CHAPTER 1

INDTRODUCTION

This chapter includes the following subtopics, namely: (1) Rationale; (2) Theoretical Framework; (3) Conceptual Framework/Paradigm; (4) Statement of the problem; (5) Hypothesis (Optional); (6) Assumption (Optional); (7) Scope and Delimitation; and (8) Importance of the study.

1.1 Rationale

Online news distribution platforms such as Microsoft News and Google News have become primary channels for efficiently accessing daily news [50]. In today's digital era, timely and relevant information is of paramount importance. Online news platforms offer fast and flexible access to a wide array of content—an advantage that is difficult to match by traditional media [42], [14]. As a result, hundreds of millions of active users rely on these platforms every day. However, the overwhelming and continuously growing volume of information makes it increasingly difficult for users to find news that aligns with their interests [42], [14], [49]. This has led to the problem of information overload, highlighting the critical need for effective news recommendation systems. Such systems not only filter excessive information in a general sense, as recommendation systems in other domains do [6], but are also specifically designed to present users with news that is relevant and engaging [50].

Recommendation systems have become an integral component of various major technology platforms such as YouTube, Netflix, Spotify, X (formerly Twitter), and Amazon. Similar systems have also been deployed across numerous other domains, including online learning, entertainment, tourism, news, e-commerce, social media [38], [7], and film [4]. In the competitive landscape of delivering high-quality recommendations, personalization plays a vital role in enhancing user experience and engagement [32], [22], [8]. By analyzing preference patterns from behavioral history, these systems can identify and select relevant news articles from a vast pool of candidates, thereby helping to mitigate the problem of information overload [42], [46].

A range of traditional recommendation algorithms—content-based methods (such as TF-IDF, CNN, DSSM), collaborative filtering techniques (such as SVD, FM), and hybrid approaches (such as DKN, LSTUR)—have been widely employed [14]. However, these methods often fall short in handling complex and dynamic data [6], especially considering that the news domain is characterized by unique user behavior and rapidly evolving content [42]. Additionally, the cold-start problem—exacerbated by the extremely short news lifecycle—remains a significant challenge [49], [41]. In general text classification tasks, conventional machine learning models (such as TF-IDF, SVM, RNN) struggle to capture deep semantic relationships and manage linguistic diversity—two critical aspects for accurately understanding and recommending content [27], [10].

Since news content is generally text-based (such as titles, abstracts, and main bodies), deep learning encoders like LSTM and Transformer architectures have proven promising in processing these features. Recently, deep learning-based pre-trained language models (PLMs) have demonstrated superior performance in various natural language processing (NLP) tasks due to their capability to capture sentence structures effectively [50]. This capability has also been applied in news recommendation systems, where PLMs are used to encode textual content and utilize the [CLS] token as a representation of news articles, thereby enhancing recommendation accuracy [46]. Among the various PLMs, BERT (Bidirectional Encoder Representations from Transformers) stands out for its deeper semantic understanding and its ability to alleviate the cold-start problem through pre-trained embeddings and parameters [49], [44], [11].

Qi Zhang, et al., [49] proposed UNBERT, which leverages BERT's pre-training techniques for news recommendation. In this approach, news headlines are represented as semantic embeddings and matched with user history through Transformer layers to generate relevant recommendations. Meanwhile, **Hongyan Xu, et al.,** [47] introduced GLSM, which explores news headlines to construct long-term, short-term, and group-based representations using BERT for generating personalized recommendations. **Z. Zhang and B. Wang,** [50] combined prompt learning with BERT in Prompt4NR (Prompt Learning for News Recommendation). This method exploits the semantic and linguistic knowledge embedded in PLMs such as BERT to capture relationships between historical headlines and candidate news, thereby improving recommendation performance by exploring latent connections between news content and user interests. However, relying solely on headlines as the primary representation may not fully capture the overall context of a news article, potentially affecting the quality of recommendations in terms of deeper content comprehension and nuance.

