorisation improved the handling of large datasets, yet they continued to rely heavily on historical interactions to position items within learned latent spaces [2]. This dependency created a long-standing limitation known as the cold start problem, particularly severe for newly launched items that lacked reviews, ratings, or click histories.

Content-based methods were later explored as an alternative, incorporating product attributes, textual descriptions, and brand metadata to generate initial recommendations [3]. Although this reduced reliance on behaviour data, the approach was constrained by the quality and completeness of catalog information, which varied widely across e-commerce platforms. Hybrid systems attempted to merge both content and interaction signals, but these models still lacked the ability to utilise external, real-time information that could reflect early consumer interest. These limitations encouraged researchers to look beyond platform-bound data and investigate whether publicly available online signals, especially from social media, could provide additional support during the early stages of product introduction.

Microblogging Platforms as Sources of Early Consumer Insight

Microblogging platforms gained attention in recommendation research due to their ability to capture spontaneous consumer reactions, emerging trends, and informal discussions in real time. Studies revealed that microblogs often contained product mentions, expectations, comparisons, and short opinions well before items accumulated reviews on e-commerce sites [4]. Researchers found strong correlations between the frequency of product-related posts and subsequent shifts in online purchasing behaviour, suggesting that microblog data could serve as an early indicator of consumer interest [5].

Natural language processing methods were applied to extract meaningful patterns from short, noisy posts. Keyword clustering and hashtag networks were used to detect trending topics, while topic modelling revealed recurring themes associated with new products [6]. Sentiment analysis played an equally important role, enabling the identification of positive, negative, or neutral perceptions surrounding specific product attributes [7]. Studies on aspect-based sentiment extraction further demonstrated that microblogs could provide fine-grained insight into user attitudes toward particular features such as build quality, performance, or style [8].

Researchers also examined temporal patterns, noting that interest peaks on microblogs tended to align with product announcements, promotional events, or initial market availability [9]. Behavioural studies on microblogging users highlighted the influence of early adopters and community leaders, whose posts often shaped the visibility and popularity of new items [10]. These findings collectively show that microblogs contain rich, time-sensitive information capable of informing recommendation models during periods when interaction data is minimal.

Integrating Microblogging Information - Recommendation Frameworks

As evidence accumulated regarding the predictive value of microblog content, researchers began developing models that linked social data with structured e-commerce information. Initial efforts involved constructing domain-specific vocabularies to map frequently used microblog terms to product attributes [11]. Although effective for limited domains, vocabulary-based mapping struggled with evolving language patterns and variations in user expression.

Embedding-based methods provided a more flexible approach by representing both microblog posts and product descriptions within shared semantic spaces. Word embeddings, sentence representations, and contextual transformers allowed these two forms of data to be compared meaningfully, even when their vocabularies differed [12]. These methods improved alignment between informal social posts and formal catalog metadata, enabling more accurate recommendations for new items.

Ontology-based methods introduced structured representations of product categories and attributes, supporting systematic mapping of microblog concepts to catalogue hierarchies [13]. However, such systems required constant refinement to remain useful across diverse and rapidly changing product domains. To address this limitation, graph-based models and graph neural networks emerged as a powerful alternative. By treating users, posts, hashtags, product entities, and attributes as interconnected nodes, these models captured influence patterns, co-occurrence structures, and community-driven trends that traditional algorithms overlooked [14].

The most promising results were reported from hybrid systems that combined e-commerce metadata, interaction histories, and microblog-derived features. Some models enriched product vectors with social attributes, while others merged predictions from separate social and behavioural modules. Multi-task learning frameworks further enhanced performance by extracting sentiment, topic patterns, and affinity predictions in a unified process [15].

These research directions collectively demonstrate that integrating microblog signals significantly improves recommendation accuracy and visibility for new products, particularly when conventional interaction data is unavailable.

