(IJACSA) International Journal of Advanced Computer Science and Applications,
Vol. 17, No. 6, 2026

Emotion-Aware Serendipity Recommendation from Textual Reviews: BERT-Based Emotional Clustering

Mariam Benayad, Ahmed Zellou
SPM Research Team, ENSIAS Mohammed V University in Rabat, Rabat, Morocco

***Abstract*—This study proposes an emotion-aware serendipity recommendation framework based on textual reviews. Six BERT-based binary classifiers are trained to detect surprise, curiosity, trust, nostalgia, frustration, and enchantment from user reviews. The predicted emotion probabilities are used to build stable review-level emotional vectors. Two complementary contributions are introduced. Contribution 1 develops an emotion-aware reranking strategy that combines calibrated emotional information with relevance, novelty, and unexpectedness. Contribution 2 introduces EC-SereRank, an emotional-combination-based reranking model that preserves co-activated emotions within reviews and groups them using K-medoids into interpretable affective meta-groups. Experiments on Amazon Product Reviews show that affective information improves serendipity-oriented recommendation. The ablation study confirms that relevance and unexpectedness alone are not sufficient, while the integration of emotional compatibility improves recommendation quality. EC-SereRank further shows that emotional combinations provide an interpretable complementary signal for affective serendipity recommendation.**

***Keywords*—Recommender systems; emotion-aware recommendation; textual reviews; BERT; K-medoids; K-means**

I. INTRODUCTION

Recommender systems have become essential tools for helping users navigate large volumes of online information, products, and services [1], [2]. In e-commerce platforms, they are mainly designed to predict user preferences and generate personalized Top-K recommendation lists. Traditional recommender systems usually focus on accuracy-oriented objectives, such as relevance and ranking quality. However, highly accurate recommendations may still remain predictable, repetitive, or too similar to the user's previous interactions. This limitation can reduce the user's opportunity to discover novel and unexpected items [3], [4].

To address this issue, recent research has emphasized beyond-accuracy objectives, particularly novelty, unexpectedness, diversity, and serendipity [3], [9]. Serendipitous recommendation aims to suggest items that are not only relevant to the user, but also unexpected and potentially valuable. In this context, a good recommendation should not only match the user's known preferences, but also support discovery by introducing useful and surprising items [3], [11].

Recent studies on serendipity in recommender systems emphasize that effective recommendation should go beyond accuracy by considering novelty, unexpectedness, usefulness, and discovery value [3], [4]. Existing approaches have explored different ways to improve serendipity, including self-serendipity preference generation for cold-start scenarios [10], opinion mining from textual reviews for generating novel and pleasantly surprising recommendations [11], and multi-criteria Top-N recommendation methods [19]. In parallel, review-aware recommender systems have shown that textual reviews can improve user and item representation learning [12], [13], while sentiment-aware approaches demonstrate the usefulness of affective information extracted from user-generated content [14], [15]. However, most of these works rely either on rating patterns, sentiment polarity, or general opinion signals, and do not explicitly exploit fine-grained emotional dimensions for serendipity-oriented reranking. With the development of transformer-based language models such as BERT and DistilBERT [16], [17], and the availability of fine-grained emotion modeling resources such as GoEmotions [18], it becomes possible to build richer affective representations from textual reviews. This motivates the proposed emotion-aware framework, which integrates calibrated BERT-based emotional signals into serendipity-oriented recommendation.

Most existing serendipity-oriented approaches rely mainly on rating patterns, item popularity, user-item similarity, or structural information [10], [11]. Although these signals are useful, they do not fully exploit the rich information contained in user-generated textual reviews. Reviews provide more than explicit preference feedback; they also express subjective experiences, affective reactions, satisfaction, surprise, trust, frustration, curiosity, or emotional attachment. These affective cues can provide a deeper understanding of user preferences. Prior review-aware recommendation models have shown that textual reviews can improve user and item representation learning [12], [13], while serendipity-oriented studies suggest that richer preference signals may support discovery-oriented recommendation [3], [11].

Several studies have integrated textual reviews or sentiment analysis into recommender systems [12], [13], [14], [15]. However, sentiment-based approaches often reduce user opinions to coarse polarity categories, such as positive, negative, or neutral. Such representations are limited when the objective is to model more complex affective experiences. For example, two positive reviews may express different emotional meanings: one may reflect trust and reliability, while another may express surprise or enchantment. Similarly, a negative review may indicate frustration, disappointment, or unmet expectations. Therefore, a richer emotional representation is needed to better capture the affective structure of user reviews [18].

This study proposes an emotion-aware serendipity recommendation framework based on textual reviews. The framework extracts latent emotional signals from Amazon Product Reviews using BERT-based emotion classifiers [16], [17]. Since the original dataset does not provide explicit emotion labels, six weakly supervised binary classifiers are trained to detect surprise, curiosity, trust, nostalgia, frustration, and enchantment from user reviews.

The proposed framework introduces two complementary contributions. The first contribution, C1, is the main quantitative component of the study. It applies K-means clustering to calibrated BERT-based emotional vectors in order to identify dominant emotional structures in the review space. These calibrated emotional vectors and cluster-based profiles are then used to construct user and item affective profiles for emotion-aware reranking based on relevance, novelty, unexpectedness, and emotional compatibility.

The second contribution, C2, extends the affective representation by modeling emotional combinations. Instead of assigning each review only to a dominant emotional profile, C2 preserves the set of simultaneously activated emotions in each review. These emotional combinations are grouped using K-medoids into interpretable affective meta-groups. This combinatorial representation allows the system to capture richer multi-emotional configurations and to integrate emotional combination compatibility into the reranking process through the proposed *EC-SereRank* model. C2 is therefore presented as a complementary representational contribution that improves the interpretability of the affective space, rather than as a replacement for C1.

The main contributions of this work are summarized as follows:

- We propose an emotion-aware serendipity recommendation framework that integrates fine-grained affective signals extracted from textual reviews into Top-K recommendation.
- We train six BERT-based binary emotion classifiers to detect surprise, curiosity, trust, nostalgia, frustration, and enchantment from user-generated reviews, and calibrate their outputs using emotion-specific thresholds.
- We propose C1, an emotion-aware reranking contribution that exploits emotional vectors together with relevance, novelty, and unexpectedness to improve serendipity-oriented recommendation. K-means clustering is also used to identify interpretable dominant emotional profiles and to support user/item affective profiling.
- We propose C2, an emotional-combination-based contribution that captures co-occurring active emotions within reviews, clusters the observed emotional combinations using K-medoids into interpretable affective meta-groups, and integrates emotional combination compatibility into the proposed EC-SereRank reranking model.
- We evaluate the proposed framework using Top-K accuracy metrics and beyond-accuracy metrics, including novelty, unexpectedness, and serendipity. The experimental results and ablation analysis show that affective information provides a useful complementary signal for serendipity-oriented recommendation.