News articles contain a variety of textual features—titles, abstracts, full content, categories, and named entities—all of which can enrich contextual understanding. Integrating these features can enhance news modeling in recommendation systems by providing more comprehensive representations [44]. Adding abstracts and full content to headlines can further improve prediction accuracy by offering deeper semantic understanding [34] and can also enhance users' trust in the system [31]. **Budi Juarto and Abba Suganda Girsang**, [15] combined title and content embeddings using Sentence-BERT, along with user and item IDs, which proved effective in increasing recommendation accuracy. **Niran A. Abdulhussein and Ahmed J. Obaid,** [1] used the MIND dataset alongside GloVe embeddings and multi-head attention to process titles and abstracts, resulting in improved relevance. **Shitao Xiao, et al.,** [45] explored pre-trained models such as BERT and RoBERTa to encode textual features of news articles (titles and content) and model user interaction, achieving highly accurate recommendations. While these approaches have enhanced semantic performance, most existing studies still overlook the temporal dimension, which is crucial in news recommendation, considering that news content has a limited window of relevance.

In terms of characteristics, news articles differ significantly from those in other domains. News content typically has a short lifecycle; its relevance can decline drastically within minutes, hours, or days after publication—unlike other domains (such as movies or e-commerce products), where relevance tends to persist over longer periods [14], [31], [48]. Therefore, integrating temporal features is essential for recommendation systems to adapt to the dynamic nature of news. **Starke, et al.,** [34] evaluated the contribution of various features in assessing similarity between news articles and found that temporal attributes—such as publication date—do exhibit some relevance, although their correlation with user-perceived similarity is relatively weak. In the aerospace industry, this feature is employed to ensure compliance with standard requirements [30]. In academic collaboration prediction, publication dates are used to identify emerging research trends over time [18]. However, the utilization of temporal features in news recommendation systems remains suboptimal, with most existing approaches failing to effectively capture the dynamics of novelty. As a result, many systems are unable to rapidly adapt recommendations in response to the latest changes and developments.

Parfenova and Clausel, [28] introduced a Time Decay Layer (TD) to assign greater weight to more recent posts within the EmoLSTMTd model, which integrates BERT and EmoBERTa with LSTM to predict pathological gambling risks based on social media activity. Conversely, **Seol, et al.,** [33] employed Unix timestamps as numerical representations of time, enabling systems to recognize and process temporal information. They also leveraged Temporal Self-Attention (TSA) to capture both short- and long-term user preferences. **Wang, et al.,** [40] proposed a conversational recommendation system that integrates a time-aware LSTM (EmoLSTMTd) with BERT and EmoBERTa. This approach incorporates time-aware attention mechanisms to emphasize newer items and processes entity-level information through R-GCN. By combining time decay mechanisms with explicit time representations such as Unix timestamps, these approaches demonstrate effectiveness in capturing temporal dimensions and prioritizing newer information, thereby enhancing the system's sensitivity to novelty dynamics in the recommendation process.

Several studies—such as [50], [49], [46], [47] and [15]—have integrated Transformer-based models like BERT to develop news recommendation systems. These methods leverage the powerful language representations learned by pre-trained models to capture contextual and textual features crucial for recommendation. BERT has shown strong performance across a range of NLP tasks [2]. In scenarios involving complex and large-scale datasets, models with higher capacity—such as BERT-Large—have the potential to provide deeper semantic understanding than BERT-Base [19]. **Ni and Kao**, [26] proposed Masked Siamese Prompt Tuning (MSP-Tuning) to enhance few-shot learning in natural language understanding. Evaluated on GLUE and SuperGLUE benchmarks (including SST-2, CoLA, RTE, and MRPC), their results demonstrated that BERT-Large significantly outperformed BERT-Base, particularly in tasks requiring deep semantic comprehension. **Eberhard, et al.,** [9] developed a text-based movie recommendation system by extracting entities from user queries on Reddit and found that BERT-Large outperformed BERT-Base in both entity recognition and contextual understanding.

In recommendation systems, BERT is not the only pre-trained model utilized; RoBERTa has also been adopted in numerous studies. RoBERTa-based models are employed to generate textual representations of news articles [51]. RoBERTa-Base contains fewer parameters than RoBERTa-Large, which affects its modeling capacity; as a result, RoBERTa-Base is less capable of capturing subtle contextual information compared to RoBERTa-Large [29]. RoBERTa possesses a larger vocabulary than BERT—50,265 versus 30,522 unique tokens [37]. Furthermore, RoBERTa was trained on a significantly larger corpus (160 GB compared to 16 GB for BERT), which potentially enhances its contextual understanding [29]. RoBERTa-Large has demonstrated superior contextual comprehension over various other methods, particularly in sentiment analysis [35], text classification, question answering, and voting-based classifier systems [16]. RoBERTa was developed as a BERT variant with improved pre-training strategies and a more flexible architecture [29].