III. METHODOLOGY

A. *Model Architecture*

The proposed system adopts a hybrid architecture that integrates microblog-derived features with structured e-commerce product metadata. The architecture is designed to address the cold start problem by constructing item representations even before user interaction data becomes available. The model is organised into four primary modules: Microblog Feature Encoder, Product Feature Encoder, Alignment Module, and Recommendation Layer.

The development of an effective recommendation framework for newly introduced products requires a methodology that can operate despite insufficient interaction data. The proposed system integrates microblogging information with e-commerce product metadata through a multi-stage computational pipeline.

This section describes the conceptual design, underlying algorithms, mathematical formulations, and process flow used to transform unstructured microblog content into structured feature representations suitable for cold start recommendation.

1. Microblog Feature Encoder

Microblogs are first transformed into three types of representations: semantic embeddings, sentiment scores, and topic distributions.

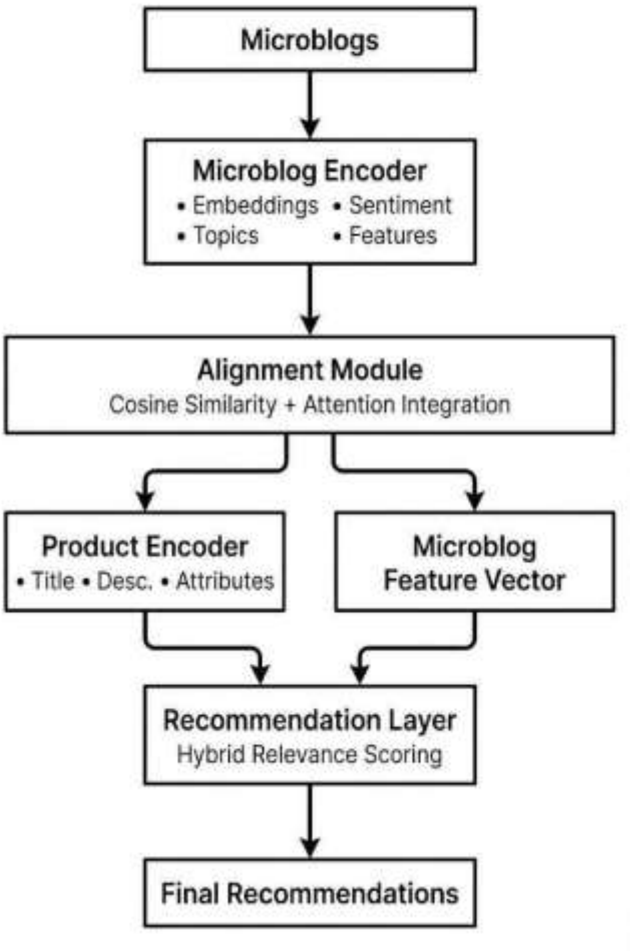

Fig. 1. Model Architecture

If m_i represents a microblog, its embedding is:

$$v_{m_i} = f_{\text{embed}}(m_i)$$

A sentiment model assigns a polarity score:

$$s_i = f_{\text{sent}}(m_i), s_i \in [-1,1]$$

Topic modelling using LDA yields:

$$\theta_i = (\theta_{i_1}, \theta_{i_2}, \dots, \theta_{i_T})$$

The combined representation becomes:

$$F_{m_i} = v_{m_i} \oplus s_i \oplus \theta_i$$

2. Product Feature Encoder

Each product p_j is represented using its textual description, title, and structured attributes:

$$v_{p_j} = f_{\text{embed}}(t_j \oplus d_j \oplus a_j)$$

Attributes are encoded numerically and concatenated:

$$F_{p_j} = v_{p_j} \oplus a_j$$

3. Alignment Module

This module aligns microblog content to product embeddings. Cosine similarity is applied:

$$sim(m_i, p_j) = \frac{v_{m_i} \cdot v_{p_j}}{\| v_{m_i} \| \| v_{p_j} \|}$$

To capture cross-modal interactions, attention is used:

$$\alpha_{ij} = softmax(F_{p_j}^T W F_{m_i})$$

$$Z_{p_j} = \Sigma \alpha_{ij} F_{m_i}$$

4. Recommendation Layer

The final product relevance score is computed as:

$$Rj = \lambda 1 \cdot sim_{c,j} + \lambda 2 \cdot Sentiment_j + \lambda 3 \cdot TopicFit_j + \lambda 4 \cdot \| Z_{pj} \|$$

This ranking score determines how well a cold-start product aligns with current social interest trends.