The main quantitative contribution of this work is the C1 emotion-aware reranking model, which evaluates the impact of calibrated emotional compatibility on serendipity-oriented Top-K recommendation. C2 is introduced as a complementary representational contribution that preserves multi-emotional co-activation patterns through emotional combinations and K-medoids meta-groups.

The remainder of this study is organized as follows: Section II reviews related work on recommender systems, serendipity, textual reviews, and affective recommendation. Section III presents the proposed methodology, including dataset preprocessing, emotion annotation, BERT-based emotion extraction, emotional calibration, C1, C2, and reranking. Section IV reports the experimental results and discusses the impact of the proposed affective representations. Finally, Section V concludes the study and outlines future research directions.

II. RELATED WORK

Recommender systems are traditionally optimized for accuracy-oriented objectives, such as relevance, precision, and ranking quality [1], [2]. However, accuracy alone may lead to repetitive and predictable recommendations, reducing the user's opportunity to discover novel and unexpected items. To overcome this limitation, recent research has increasingly focused on beyond-accuracy objectives, including novelty, diversity, unexpectedness, coverage, and serendipity [5], [6].

Serendipity is generally associated with recommendations that are relevant, unexpected, and useful to the user. Kotkov et al. [3] provide a comprehensive survey of serendipity in recommender systems and highlight the diversity of its definitions and evaluation strategies. Similarly, Ziarani and Ravanmehr [4] show that serendipity remains a complex concept because its formalization and measurement vary across studies. Earlier work by Murakami et al. [7] also emphasized the importance of evaluating serendipity through recommendation-list metrics, while Adamopoulos and Tuzhilin [8] focused specifically on the role of unexpectedness in improving user satisfaction.

Several approaches have attempted to improve serendipity through popularity debiasing, long-tail recommendation, weak-tie structures, unexpectedness modeling, or multi-objective optimization. Xu et al. [10] proposed GS2-RS, a serendipity-oriented recommendation framework that generates self-serendipity preferences from rating data in order to address cold-start and filter-bubble problems. Shrivastava et al. [11] proposed an optimized framework exploiting textual reviews and opinion mining to generate recommendations that are novel, unexpected, and relevant. Duricic et al. [9] further reviewed beyond-accuracy objectives, including diversity, serendipity, and fairness, in graph neural network-based recommender systems. These works confirm the importance of balancing relevance with discovery-oriented dimensions.

More recent studies have extended serendipity-oriented recommendation beyond classical rating-based settings. Al Jurdi et al. [20] explored the role of weak ties and network-inspired clustering in improving unexpected but relevant recommendations. Kaya and Kaleli [19] proposed a Top-N recommendation method for multi-criteria collaborative filtering and evaluated beyond-accuracy dimensions such as novelty, diversity, and serendipity. Chen et al. [21] introduced a cognitive-based knowledge learning framework that integrates cognitive psychology and knowledge graphs to improve both accuracy and beyond-accuracy objectives, including diversity and serendipity. Yu et al. [22] addressed long-tail service recommendation using cross-view and contrastive learning, emphasizing that long-tail items can improve novelty and serendipity while reducing popularity bias. From a content-oriented perspective, Yang et al. [23] showed that recommendation content attributes, including content serendipity, can influence consumers' recommendation adoption intention.

Despite these advances, most existing serendipity-oriented approaches still rely mainly on rating data, item popularity, user-item similarity, structural signals, cognitive signals, or general opinion information. They rarely integrate fine-grained emotional dimensions extracted from textual reviews into the serendipity reranking process. This limitation motivates the proposed emotion-aware framework, which uses calibrated BERT-based affective signals to enrich serendipity-oriented recommendation beyond classical relevance, novelty, and unexpectedness components.

A. Textual Reviews and Opinion-Aware Recommendation

Textual reviews provide rich information about user preferences and item characteristics. Unlike ratings, which summarize user feedback through a numerical value, reviews contain explanations, subjective opinions, product experiences, and contextual details. Review-aware recommender systems, therefore, exploit textual reviews to improve user and item representation.

Zheng et al. [12] proposed DeepCoNN, a deep cooperative neural network that jointly learns user behaviors and item properties from review text. Chen et al. [13] introduced NARRE, an attention-based review-aware model that learns the usefulness of reviews and provides review-level explanations for recommendations. These works demonstrate that textual reviews can improve recommendation performance by providing richer user-item representations than ratings alone.

III. METHODOLOGY

A. Overview of the Proposed Framework

The proposed framework aims to enhance serendipity-oriented recommendation by exploiting latent emotional signals extracted from textual reviews. Instead of relying only on explicit ratings or conventional relevance scores, the proposed approach introduces an affective layer based on BERT-inferred emotions. These emotions are then used to construct two complementary recommendation contributions.

The overall framework is composed of four main stages. First, textual reviews are processed using BERT-based emotion classifiers in order to infer six affective dimensions: surprise, curiosity, trust, nostalgia, frustration, and enchantment. Second, the predicted emotional probabilities are used to construct review-level emotional vectors. These emotional vectors prepare the data for clustering and profile construction.

Third, two affective modeling strategies are introduced. The first contribution, C1, develops an emotion-aware reranking strategy based on review-level emotional vectors. K-means clustering is also applied to these emotional vectors in order to identify interpretable dominant emotional profiles and to support the construction of user and item affective profiles. The second contribution, C2, extends the emotional representation by preserving multiple active emotions within each review. It builds combinatorial emotional signatures and groups them using K-medoids to obtain interpretable affective meta-groups.

Finally, the affective profiles obtained from C1 and C2 are integrated into a serendipity-oriented reranking stage. The reranking function combines relevance, unexpectedness, novelty, and affective compatibility in order to generate Top-K recommendations that are both useful and discovery-oriented.

The general architecture of the proposed framework is summarized in Fig. 1.

B. Dataset and Preprocessing

The experimental study is conducted on the Amazon Product Reviews dataset obtained from Kaggle. The raw dataset initially contains 568,454 reviews and 10 attributes: *Id*, *ProductId*, *UserId*, *ProfileName*, *HelpfulnessNumerator*, *HelpfulnessDenominator*, *Score*, *Time*, *Summary*, and *Text*. Each record represents a user-item review associated with a given product.