Khosa, et al., [16] found that RoBERTa-Large achieved superior performance in news category classification compared to other Transformer-based models, including BERT, MPNet, and T5. The model attained the highest F1 scores using both Softmax Regression and voting classifiers, indicating its effectiveness in accurately capturing news context. **Timoneda and Vallejo Vera,** [37] evaluated BERT-Large, RoBERTa-Large, and DeBERTa-Large on political text classification tasks, including politeness analysis in political tweets. Their findings revealed that RoBERTa-Large consistently achieved the best performance across multiple classification tasks—particularly in identifying politeness in political discourse—outperforming both DeBERTa-Large and BERT-Large. While DeBERTa-Large demonstrated comparable performance in certain cases, RoBERTa-Large remained the most effective across various political text classification applications. With its greater modeling capacity and richer language representations, RoBERTa-Large has demonstrated strong advantages in numerous NLP tasks that require deep linguistic understanding.

To address a gap that remains underexplored in prior research—namely, the explicit integration of semantic modeling and temporal novelty, this study proposes a modular approach that combines the representational power of RoBERTa-Large with a time decay score based on Unix timestamps. The system computes two separate scores: a semantic relevance score obtained by encoding the news headline and abstract using RoBERTa-Large, and a temporal novelty score calculated through an exponential decay function. These two scores are combined at the inference stage using alpha (α) and lambda (λ) parameters to adaptively balance content relevance and recency, without requiring model retraining. This design not only supports scalability and efficiency for real-time deployment but also provides flexibility in controlling the system's focus on newly published news. Evaluation was conducted using the Microsoft News Dataset (MIND) benchmark, with metrics such as AUC, MRR, nDCG@5, and nDCG@10 used to assess the system's effectiveness in delivering relevant, accurate, and timely recommendations in response to dynamic information landscapes.

1.2 Theoretical Framework

This study is grounded in a set of integrated pillars and theoretical concepts that guide the development of a news recommendation system combining both semantic and temporal dimensions. Each component within this framework plays a crucial role in achieving the main objective of the study: enhancing the accuracy of timely news recommendations.

1. **Natural Language Processing (NLP) with RoBERTa-Large**

   Natural Language Processing (NLP) focuses on the interaction between computers and human language. In the context of news recommendation, NLP enables deeper understanding of textual content, thereby allowing for more precise measurement of news relevance. RoBERTa (Robustly Optimized BERT Approach) was developed as an advanced variant of BERT, incorporating several improvements, including a significantly larger pre-training corpus (approximately 160 GB), a more extensive vocabulary (50,265 tokens compared to BERT's 30,522), and dynamic masking strategies based on epochs. RoBERTa-Large, the high-capacity variant, typically consists of 24 Transformer layers with around 355 million parameters, enabling it to capture linguistic nuances and inter-word relationships more comprehensively. Owing to its extensive pre-training and self-attention mechanisms at every layer, RoBERTa-Large effectively reduces the cold-start problem and generates rich semantic representations, making it a foundational component for enhancing the accuracy of content-based news recommendation systems.

2. **Personalised Recommendation Systems**

   Once semantic representations are obtained from RoBERTa-Large, personalized recommendation systems integrate this information into a user-centered framework. These systems combine user preferences—both explicit (e.g., ratings) and implicit (e.g., click history)—with content features to estimate the relevance of news articles. Unlike traditional approaches that rely solely on collaborative filtering or content-based filtering, personalization in the news domain integrates multiple dimensions to build a more comprehensive user profile. Moreover, this integration takes into account temporal dynamics, where user preferences may shift rapidly in response to the latest developments. In other words, personalization is achieved by combining the semantic representations generated by RoBERTa-Large (for content relevance) with temporal features (for information novelty.

3. **Temporal Information Processing**

   Beyond aligning recommendations with individual preferences, recency is a critical factor due to the short lifecycle of news content. To maintain the relevance of recommendations, temporal information must be explicitly integrated. One common method is time decay, which assigns greater weight to more recently published articles. By considering the elapsed time between publication and recommendation, the system can adjust relevance scores to ensure that recent news is not overlooked. This approach ensures that recommendations remain up-to-date and minimizes the risk of presenting outdated content. The uniqueness of this approach compared to previous studies lies in the explicit and modular integration of temporal and semantic scores, where the temporal score emphasizes recency (novelty of information), while the semantic score from RoBERTa-Large focuses on content relevance. This modular design enables flexibility in tuning the system's sensitivity to novelty without requiring retraining of the model.