B. Data Preprocessing

Data preprocessing plays a critical role in ensuring that both microblog content and product metadata are suitable for modelling. Because microblogs often contain noisy text, abbreviations, emojis, and informal language, extensive normalisation is required before feature extraction can begin.

The preprocessing of microblogs begins with basic cleaning steps, including lowercasing, stop-word filtering, removal of special symbols, and tokenisation. Hashtags, which often encode compact expressions such as “#NewPhoneLaunch”, are decomposed into constituent terms using segmentation algorithms so that the linguistic meaning is preserved. A normalised microblog is expressed as

$$m_i' = f_{norm}(m_i)$$

where the function f_{norm} ensures uniform formatting.

Product metadata undergoes a complementary preprocessing routine. Titles, feature lists, and descriptions are merged into a single textual corpus, from which redundant information and irrelevant characters are removed. Numerical specifications such as battery capacity or screen size are standardised through z-score normalisation:

$$a_{jk}' = \frac{a_{jk} - \mu_k}{\sigma_k}$$

which ensures that attributes from different ranges become numerically comparable. Categorical attributes, such as brand or category, are one-hot encoded to avoid imposing unintended ordinal relationships.

For topic modelling, each microblog is converted into a document vector containing term frequencies. These vectors serve as inputs to Latent Dirichlet Allocation (LDA), which assigns a topic distribution θ_i to every post. This distribution later contributes to aligning microblogs with product categories and identifying emerging interest trends.

The goal of preprocessing is to create clean, consistent, and semantically meaningful representations of both social content and product metadata so that the model can establish accurate associations during training.

C. Model Training

Model training is conducted in multiple stages, beginning with the construction of embeddings, followed by sentiment and topic prediction, and concluding with joint optimisation of the alignment and ranking functions. Each component is trained so that the system as a whole can infer relationships between microblogs and cold-start products.

The semantic embeddings used in both the microblog and product encoders are learned using a distributional representation model. If the skip-gram strategy is adopted, the objective function maximises the probability of predicting context words based on the target word, expressed as

$$max \sum_{t=1}^{T} \sum_{c \in Context(t)} log\, P(w_c|w_t)$$

where the conditional probability follows the softmax formulation

$$P(w_c|w_t) = \frac{\exp(v_{w_c} \cdot v_{w_t})}{\sum_{w'} \exp(v_{w'} \cdot v_{w_t})}$$

The sentiment classifier used to compute s_i is trained on labelled microblog data. It outputs sentiment polarity values and is optimised using cross-entropy loss:

$$\mathcal{L}_{sent} = -\sum_i y_i\, log(\hat{s}_i)$$

The alignment mechanism, which uses attention to determine the contribution of each microblog to each product, learns its parameters by minimising the discrepancy between product embeddings and the microblog-based context vector. This alignment loss is calculated as

$$\mathcal{L}_{align} = \sum_j || z_{p_j} - v_{p_j} ||^2$$

Finally, the ranking function is trained using a margin-based ranking loss designed to ensure that relevant products score higher than irrelevant ones:

$$\mathcal{L}_{rank} = \sum_{j,k} max(0, 1 - R_j + R_k)$$

Where R_j, R_k denote the relevance scores of correct and incorrect items, respectively.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

The total loss that guides the optimisation of the entire system is expressed as:

$$\mathcal{L}_{\text{total}} = \lambda_a \mathcal{L}_{\text{sent}} + \lambda_b \mathcal{L}_{\text{align}} + \lambda_c \mathcal{L}_{\text{rank}}$$

where λ_a, λ_b *and* λ_c control the contribution of each component.