Although the raw dataset includes several metadata and rating-related fields, the proposed preprocessing stage mainly relies on *UserId*, *ProductId*, *Text*, and *Time*. The *UserId* and *ProductId* fields are used to preserve the user-item interaction structure. The *Text* field is used as the main textual source for affective information extraction. The *Time* field is preserved for the recommendation stage in order to maintain the chronological order of user interactions.

The first preprocessing step consists of preparing the textual content of each review. Missing values in the *Text* field are replaced with empty strings, and the field is converted into textual format. Reviews with empty, very short, or non-informative textual content are removed by keeping only reviews whose length exceeds 20 characters. This filtering step ensures that the retained reviews contain enough meaningful textual information for the subsequent emotion extraction and recommendation stages.

A light text normalization step is then applied. The text is converted to lowercase, leading and trailing spaces are removed, and repeated whitespace characters are replaced with a single space. This normalization is used to make the text representation more consistent while avoiding aggressive preprocessing. In particular, no stemming, lemmatization, or heavy stop-word removal is applied, since preserving the original wording remains important for capturing contextual and affective cues in user-generated reviews.

After text cleaning, the user-item interaction structure is filtered in order to support a reliable Top-K recommendation protocol. Since users with only one or two reviews do not provide enough interaction history for constructing user-level affective profiles or for chronological evaluation, only users with at least three interactions are retained. This filtering step produced a pool of eligible users from which the experimental subset was constructed.

The final experimental subset contains 10,000 reviews. It was not obtained by uniformly sampling isolated reviews

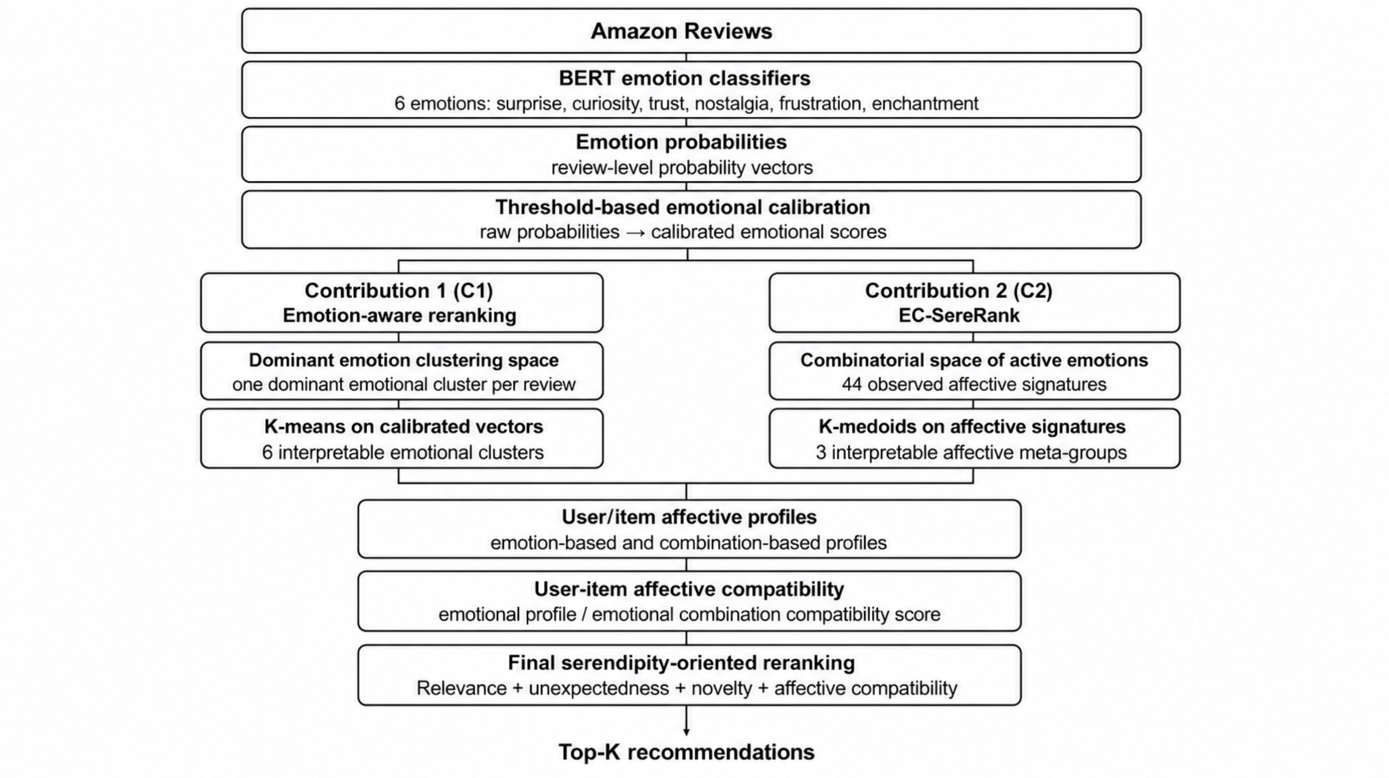

Fig. 1. General architecture of the proposed emotion-aware and combinatorial serendipity recommendation framework.

from the raw dataset. Instead, the subset was constructed from eligible users with at least three interactions, while preserving their user-item structure and temporal information. This strategy ensures that each retained user has sufficient historical interactions to support profile construction and recommendation evaluation. As shown in Table I, the resulting subset contains 1,582 users and 4,899 products. This controlled subset size also makes the BERT-based emotion extraction, clustering, user/item profile construction, and reranking stages computationally feasible while maintaining a meaningful user-item recommendation setting.

TABLE I. Dataset Filtering Summary

Step	Value
Raw reviews	568,454
Raw attributes	10
Minimum text length	$>$ 20 characters
Minimum interactions per retained user	≥ 3
Final experimental reviews	10,000
Final users	1,582
Final products	4,899

Overall, the preprocessing pipeline transforms the raw Amazon review dataset into a structured experimental subset containing user identifiers, product identifiers, cleaned review texts, ratings, and temporal information. These prepared data are then used in the following stages of the methodology, including weakly supervised emotion annotation, BERT-based emotion extraction, affective representation learning, and Top-K recommendation evaluation.