4. **Hybrid Semantic–Temporal Feature Integration**

   To fully leverage the available data—comprising semantic representations (e.g., headlines and abstracts) and temporal features (e.g., Unix timestamps)—a hybrid feature integration is performed. This step typically involves normalization, concatenation, and weighting of each dimension so that their contributions are proportional to their impact on relevance. By balancing textual understanding and temporal awareness, the recommendation system can prioritize fresh content while remaining accurate in assessing news relevance. In this study, the integration explicitly follows the previously described modular approach, which adaptively combines two scores—the semantic score from RoBERTa-Large and the temporal score derived via time decay—to generate recommendations that are both precise and responsive to recent news developments.

5. **Performance Evaluation**

   Once all components and integration mechanisms have been assembled, the system's performance must be evaluated using a variety of metrics to capture different dimensions of effectiveness. Ranking metrics such as Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (nDCG) assess the ordering quality of recommendations, while Area Under the Curve (AUC) measures the model's ability to distinguish between relevant and irrelevant items. Beyond accuracy, the evaluation also considers timeliness, ensuring that the system not only delivers content aligned with user interests but also provides the most current information available. The use of the Microsoft News Dataset (MIND)—which contains anonymized real-world user interaction data—ensures that the evaluation reflects realistic usage conditions actual usage.

1.3 Statement of the Problem

Current news recommendation systems often place a greater emphasis on content relevance while inadequately addressing the critical dimension of recency. Previous studies, particularly those leveraging pre-trained language models such as BERT-Base, typically process only a single textual feature (e.g., news headlines) and incorporate temporal information either separately or by relying solely on popularity signals. Therefore, these systems frequently recommend semantically relevant articles that are no longer timely, thereby failing to reflect the evolving nature of the news landscape. This limitation is particularly pressing in the news domain, which is inherently characterized by short content lifespans. Most prior works have utilized BERT-Base on the MIND dataset—a large-scale, real-world dataset provided by Microsoft. However, due to the scale and complexity of this dataset, BERT-Base has struggled to achieve optimal performance, with AUC scores typically falling below 73%. This limited modeling capacity negatively impacts the system's ability to deliver highly personalized, accurate, and time-sensitive news recommendations. To address these limitations, this study systematically formulates the following key research questions, progressing from practical considerations to analytical depth:

1. How does the use of different configurations of textual and temporal features—ranging from headlines only, to combinations of headlines and abstracts, to the explicit integration of temporal features such as publication date—affect the capability of news recommendation systems to balance content relevance with information novelty?

2. To what extent does increasing the modeling capacity using RoBERTa-Large improve recommendation performance—compared to baselines such as BERT-Base, BERT-Large, DeBERTa-Base, and RoBERTa-Base—on large-scale and complex datasets like MIND, particularly in the context of explicit integration of semantic and temporal features?

This study will be evaluated using AUC, MRR, and NDCG metrics to assess the system's effectiveness in recommending news that is both relevant and up to date. Through a comprehensive approach that explicitly integrates textual and temporal features, and by leveraging a high-capacity pre-trained language model (RoBERTa-Large), this research aims to address the identified gaps in prior studies and enhance user satisfaction by delivering recommendations that are more responsive to the dynamics of real-time information.