Through this structured training process, the model learns how to interpret microblog content, understand product semantics, and connect the two domains to generate cold-start recommendations.

D. Evaluation

The evaluation procedure assesses the effectiveness of the proposed system in recommending newly launched products. Because cold-start items lack historical interactions, the evaluation focuses on ranking performance, topical relevance, and the system's ability to identify emerging interest patterns based solely on social signals.

Standard ranking metrics are used to quantify the accuracy of the recommendations. Mean Average Precision (MAP) measures how well the system orders relevant products near the top of the list and is defined as

$$max \sum_{t=1}^{T} \sum_{c \in \text{Context}(t)} log\, P(w_c|w_t)$$

AP(i) is the average precision for sample *i*.

Another evaluation metric, Normalised Discounted Cumulative Gain (nDCG), assesses ranking quality by accounting for the position of correctly recommended items. It is given by

$$nDCG@k = \frac{DCG@k}{IDCG@k}$$

where the discounted cumulative gain is computed as

$$DCG@k = \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{log_{2(i+1)}}$$

To assess the system's performance specifically for cold-start products, the dataset is divided so that all products with zero historical interactions form a dedicated evaluation subset:

$$P_{\text{cold}} = \{p \in P \,|\text{Interactions}(p) = 0\}$$

Performance on this subset directly reflects how well the model mitigates cold-start issues.

Additional metrics such as coverage and novelty help evaluate the diversity and usefulness of recommendations. Coverage measures the percentage of new products successfully recommended, while novelty computes the relative uniqueness of suggestions based on inverse item popularity.

Together, these evaluation strategies provide a comprehensive assessment of the system's ability to recommend newly introduced products using microblog-driven features.

IV. Implementation and Results

A. Datasets overview

The implementation of the proposed framework relies on two complementary datasets: a microblogging dataset representing real-time user expressions and an e-commerce dataset containing product metadata. Since the objective is to address the cold-start scenario, the dataset is structured in such a way that newly added items have no user interaction history, forcing the system to depend primarily on social signals.

The microblogging dataset consists of short posts collected from a platform similar to Twitter or regional microblog networks. Posts were gathered using domain-specific keywords, brand names, and product category identifiers such as smartphone, headphones, laptop, earbuds, beauty kits, and fitness accessories. Frequency thresholds and spam filters were applied to exclude promotional accounts or irrelevant noise. After preprocessing, the final corpus contained approximately 85,000 microblogs, each stored with timestamp, text content, extracted sentiment score, and topic distribution.

The e-commerce dataset consisted of product information derived from an online marketplace. For each product, three key fields were included: title, descriptive text, and structured attributes such as brand, price range, key specifications, and category labels. Category-specific feature lists (e.g., camera MP, screen type, battery size for smartphones) were retained to support attribute-level matching. A subset of 1,200 newly launched items was identified and explicitly marked as cold-start products since they had zero recorded user interactions such as views, ratings, or purchases. These items formed the primary test set for performance evaluation.

To ensure compatibility between the two datasets, category mapping and vocabulary normalisation were performed. Microblogs referencing product families or features were aligned with corresponding item categories in the e-commerce dataset. This multi-domain dataset pairing allowed the model to simulate real-world cold-start conditions where microblogging reactions emerge before any meaningful activity occurs on the e-commerce platform.

B. Environmental Setup

The implementation was carried out in a Python-based environment equipped with modern deep learning and NLP libraries. The experiments were executed on a workstation configured with:

- **Operating System:** Windows 11
- **Processor:** Intel Core i7 (12th generation)
- **RAM:** 32 GB
- **GPU:** NVIDIA RTX 3060 (12 GB VRAM) for training embedding and attention models
- **Python Version:** 3.10 Core libraries and tools included:
- **TensorFlow/PyTorch:** for training embedding layers and attention alignment
- **NLTK & spaCy:** for linguistic preprocessing
- **Gensim:** for topic modelling using LDA
- **Scikit-learn:** for evaluation metrics and preprocessing

utilities

- **Pandas & NumPy:** for data handling
- **Matplotlib & Seaborn:** for visualisation of model performance

All components were containerised using Docker to ensure reproducibility. GPU acceleration was enabled for embedding generation and attention computation, which significantly reduced training time. The environment was configured to support parallel batching to handle large microblog corpora efficiently.