C. Emotion Annotation Strategy

Since the original Kaggle dataset does not provide explicit emotion labels, a weak supervision strategy is adopted to generate initial affective labels from review texts. Six target emotions are considered: surprise, curiosity, trust, nostalgia, frustration, and enchantment. These emotions are selected because they cover both discovery-oriented affective states, such as surprise and curiosity, and experience-oriented affective states, such as trust, nostalgia, frustration, and enchantment.

For each emotion, a manually curated list of lexical patterns is defined using regular expressions. These patterns are designed to capture representative linguistic cues associated with each emotional dimension. For example, surprise is associated with expressions related to unexpectedness, curiosity with exploration-oriented terms, trust with reliability-related expressions, nostalgia with memory-related cues, frustration with dissatisfaction-related terms, and enchantment with delight-oriented expressions.

Before annotation, each review is normalized by converting the text to lowercase and removing redundant whitespace. The annotation function then scans the normalized text using the emotion-specific patterns. A review receives a positive weak label for an emotion when at least one corresponding lexical pattern is detected. Therefore, the minimum-hit threshold is set to one for all six emotions.

The annotation is applied independently for each emotion, making the task multi-label rather than multi-class. As a result, the same review can be associated with several emotions when it contains cues related to multiple affective dimensions. This design is important because user reviews may naturally express mixed emotional states rather than a single isolated emotion.

After this weak annotation process, the dataset contains six binary emotion columns, one for each target emotion. These weak labels are then used to prepare emotion-specific training datasets for the BERT-based binary classifiers. One classifier is trained for each emotion using a one-vs-rest strategy. This allows the model to learn contextual affective patterns beyond the initial lexical cues used for weak labeling.

Since weak supervision may introduce noisy labels, an emotion-specific threshold calibration step is applied after BERT training. For each emotion classifier, the decision threshold is optimized on the validation set by selecting the threshold that maximizes the F1-score. The calibrated thresholds are then used to generate the final emotion vectors used in the clustering and recommendation stages.

Although this weak supervision strategy enables scalable emotion labeling without requiring a fully human-annotated corpus, it remains a limitation of the study. The generated labels should therefore be interpreted as weak affective labels rather than gold-standard human annotations. This limitation is mitigated by using contextual BERT classifiers, emotion-specific validation, and threshold calibration.

D. BERT-Based Emotion Extraction

After the weak annotation stage, emotion extraction is formulated as six independent binary classification tasks, corresponding to surprise, curiosity, trust, nostalgia, frustration, and enchantment. This formulation is consistent with the multi-label nature of user reviews, since the same review may express several emotions simultaneously. Therefore, instead of training a single multi-class model, a separate BERT-based binary classifier is trained for each emotional dimension.

For each emotion, the dataset is restricted to the review text and the corresponding binary label. The annotated data are split into training, validation, and test sets using stratified sampling in order to preserve the distribution of positive and negative classes. In the implemented pipeline, *distilbert-base-uncased* is used as the base language model. Review texts are tokenized using the BERT tokenizer, with padding and truncation applied to a maximum length of 64 tokens.

Because the emotion labels are imbalanced, a weighted training strategy is adopted. The standard cross-entropy loss is replaced with a class-weighted cross-entropy loss, giving higher importance to the minority class. This reduces the tendency of the classifier to favor the majority class, especially for less frequent emotions such as curiosity, nostalgia, and frustration.

During training, performance is monitored using accuracy and F1-score. Although a default threshold of 0.5 is commonly used for binary classification, a validation-based threshold optimization is applied after training in order to account for emotion-specific class imbalance and probability distributions. For each emotion, predicted probabilities on the validation set are evaluated over thresholds ranging from 0.10 to 0.90 with a step of 0.05. The threshold that maximizes the F1-score is selected as the optimal decision threshold; in case of equal F1-scores, the threshold with the higher accuracy is retained.

After training, the six BERT-based classifiers are applied to each review. Each classifier outputs the probability that its corresponding emotion is expressed in the review. The six predicted probabilities are then concatenated to form a six-dimensional review-level emotional probability vector. This vector represents the affective profile of the review and is used as input for the subsequent threshold-based emotional calibration stage.

E. BERT-Based Emotion Classification Results

Table II reports the final test performance of the six BERT-based emotion classifiers. The results show that the affective extraction layer provides reliable emotion signals for the downstream recommendation pipeline. Enchantment and surprise obtain the strongest F1-scores, while nostalgia, frustration, trust, and curiosity remain sufficiently robust for affective profile construction, emotional clustering, and emotion-aware reranking.

TABLE II. BERT-BASED EMOTION CLASSIFICATION RESULTS

Emotion	Threshold	Test F1
Enchantment	0.85	0.9383
Surprise	0.30	0.9207
Nostalgia	0.20	0.8418
Frustration	0.20	0.8176
Trust	0.45	0.8112
Curiosity	0.20	0.7955

F. Contribution 1: BERT-Based Emotional Clustering with K-means

The first contribution aims to structure the calibrated emotional space produced by the BERT-based emotion extraction layer. Each review is represented by a six-dimensional calibrated emotional vector composed of surprise, curiosity, trust, nostalgia, frustration, and enchantment scores. This representation provides a continuous affective description of each review.

Before applying clustering, the calibrated emotional vectors are standardized in order to ensure that all emotional dimensions contribute equally to the clustering process. K-means is then applied to the standardized calibrated vectors to identify dominant affective structures in the review space.

Given a set of calibrated emotional vectors, K-means partitions the reviews into K clusters by minimizing the within-cluster sum of squared distances:

$$\arg\min_{C} \sum_{k=1}^{K} \sum_{r \in C_k} \|E_{cal}(r) - \mu_k\|^2 \tag{1}$$

where, C_k denotes the k-th emotional cluster, μ_k is its centroid, and $E_{cal}(r)$ represents the calibrated emotional vector of review r.

In this contribution, each review is assigned to one dominant emotional cluster. The number of clusters is set to six in order to reflect the six emotional dimensions considered in the study. Each cluster is then interpreted according to the dominant dimension of its centroid. This produces interpretable emotional groups associated with high enchantment, high trust, high nostalgia, high frustration, high curiosity, and high surprise.

This clustering step transforms individual review-level emotion vectors into structured affective groups. Instead of using only isolated probability scores, C1 provides an interpretable emotional segmentation of the review space. These clusters are then used to support the construction of user and item emotional profiles in the recommendation pipeline.

G. C1 User and Item Emotional Profiles

After assigning each review to a dominant emotional cluster, user and item emotional profiles are constructed from the calibrated emotional vectors and the K-means cluster assignments.