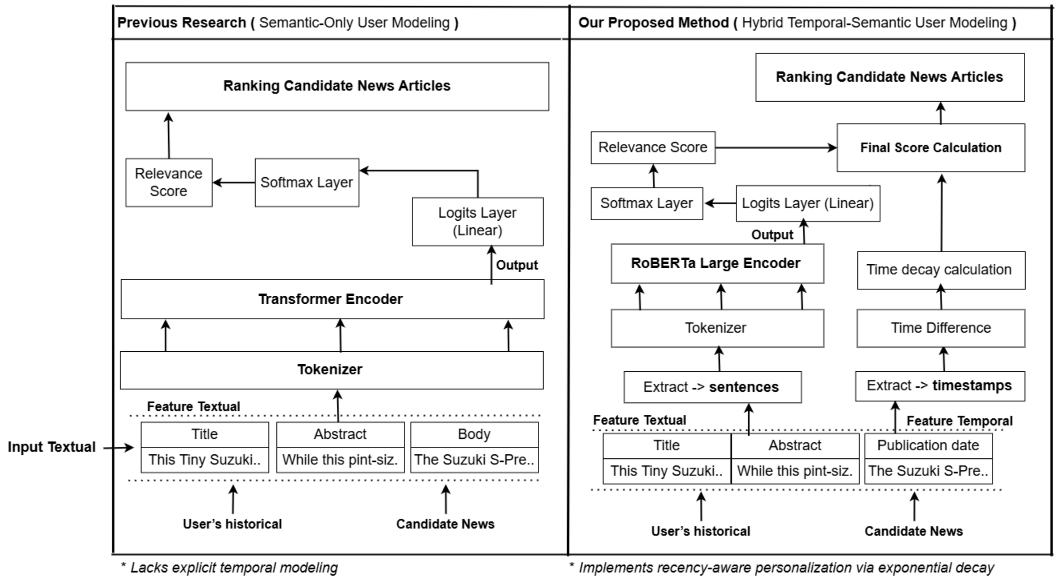

Figure 1.1: Comparison of previous and proposed user modeling approaches

1.4 Objective and Hypotheses

This study aims to develop a news recommendation system that integrates textual features (headlines and abstracts) with temporal features (publication dates) in order to enhance relevance without neglecting recency. This approach is expected to overcome the limitations of previous methods, which typically rely solely on news headlines and tend to ignore the fast-paced dynamics of the news domain. Additionally, the study leverages RoBERTa-Large—a high-capacity pre-trained language model—based on the assumption that it can effectively process large-scale data, provide comprehensive semantic understanding, and optimize the recommendation system to remain both relevant and up to date.

1.4.1 Objective

1. To develop a news recommendation system that explicitly integrates headlines, abstracts, and publication dates using a time-decay weighting mechanism, thereby effectively accounting for both content relevance and information recency.

2. To evaluate the capability of RoBERTa-Large in processing the integration of textual and temporal features and to compare its performance against baseline models (BERT-Base, BERT-Large, DeBERTa-Base, RoBERTa-Base) using AUC, MRR, and NDCG as evaluation metrics.

1.4.2 Hypotheses

Based on the theoretical and empirical premises identified, this study formulates the following hypotheses:

1. **Premise**: Previous studies have shown that combining features such as headlines and abstracts can enrich semantic representations in recommendation systems [15], [1], [45], [43]. Furthermore, the incorporation of temporal features through a time-decay mechanism can assign greater weight to more recent items, making the resulting recommendations more adaptive to current information dynamics [28], [33], [40].

   Hypothesis: Therefore, the integration of headlines, abstracts, and publication dates (with a time-decay weighting mechanism) is expected to produce news recommendations that are significantly more relevant and timely compared to approaches that rely solely on headlines.

   - **(H0)**: There is no significant difference in recommendation performance between a system that uses only news headlines without temporal feature integration and a system that integrates headlines, abstracts, and publication dates using a time-decay weighting mechanism, as measured by AUC, MRR, and NDCG on the MIND dataset.
   - **(H1)**: A system that integrates headlines, abstracts, and publication dates using a time-decay weighting mechanism produces significantly more relevant and timely recommendations compared to a system that uses only news headlines without temporal feature integration, as measured by AUC, MRR, and NDCG on the MIND dataset.

2. **Premise**: Prior studies have also demonstrated that RoBERTa-Large outperforms other models in various NLP tasks, particularly when applied to large-scale and complex datasets [16], [37]. RoBERTa-Large is trained on a 160 GB corpus with a vocabulary of approximately 50,000 tokens, enabling it to capture deep semantic understanding and contextual relationships.

   Hypothesis: Therefore, RoBERTa-Large is expected to effectively handle the complexity of large-scale datasets such as MIND, and to generate news recommendations that are more accurate, relevant, and sensitive to temporal aspects compared to baseline models (BERT-Base, BERT-Large, DeBERTa-Base, RoBERTa-Base).

   - **(H0)**: There is no significant difference in news recommendation performance between the RoBERTa-Large model and the baseline models (BERT-Base, BERT-Large, DeBERTa-Base, RoBERTa-Base), based on the evaluation metrics AUC, MRR, and NDCG on the MIND dataset, under conditions involving the use of combined semantic features (headlines and abstracts) and temporal features (publication dates with time decay).

- **(H1)**: The RoBERTa-Large model demonstrates significantly improved news recommendation performance compared to the baseline models (BERT-Base, BERT-Large, DeBERTa-Base, RoBERTa-Base), based on the evaluation metrics AUC, MRR, and NDCG on the MIND dataset, under conditions involving the use of combined semantic features (headlines and abstracts) and temporal features (publication dates with time decay).