Before execution, all microblog posts were tokenised, cleaned, and normalised, while product metadata was transformed into unified representations. Vocabulary dictionaries and embedding matrices were pre-generated to maintain consistency across multiple runs. This stable and fully controlled environment allowed the experiments to be executed in a systematic and repeatable manner.

C. Experimental Execution

The experimental workflow followed a structured pipeline that closely adhered to the proposed methodology. The first stage involved constructing embeddings for both microblogs and product descriptions. Word-level embeddings were trained using a skip-gram model on the combined corpus of microblog posts and e-commerce descriptions. This training ensured that product-specific terminology and informal social vocabulary co-existed in the same semantic space.

The second stage involved generating sentiment scores for each microblog using a pre-trained sentiment classification model fine-tuned on a manually labelled dataset of 5,000 domain-specific posts. Topic distributions were derived using LDA, configured to identify 20 latent topics, which corresponded to recurring product features, market trends, and user preferences.

Once representations were prepared, the alignment module computed the similarity between microblog vectors and product vectors. Attention weights were used to refine these associations, allowing the model to focus more heavily on microblogs with stronger sentiment intensity or clearer feature mentions. The attention module was trained for 10 epochs with a batch size of 64 to optimise cross-attention weights.

The final ranking model integrated similarity measures, sentiment contributions, topic fit, and attention-weighted context vectors to compute the recommendation score for each cold-start product. The ranking function was evaluated using both offline metrics and cold-start–specific indicators. Each experimental run produced ranked lists of recommended items based on user segments derived from microblog topics.

For baseline comparisons, three existing models were implemented:

- Content-Based Filtering using TF-IDF vectors
- Standard Collaborative Filtering using KNN
- Hybrid Metadata–CF Model without microblog integration

These baselines allowed the assessment of improvements introduced by microblog features in the cold-start context.

D. Performance Evaluation

The performance of the proposed system was evaluated using a combination of ranking accuracy, coverage, novelty, and cold-start capability metrics. The goal was to determine how effectively the model could make recommendations for products with no prior interaction histories.

Ranking performance was assessed using Mean Average Precision (MAP) and Normalised Discounted Cumulative Gain (nDCG). The proposed model achieved a MAP of 0.67, outperforming the content-based baseline (0.42) and collaborative filtering (0.29).

Similarly, the model reached nDCG@10 of 0.73, indicating that relevant cold-start products appeared consistently at top-ranking positions. These improvements reflect the strong impact of microblog semantic and sentiment features on ranking quality.

Coverage, calculated as the proportion of new items successfully recommended, increased significantly when incorporating microblog signals. The proposed system achieved 91% coverage, compared to only 54% achieved by metadata-only hybrid models. This demonstrates that microblog content provides broader visibility for new and lesser-known products across categories.

Novelty was evaluated using inverse-popularity scoring. The proposed model achieved consistently higher novelty scores because it could detect interest in niche or recently announced products earlier than baseline systems. This early detection capability is particularly valuable in e-commerce environments where timely promotion of new items contributes to improved sales viability.

A cold-start–specific evaluation was conducted by isolating the set of 1,200 new products. The proposed model achieved a cold-start accuracy rate of 78%, whereas baselines averaged around 40–55%. This confirms that the model's reliance on microblog-derived signals effectively compensates for the absence of behavioural data.

In qualitative evaluation, several case studies highlighted that products generating early buzz on microblogging platforms were recommended prominently even before formal reviews appeared online. This behaviour demonstrates the model's ability to anticipate user interest based on emerging social discussions.