For each user u, the emotional profile is obtained by aggregating the calibrated emotional vectors of the reviews associated with that user:

$$P_u^{C1} = \frac{1}{|R_u|} \sum_{r \in R_u} E_{cal}(r) \tag{2}$$

where, R_u denotes the set of reviews associated with user u.

Similarly, for each item i, the emotional profile is computed by aggregating the calibrated emotional vectors of the reviews associated with that item:

$$P_i^{C1} = \frac{1}{|R_i|} \sum_{r \in R_i} E_{cal}(r) \tag{3}$$

where, R_i denotes the set of reviews associated with item i.

In addition to these continuous emotional profiles, the distribution of K-means cluster assignments is also used to describe the dominant affective tendency of users and items. This allows the model to capture both the average emotional intensity and the dominant emotional group structure.

The affective compatibility between a user and an item is then estimated by comparing their emotional profiles:

$$AffectAlign_{C1}(u,i) = sim(P_u^{C1}, P_i^{C1}) \tag{4}$$

where, $sim(.)$ denotes a similarity function, such as cosine similarity or dot product. This affective alignment score is later integrated into the serendipity-oriented reranking stage.

H. Contribution 2: BERT-Based Combinatorial Emotional Modeling with K-Medoids

The second contribution extends the affective representation by preserving the multi-emotional nature of user reviews. While C1 assigns each review to one dominant emotional cluster, C2 represents each review through the set of emotions that are simultaneously activated. This set is referred to as an *emotional combination* and is denoted by $EC(r)$ for a given review r. This representation allows the model to capture richer affective configurations, such as trust-enchantment, surprise-curiosity, or surprise-curiosity-trust-frustration.

For each review r, an emotion e is considered active when its BERT-predicted probability exceeds its emotion-specific optimal threshold:

$$p_e(r) \geq \theta_e^* \tag{5}$$

where, $p_e(r)$ denotes the predicted probability of emotion e for review r, and θ_e^* is the optimal threshold selected during validation.

The emotional combination of a review is then formally defined as:

$$EC(r) = \{e \in E \mid p_e(r) \geq \theta_e^*\} \tag{6}$$

where, E denotes the set of the six emotions: surprise, curiosity, trust, nostalgia, frustration, and enchantment. Thus, $EC(r)$ represents the set of active emotions associated with review r.

Since six emotions are considered, the theoretical number of possible emotional combinations is:

$$2^6 = 64 \tag{7}$$

including the empty combination. In practice, only the emotional combinations observed in the dataset are retained. In the implemented pipeline, this process produces 44 observed affective signatures.

After constructing the emotional combination space, K-medoids clustering is applied to group similar emotional combinations into compact and interpretable affective meta-groups. K-medoids is selected because, unlike centroid-based clustering methods, it uses real observed emotional combinations as representative medoids. This property makes the resulting affective meta-groups easier to interpret.

Each emotional combination is represented in the six-dimensional calibrated emotional space. The dissimilarity between two emotional combinations x and y is computed using Euclidean distance:

$$d(x,y) = \sqrt{\sum_{j=1}^{6} (x_j - y_j)^2} \tag{8}$$

The K-medoids objective is to minimize the sum of distances between each emotional combination and the medoid of its assigned group:

$$\arg\min_{M} \sum_{k=1}^{K} \sum_{x \in G_k} d(x, m_k) \tag{9}$$

where, G_k denotes the k-th affective meta-group and m_k is its medoid. In this contribution, the observed emotional combinations are grouped into three interpretable affective meta-groups.

This contribution allows the framework to move beyond a single dominant emotion per review. Instead, it models emotional co-occurrences and captures affective configurations that may be particularly useful for serendipity-oriented recommendation.

I. C2 User and Item Emotional Combination Profiles

After constructing the emotional combinations and their K-medoids meta-groups, user and item profiles are built from the distribution of emotional combinations. Unlike C1, where profiles are based on averaged calibrated emotional vectors and dominant emotional clusters, C2 represents users and items through the frequency distribution of emotional combinations.

For each user u, the emotional combination profile is defined as the normalized distribution of the emotional combinations observed in the user's review history:

$$P_u^{C2} = [q_{u,1}, q_{u,2}, ..., q_{u,m}] \tag{10}$$

where, $q_{u,k}$ denotes the proportion of the k-th emotional combination in the reviews associated with user u, and m is the number of observed emotional combinations.

Similarly, for each item i, the emotional combination profile is computed as:

$$P_i^{C2} = [q_{i,1}, q_{i,2}, ..., q_{i,m}] \tag{11}$$

where, $q_{i,k}$ represents the proportion of the k-th emotional combination in the reviews associated with item i.

The compatibility between a user and an item is estimated through *emotional combination compatibility*, denoted as $ECC(u,i)$. This score is computed using cosine similarity between the user and item emotional combination profiles:

$$ECC(u,i) = \frac{P_u^{C2} \cdot P_i^{C2}}{\|P_u^{C2}\| \, \|P_i^{C2}\|} \tag{12}$$

where, P_u^{C2} and P_i^{C2} denote the user and item emotional combination profiles, respectively. The resulting score measures the degree of compatibility between the emotional combination preferences of the user and the emotional combination structure of the item.

In addition to fine-grained emotional combination profiles, the K-medoids meta-groups provide a more compact and interpretable representation of the combinatorial emotional space. Thus, C2 captures two levels of affective information: the detailed level of emotional combinations and the higher-level structure of affective meta-groups.

The resulting compatibility score $ECC(u,i)$ is later integrated into the EC-SereRank reranking model as a combinatorial affective compatibility signal

J. Serendipity-Oriented Reranking

After constructing the affective representations in C1 and C2, the final stage consists of reranking candidate items in order to promote serendipitous recommendations. The objective is to preserve relevance while introducing additional beyond-accuracy dimensions, namely unexpectedness, novelty, and affective compatibility.

For a given user u and candidate item i, the general reranking function is defined as:

$$Score(u,i) = w_r R(u,i) + w_u U(u,i) + w_n N(i) + w_a Aff(u,i) \tag{13}$$

where, $R(u,i)$ denotes relevance, $U(u,i)$ denotes unexpectedness, $N(i)$ denotes novelty, and $Aff(u,i)$ denotes the affective compatibility signal. The parameters w_r, w_u, w_n, and w_a control the contribution of each component.