The evaluation of both hypotheses will be conducted using evaluation metrics such as AUC, MRR, and NDCG, which comprehensively measure aspects of accuracy and ranking quality.

1.5 Assumption

This study is conducted based on several fundamental assumptions to ensure that the proposed news recommendation system can be implemented effectively and validly in achieving its stated objectives. The assumptions are as follows:

1. **Complete metadata for each article**: Each news item is assumed to possess complete metadata—including a news ID, title, abstract, and publication date—as well as records of user behavior in the form of user IDs and their corresponding article reading history.
2. **Sufficient user reading history**: Each user is assumed to have an adequately rich reading history, enabling the system to accurately infer user preferences and recommend articles that are topically and contextually relevant.
3. **Accurate publication dates**: All news articles are assumed to have valid, precise, and accurate publication dates, allowing the time-decay mechanism to assign higher weights to more recent articles. As a result, the system's recommendations are expected to be not only semantically relevant but also responsive to the novelty of information.

1.6 Scope and Delimitation

This study focuses on the development of a news recommendation system utilizing the MIND dataset, which comprises news article logs from the U.S. edition of MSN News. The independent variables in this study include: (i) content features—specifically, headlines and abstracts transformed into textual representations using RoBERTa-Large, and (ii) temporal features—publication dates converted into Unix timestamps and integrated via a time-decay function. The dependent variables consist of the relevance score (the softmax output of the model) and the final recommendation score, which is a combination of the relevance score and the time-decay weight. Technical parameters such as learning rate, batch size, weight decay, and input configurations (click history and candidate articles) are controlled to ensure consistency in evaluation. The data is restricted to news articles that possess complete metadata—including news ID, headline, abstract, and publication date—and to user behavior logs that record up to the 50 most recent clicks, allowing optimal modeling of user preferences. The scope of the study is further limited to English-language articles within the MIND dataset; publication date information is supplemented through web scraping techniques using Apify to ensure data completeness.

Data collection and analysis cover the period from November 9 to 15, 2019. The training and validation datasets span November 9–14, 2019, while the test data is drawn from November 15, 2019. However, November 20, 2019 is used as the reference "current day" when calculating publication time differences, enabling offline novelty evaluation independently of the actual calendar. The recommendation system is evaluated in an offline setting to assess its ability to provide relevant and timely recommendations. Real-time deployment is not within the scope of this study; thus, the findings are applicable solely within the context of controlled offline evaluation. The chosen scope and constraints ensure that the data used is of high quality and representative, enabling the system to recommend news articles that are aligned with user preferences while maintaining novelty through temporal features. Limiting the analysis to English-language articles also helps reduce linguistic variation, resulting in a more homogeneous and optimized evaluation setting. Additionally, the system is explicitly constrained to avoid recommending outdated news content. This is achieved by integrating a time-decay mechanism that assigns lower weights to older articles, thereby ensuring that only recent and relevant news are prioritized in the recommendation process.

1.7 Significance of the Study

This study has the potential to enrich the existing body of knowledge by proposing an innovative news recommendation framework that explicitly integrates content features—namely headlines and abstracts—with temporal information represented through publication dates. By leveraging the RoBERTa-Large text encoder and adopting a time decay mechanism based on Unix timestamps, the proposed model is expected to enhance semantic alignment between user preferences and news content, while also prioritizing the most up-to-date information. This approach aims to address the limitations of traditional recommendation systems that often focus solely on semantic relevance, thereby offering a more balanced and accurate solution to the problem of information overload. Moreover, the proposed modular design, namely the separate computation of semantic relevance scores and temporal novelty scores, combined adaptively using α and λ parameters at the inference stage, renders the system more flexible, efficient, and easily adjustable without the need for retraining the model. Preliminary findings from the evaluation using the MIND dataset and the metrics AUC, MRR, and NDCG indicate the potential for improved recommendation accuracy, providing explicit support for the conclusion that integrating temporal and semantic features is crucial for generating higher-quality news recommendations.

This research offers tangible benefits for several key stakeholders: (i) users, who require news that is both relevant and timely; (ii) recommendation system developers, who can apply this method to build more accurate and responsive engines; and (iii) news platforms, which aim to optimize user engagement by delivering content that aligns with both preference and novelty. Overall, this study promises to deliver valuable practical contributions and significant scientific insights, serving as a strong foundation for future advancements in the field of news recommendation systems.