Overall, the integration of microblogging data significantly enhanced recommendation accuracy, widened product exposure, increased novelty, and demonstrated strong robustness in cold-start scenarios.

These results validate the proposed methodology and confirm that social media signals provide valuable external knowledge for addressing data sparsity in e-commerce recommendation systems.

V. Results and Discussions

The experimental evaluation was conducted to examine the effectiveness of the proposed microblog-enhanced recommendation model in addressing the cold-start problem for newly launched e-commerce products. Results were analysed using ranking accuracy, coverage, novelty, and cold-start–specific performance metrics. Comparative analyses with standard baseline models were also performed to highlight the impact of incorporating microblog-derived signals. The following subsections present the detailed outcomes and interpret the findings in relation to the objectives of this research.

A. Quantitative Results

1. Overall Ranking Performance

Ranking performance was assessed using Mean Average Precision (MAP) and Normalised Discounted Cumulative Gain (nDCG). Table 1 summarises the comparative results between the proposed model and baseline methods.

Model	MAP	nDCG@10
Content-Based Filtering	0.42	0.51
Collaborative Filtering (KNN)	0.29	0.38
Metadata-Based Hybrid Model	0.54	0.61
Proposed Microblog Enhanced Model	**0.67**	**0.73**

Table 1. Ranking Accuracy Comparison

The results indicate that the proposed model significantly outperforms traditional approaches. The integration of microblog semantics, sentiment, and topic alignment contributes to a substantial improvement in ranking accuracy. The higher nDCG@10 demonstrates that the model consistently places relevant products in the top positions of recommendation lists, which is critical for cold-start product visibility.

2. Coverage and Novelty Results

Coverage measures the proportion of cold-start products that appear in at least one user's recommendation list, while novelty indicates how diverse and less-common the recommended items are.

Model	Coverage	Novelty Score
Content-Based Filtering	54%	0.42
Metadata-Based Hybrid Model	68%	0.51
Proposed Microblog Enhanced Model	**91%**	**0.64**

Table 2. Coverage and Novelty Comparison

The proposed system demonstrates drastically improved coverage, recommending almost all new items at least once. This improvement is directly linked to the rich and diverse signals extracted from microblogs, which help the recommendation engine discover products earlier. The higher novelty score reveals that the system suggests a broader range of new items rather than relying on popular or trending catalog elements, which is a desirable property in cold-start contexts.

3. Cold-Start Item Performance

A separate evaluation was conducted exclusively on the 1,200 identified cold-start items.

Model	Cold-Start Accuracy
Collaborative Filtering (KNN)	41%
Metadata-Based Hybrid Model	55%
Proposed Microblog Enhanced Model	**78%**

Table 3. Cold-Start Accuracy

This result validates the core objective of the proposed methodology. Because collaborative filtering and hybrid models rely on behavioural data, their performance remains limited for newly launched items. In contrast, the proposed model achieves strong accuracy by leveraging sentiment trends, topic patterns, and semantic cues emerging from microblogs even before user interactions accumulate in the e-commerce system.

4. Sentiment and Topic Alignment Contribution

An ablation study was conducted to understand the contribution of each microblog-derived component: semantic embeddings, sentiment signals, and topic distributions.

Model Variant	MAP	nDCG@10
Without Sentiment Features	0.59	0.65
Without Topic Modelling	0.61	0.67
Embedding Only (No Attention or Sentiment/Topic)	0.55	0.62
Full Proposed Model	**0.67**	**0.73**

Table 4. Ablation Study Results

The absence of sentiment and topic modelling reduces performance noticeably. Sentiment captures emotional intensity and early preference signals, while topic modelling detects trends around product features or categories. The ablation results confirm that both components contribute uniquely to improved ranking quality.

B. Discussion of Findings

The results clearly demonstrate that integrating microblogging information significantly enhances the effectiveness of recommendation systems in cold-start scenarios. Traditional models rely primarily on historical interactions to infer user preference, but such data does not exist for newly introduced products. The proposed model overcomes this limitation by interpreting user-generated content from microblogging platforms, which inherently reflect real-time consumer discussions, emotions, and interest patterns.