In C1, the affective signal is derived from the compatibility between the continuous emotional profile of the user and the continuous emotional profile of the item. This score is denoted as emotional profile compatibility, $EPC(u,i)$:

$$Aff_{C1}(u,i) = EPC(u,i) \tag{14}$$

Thus, the C1 reranking model, denoted as *SereEmotion*, is defined as follows:

$$SereEmotion(u,i) == w_r R(u,i) + w_u U(u,i) + w_n N(i) + w_e EPC(u,i) \tag{15}$$

In C2, the affective signal is derived from the compatibility between the emotional combination profile of the user and that of the item. This score is denoted as emotional combination compatibility, $ECC(u,i)$:

$$Aff_{C2}(u,i) = ECC(u,i) \tag{16}$$

Accordingly, the C2 reranking model, denoted as *EC-SereRank*, is defined as follows:

$$EC{-}SereRank(u,i) = w_r R(u,i) + w_u U(u,i) + w_n N(i) + w_c ECC(u,i) \tag{17}$$

In the controlled version of *EC-SereRank*, the emotional combination compatibility signal is injected with a small weight in order to preserve the relevance structure of the recommendation list while adding a fine-grained combinatorial affective signal:

$$EC - SereRank(u,i) = 0.95 \cdot Rel(u,i) + 0.05 \cdot ECC(u,i) \tag{18}$$

This formulation allows the model to introduce emotional combination compatibility without excessively disrupting ranking relevance. The final recommendation list is obtained by sorting candidate items according to the reranking score and selecting the top-K items.

K. Evaluation Metrics

The proposed framework is evaluated using both accuracy-oriented metrics and beyond-accuracy metrics. Accuracy-oriented metrics assess the relevance and ranking quality of the recommended items, while beyond-accuracy metrics evaluate the ability of the system to promote novelty, unexpectedness, and serendipity.

For each user u, let Rec_u^K denote the list of top-K recommended items, and let GT_u denote the set of relevant ground-truth items. Precision@K measures the proportion of relevant items among the recommended items:

$$Precision@K = \frac{1}{|U|}\sum_{u\in U}\frac{|Rec_u^K \cap GT_u|}{K} \tag{19}$$

Recall@K measures the proportion of relevant items successfully retrieved in the top-K list:

$$Recall@K = \frac{1}{|U|}\sum_{u\in U}\frac{|Rec_u^K \cap GT_u|}{|GT_u|} \tag{20}$$

HitRate@K evaluates whether at least one relevant item appears in the recommendation list:

$$HitRate@K = \frac{1}{|U|}\sum_{u\in U}\mathbb{I}(|Rec_u^K \cap GT_u| > 0) \tag{21}$$

where, $\mathbb{I}(.)$ is an indicator function.

To evaluate ranking quality, Mean Average Precision (MAP) is used. For each user, Average Precision at K is computed as:

$$AP@K(u) = \frac{1}{|GT_u|}\sum_{j=1}^{K} Precision@j(u)\cdot rel_u(j) \tag{22}$$

where, $rel_u(j) = 1$ if the item ranked at position j is relevant for user u, and 0 otherwise. MAP@K is then defined as:

$$MAP@K = \frac{1}{|U|}\sum_{u\in U} AP@K(u) \tag{23}$$

Normalized Discounted Cumulative Gain (NDCG@K) is also used to evaluate whether relevant items are ranked near the top of the recommendation list:

$$NDCG@K = \frac{1}{|U|}\sum_{u\in U}\frac{DCG@K(u)}{IDCG@K(u)} \tag{24}$$

where,

$$DCG@K(u) = \sum_{j=1}^{K}\frac{rel_u(j)}{\log_2(j+1)} \tag{25}$$

and $IDCG@K(u)$ denotes the ideal discounted cumulative gain for user u.

In addition to accuracy-oriented metrics, novelty is evaluated in order to measure the ability of the system to recommend less popular items. The novelty score of an item i is defined as:

$$Nov(i) = 1 - Pop(i) \tag{26}$$

where, $Pop(i)$ denotes the normalized popularity of item i. The novelty of a recommendation list is computed as the average novelty of its top-K items:

$$Novelty@K(u) = \frac{1}{K}\sum_{i\in Rec_u^K} Nov(i) \tag{27}$$

Unexpectedness measures the degree to which a recommended item differs from the user's expected preference space. In this work, unexpectedness is computed from the distance between user and item representations:

$$Unexp(u,i) = 1 - Sim(P_u, P_i) \tag{28}$$

where, P_u and P_i denote the user and item profiles, respectively, and $Sim(.)$ is a similarity function.

Finally, serendipity is evaluated by combining relevance with unexpectedness. The standard serendipity score is defined as:

$$Serendipity_{std}(u,i) = Rel(u,i)\cdot Unexp(u,i) \tag{29}$$

To better reflect the objective of the proposed framework, an extended serendipity metric is also considered. It incorporates relevance, unexpectedness, novelty, and affective compatibility:

$$Serendipity_{common}(u,i) = Rel(u,i)\cdot Unexp(u,i)\cdot Nov(i)\cdot Aff(u,i) \tag{30}$$

where, $Aff(u,i)$ corresponds to the affective compatibility score. In C1, this term is represented by the emotional profile compatibility $EPC(u,i)$, while in C2 it is represented by the emotional combination compatibility $ECC(u,i)$.

For a recommendation list, the serendipity score is computed as the average score over the top-K recommended items:

$$Serendipity@K(u) = \frac{1}{K}\sum_{i\in Rec_u^K} Serendipity(u,i) \tag{31}$$

L. Experimental Protocol and Ablation Study

The experimental protocol is designed to evaluate the contribution of each component of the proposed framework to Top-K recommendation performance and serendipity. The final working subset contains 10,000 reviews and is used consistently across the emotion extraction, affective profiling, and recommendation stages. Only users with sufficient interaction history are retained in order to build meaningful user profiles and support reliable evaluation.

A temporal evaluation strategy is adopted. For each user, interactions are ordered chronologically using the *Time* field. The most recent interaction is held out as the test item, while the remaining interactions are used to construct the user's historical profile. This protocol simulates a realistic recommendation scenario in which the model recommends items based on past user behavior and is evaluated on a future interaction.

The candidate items are reranked according to different scoring functions. To analyze the impact of each component, an ablation study is conducted by progressively adding unexpectedness, novelty, and affective compatibility to the relevance-based baseline. The evaluated variants are defined as follows:

- Relevance-only: uses only the relevance score $Rel(u,i)$.
- Relevance + Unexpectedness: combines relevance with unexpectedness.
- Relevance + Unexpectedness + Emotion: integrates the emotional compatibility signal from C1.
- Relevance + Novelty + Unexpectedness + Emotion: combines all C1 components.
- EC-SereRank: integrates emotional combination compatibility from C2 as a controlled complementary signal to relevance.