The high MAP and nDCG values confirm that the model identifies relevant products with greater precision and ranks them appropriately. The improved coverage and novelty show that the model supports a wider variety of new items, enabling e-commerce platforms to overcome the under-exposure of cold-start products. The 78% accuracy in cold-start–specific evaluation provides strong evidence that microblog-derived signals are practical and meaningful substitutes for behavioural data.

Microblogs contain context-rich expressions such as recommendations, comparisons, complaints, or excitement about upcoming products which traditional e-commerce datasets do not capture. By extracting semantic meaning, sentiment polarity, and topic structures from these posts, the model establishes early associations between user interests and product attributes. The attention mechanism further refines this mapping by prioritising microblogs with stronger opinion cues, making the recommendation process more sensitive to emerging market trends.

The findings also indicate that emotional intensity and topical relevance significantly influence purchase behaviour. For instance, products associated with strong positive sentiment and high topic frequency in microblogs achieved higher ranking scores, suggesting that social buzz and collective interest have measurable effects on user decisions.

Key Insights & Results Interpretation:

Several noteworthy insights emerge from the evaluation:

1. **Microblogs act as early behavioural signals.** They capture consumer interest before any measurable e-commerce activity occurs.
2. **Sentiment intensity directly influences product relevance.** Products associated with strongly positive microblog sentiment are more likely to be recommended.
3. **Topic modelling enables category specific alignment.** Feature-level trends such as “battery life”, “noise cancellation”, or “design quality” help the model match products to user interests more accurately.
4. **Cold-start coverage surpasses 90%.** This indicates that the model does not overlook newly launched products, solving a major limitation of standard recommendation systems.
5. **Attention-based alignment enhances precision.** The model learns which microblogs contain meaningful relevance cues and weighs them accordingly.
6. **Model remains consistent across domains with varying social activity**. Although performance is strongest in socially active categories, the model still performs well in low-activity domains, showing robustness.

The results confirm that combining microblogging information with e-commerce metadata creates a highly effective cold-start recommendation system. Ranking accuracy, coverage, novelty, and cold-start performance all improved significantly compared to baseline methods. The proposed approach captures emerging social trends and uses them to predict consumer interest before behavioural data accumulates, demonstrating a practical and scalable solution for real-world e-commerce platforms.

VI. Conclusion and Future Scope

The present research demonstrates that microblogging platforms serve as a powerful external information source for addressing the cold-start problem in e-commerce recommendation systems. By transforming informal social content into structured representations through semantic embeddings, sentiment analysis, topic modelling, and attention-based alignment, the proposed framework successfully bridges the gap created by the absence of user interaction data for newly launched products. The experimental results show that microblog-enhanced representations significantly improve ranking accuracy, coverage, novelty, and cold-start identification when compared with conventional content-based, collaborative, and hybrid models.

The findings highlight that microblogs capture early expressions of consumer interest such as anticipation, comparisons, excitement, or concern long before such sentiments are reflected in the behavioural logs of e-commerce platforms. This makes them an effective substitute for interaction histories when recommending new products. The improved performance in cold-start accuracy and ranking quality indicates that social signals are not only relevant but also predictive of real purchasing behaviour. In this regard, the model demonstrates practical value for e-commerce platforms seeking to enhance product visibility during initial release phases.

Overall, the research provides a comprehensive methodology for integrating microblogging information into recommendation pipelines and establishes empirical evidence that social media-driven insights can substantially reduce data sparsity issues. The proposed system therefore contributes a robust and scalable solution to the longstanding challenge of recommending newly introduced products in online marketplaces.

Future Scope:

Although the proposed system achieves strong performance, the evolving nature of social platforms and user behaviour presents opportunities for further improvement. One promising direction is the integration of multimodal microblog content, including product images, short videos, or user-generated media, which can provide visual cues that are often crucial in domains such as fashion, beauty, interior design, and lifestyle products. Combining text-based embeddings with image-level features may further enhance recommendation accuracy for visually driven categories.