The purpose of this ablation study is to determine whether the proposed affective components improve recommendation quality beyond a standard relevance-driven ranking. In particular, the comparison between the relevance-only baseline and the emotion-aware variants makes it possible to quantify the added value of emotional modeling. Similarly, the comparison between C1-based emotional compatibility and C2-based emotional combination compatibility allows us to evaluate whether modeling active emotional combinations provides an additional benefit for serendipity-oriented recommendation.

All models are evaluated at different cutoff values, namely $K = 10$, $K = 20$, and $K = 50$. The evaluation includes both accuracy-oriented metrics, such as Precision@K, Recall@K, HitRate@K, MAP@K, and NDCG@K, and beyond-accuracy metrics, such as Novelty@K, Unexpectedness@K, and Serendipity@K. This dual evaluation makes it possible to assess whether the proposed models improve discovery-oriented recommendation while preserving acceptable ranking quality. The results of this ablation protocol are reported and discussed in the following section.

IV. RESULTS AND DISCUSSION

This section presents the main experimental results of the proposed emotion-aware serendipity recommendation framework. The analysis focuses on three aspects: the performance of the BERT-based emotion classifiers, the interpretability of the affective structures obtained in C1 and C2, and the impact of the proposed affective signals on Top-K recommendation and serendipity.

A. Performance of BERT-Based Emotion Classification

The first step evaluates the quality of the BERT-based emotion extraction layer. Six independent binary classifiers were trained for surprise, curiosity, trust, nostalgia, frustration, and enchantment. Table III summarizes the final test F1-scores and the optimal thresholds selected on the validation set.

TABLE III. BERT-BASED EMOTION CLASSIFICATION RESULTS

Emotion	**Optimal Threshold**	**Test F1-Score**
Surprise	0.30	0.9207
Curiosity	0.20	0.7955
Trust	0.45	0.8112
Nostalgia	0.20	0.8418
Frustration	0.20	0.8176
Enchantment	0.85	0.9383

The results show that the emotional layer is globally reliable. Enchantment and surprise obtain the strongest performance, while nostalgia, frustration, and trust also provide robust results. Curiosity is more difficult to detect, but remains exploitable for the recommendation pipeline. These results confirm that the extracted emotional probabilities can be used as meaningful affective signals.

B. Confusion Matrices of the Six BERT-Based Emotion Classifiers on the Test Set

To further validate the reliability of the BERT-based emotion classifiers, the confusion matrix components were visualized using two complementary figures. Fig. 2 presents the correct classifications, namely true negatives (TN) and true positives (TP), for each emotion. The results show that the models correctly identify a large number of non-active emotion instances, while also detecting positive emotion cases across the six affective dimensions. In particular, enchantment obtains the highest true positive count, which confirms the strong detection capacity already observed through its F1-score.

Fig. 3 focuses on the classification errors by reporting false positives (FP) and false negatives (FN). This representation makes it easier to identify the most challenging emotions. Trust presents the highest number of false negatives and false positives, indicating that it remains the most difficult emotion to classify. Curiosity also appears more challenging than surprise and nostalgia, but its error level remains acceptable for downstream affective recommendation. Overall, these two figures confirm that the proposed emotion classifiers do not simply predict the majority class, but provide exploitable emotional signals for the calibration, clustering, and serendipity-oriented reranking stages.

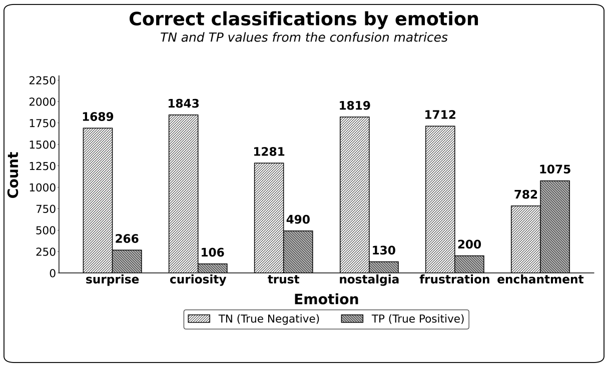

Fig. 2. Correct classification components of the six BERT-based emotion classifiers.

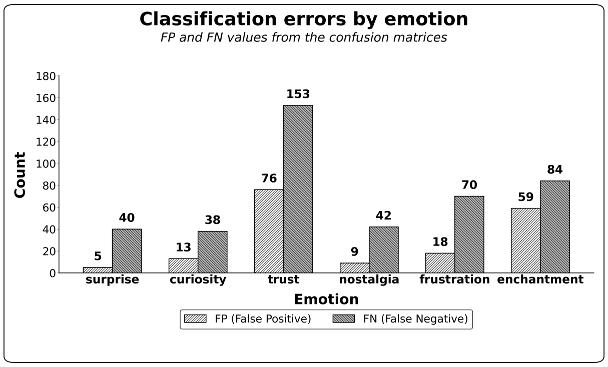

Fig. 3. Classification error components of the six BERT-based emotion classifiers.

C. Literature Positioning of C1 Against Existing Systems

Table IV positions C1 with respect to existing Top-10 recommendation systems reported in the literature. The values of the existing systems correspond to the average results over the six datasets reported in the comparative study. However, this table is provided only as a literature positioning reference. The compared systems were evaluated on different datasets and under different experimental protocols, whereas C1 is evaluated on the Amazon Fine Food Reviews subset used in this study. Therefore, the results should not be interpreted as a direct empirical comparison.

TABLE IV. LITERATURE POSITIONING OF C1 AGAINST EXISTING TOP-10 SYSTEMS.