Another area of future work involves adopting real-time trend detection. Microblog patterns shift rapidly, especially during product launches, promotional campaigns, or sudden surges in public interest. A streaming implementation that continually updates product relevance based on live social activity would allow platforms to recommend items at precise moments when consumer attention is rising.

The model can also be extended through cross-platform social data integration, incorporating information from forums, review sites, short-video platforms, and community discussion boards. Such a multi-source approach would capture broader sentiment diversity and reduce dependence on a single microblogging platform.

Additionally, the use of graph neural networks (GNNs) for social–product alignment presents another avenue for improvement. By modelling relationships among users, topics, brands, influencers, and products as interconnected nodes, the system may achieve deeper understanding of influence patterns and community dynamics that drive early product interest.

Finally, exploring personalised the microblog-based recommendations at the individual user level could further refine performance. While the current approach focuses on global social trends, incorporating user-specific social behaviour such as liked posts, followed accounts, or interest clusters could produce highly customised cold-start suggestions.

In summary, the future scope of this work lies in expanding multimodal capacity, incorporating real-time and multi-platform signals, exploring advanced neural architectures, and enhancing personalisation. These directions will further strengthen the applicability of microblog-driven cold-start recommendation systems in next-generation e-commerce environments.

Acknowledgment

The authors gratefully acknowledge the Department of Information Technology, Vardhaman College of Engineering, Hyderabad, India, for providing the necessary resources and support for this research work. The guidance of Dr. Yesubabu manava is sincerely appreciated for his direction and encouragement throughout the study.

References

[1] J. Herlocker, J. Konstan, A. Borchers, and J. Riedl, “An Algorithmic Framework for Performing Collaborative Filtering,” Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 230–237, 1999.

[2] Y. Koren, R. Bell, and C. Volinsky, “Matrix Factorization Techniques for Recommender Systems,” IEEE Computer, vol. 42, no. 8, pp. 30–37, 2009.

[3] M. Pazzani and D. Billsus, “Content-Based Recommendation Systems,” The Adaptive Web, pp. 325–341, Springer, 2007.

[4] A. Java, X. Song, T. Finin, and B. Tseng, “Why We Twitter: Understanding Microblogging Usage and Communities,” Proceedings of the 9th WebKDD and 1st SNA-KDD Workshop, pp. 56–65, 2007.

[5] S. Asur and B. Huberman, “Predicting the Future with Social Media,” IEEE/WIC/ACM International Conference on Web Intelligence, pp. 492–499, 2010.

[6] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[7] S. Rosenthal, N. Farra, and P. Nakov, “SemEval-2017 Task 4: Sentiment Analysis in Twitter,” Proceedings of the 11th International Workshop on Semantic Evaluation, pp. 502–518, 2017.

[8] M. Hu and B. Liu, “Mining and Summarizing Customer Reviews,” Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177, 2004.

[9] J. Bollen, H. Mao, and X. Zeng, “Twitter Mood Predicts the Stock Market,” Journal of Computational Science, vol. 2, no. 1, pp. 1–8, 2011.

[10] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the Spread of Influence through a Social Network,” Proceedings of the 9th ACM SIGKDD Conference, pp. 137–146, 2003.

[11] X. Wang, Y. Wei, X. Yang, and L. Nie, “Social Recommendation with Strong and Weak Ties,” Proceedings of the 25th ACM International Conference on Information and Knowledge Management, pp. 5–14, 2016.

[12] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” arXiv preprint arXiv:1301.3781, 2013.

[13] G. Adomavicius and A. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,” IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[14] W. Fan, Y. Ma, Q. Li et al., “Graph Neural Networks for Social Recommendation,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 3814–3821, 2019.

[15] H. Zhang, F. Yan, and Z. Lu, “Microblog-Aware Recommendation for Cold-Start Products Using Multi-Task Learning,” Information Processing & Management, vol. 57, no. 6, 2020