System	**Prec.**	**Recall**	**Ser.**	**nDCG**
Matrix factorization based recommendations	0.340	0.322	0.021	–
Hybrid recommendations	0.410	0.383	0.027	0.335
Sankar et al.	0.415	0.408	0.036	0.373
FBPRR	0.422	0.445	0.030	0.382
C1: Emotion-aware reranking	0.061	**0.611**	0.034	**0.449**

The results show that C1 follows a different performance profile from the systems reported in the literature. In terms of Precision@10, C1 obtains a lower value than the compared systems. This difference should be interpreted carefully, since C1 is evaluated under a leave-one-out Top-K protocol in which each user has only one held-out relevant item. Under this setting, Precision@10 is bounded by 0.10 and corresponds to Recall@10 divided by 10. Despite this lower precision value, C1 achieves a high Recall@10 of 0.611, indicating that the proposed model is able to recover the held-out relevant item for approximately 61% of the evaluated users. C1 also obtains a competitive nDCG@10 value of 0.449, suggesting that the relevant item is often ranked in a favorable position within the Top-10 list. Regarding serendipity, C1 reaches a value of 0.034, which is competitive with the values reported by the compared systems. Nevertheless, because serendipity definitions and evaluation protocols differ across studies, this result should be interpreted as a positioning indicator rather than as a direct superiority claim. Overall, the table shows that C1 is not optimized for precision alone, but provides a strong recall-oriented and serendipity-oriented recommendation profile.

D. Ablation Analysis of C1: Added Value of the Emotional Component

To evaluate the contribution of the emotional component in C1, an ablation analysis was conducted by progressively integrating relevance, unexpectedness, emotion, and novelty into the reranking function. Table V reports the serendipity-oriented results at $K = 10$.

TABLE V. ABLATION RESULTS OF C1 AT K=10

C1 Variant	**Ser_std**	**Ser_full**
Relevance only	0.000488	0.012546
Relevance + Unexpectedness	0.000715	0.012714
Relevance + Unexpectedness + Emotion	0.002509	0.024797
Relevance + Novelty + Unexpectedness + Emotion	**0.005631**	**0.033690**

While Fig. 4 and Fig. 5 provide a graphical representation of the same ablation behavior for the standard and full serendipity metrics, respectively.

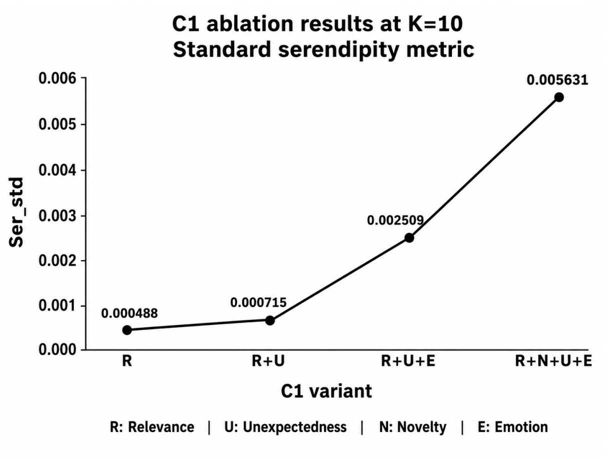

Fig. 4. Ablation results of C1 at $K = 10$ according to the standard serendipity metric.

The ablation results show that relevance alone is not sufficient to generate strong serendipitous recommendations. Adding unexpectedness slightly improves the serendipity score, but the gain remains limited. In contrast, the integration of the emotional component produces a clear improvement: the model combining relevance, unexpectedness, and emotion achieves a substantially higher serendipity score than the relevance–unexpectedness variant. This confirms that emotional alignment provides an effective complementary signal for serendipity-oriented reranking. The best performance is obtained when novelty is added together with relevance, unexpectedness, and emotion, which indicates that C1 benefits most from the joint integration of affective, novel, and unexpected recommendation signals.

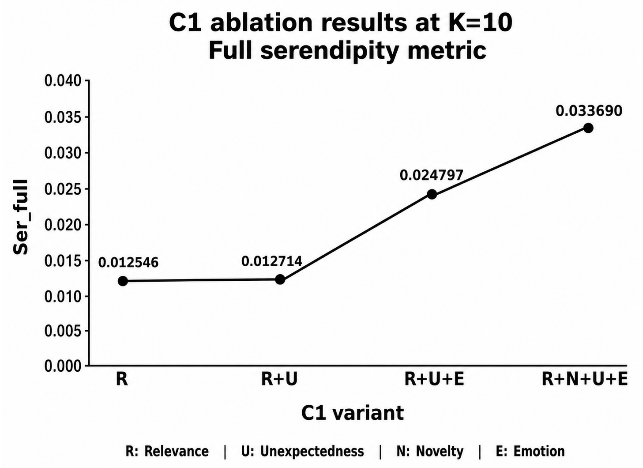

Fig. 5. Ablation results of C1 at $K = 10$ according to the full serendipity metric.

E. Analysis of C2: Emotional Combination Modeling with K-Medoids

C2 extends C1 by moving from dominant emotional clustering to fine-grained emotional combination modeling. While C1 assigns each review to one dominant emotional cluster, C2 preserves the multi-emotional nature of reviews by representing each review through its active emotional combination, denoted as $EC(r)$. With six calibrated emotions, the theoretical number of possible combinations is $2^6 = 64$. In practice, 44 emotional combinations were observed in the dataset.

K-medoids clustering was then applied to these observed emotional combinations in order to obtain compact and interpretable affective meta-groups. The final retained solution produced three meta-groups, summarized in Table VI.

TABLE VI. K-MEDOIDS AFFECTIVE META-GROUPS IN C2

Meta-Group	Reviews	Percentage
Positive affective core	5528	55.28%
Surprise discovery	2361	23.61%
Trust exploratory tension	2111	21.11%

These results show that C2 captures richer affective structures than C1. Instead of reducing each review to a single dominant emotional cluster, C2 keeps the interaction between multiple activated emotions. The first meta-group represents a positive affective pole mainly associated with enchantment, the second captures a discovery-oriented surprise profile, and the third reflects a trust-based exploratory tension. Therefore, C2 provides an additional level of affective interpretation by modeling emotional configurations rather than isolated dominant emotions.

F. Added Value of C2 Compared with C1

The added value of C2 compared with C1 lies mainly in the way affective information is represented and exploited in the recommendation process. C1 provides a clear and interpretable segmentation of the emotional space by assigning each review to a dominant affective cluster. This representation is useful because it identifies the main emotional tendency of each review. However, it may simplify reviews that contain several simultaneous emotional signals.

C2 addresses this limitation by preserving emotional co-activation patterns through emotional combinations. Instead of reducing a review to a single dominant emotional cluster, C2 represents each review by the set of active emotions detected by the calibrated BERT-based classifiers. This set is formalized as an emotional combination, denoted by $EC(r)$. Thus, a review can be represented not only by a single emotion such as *trust* or *surprise*, but also by richer affective signatures such as *trust-enchantment*, *surprise-curiosity*, or *surprise-curiosity-trust-frustration*.

TABLE VII. CONCEPTUAL ADDED VALUE OF C2 COMPARED WITH C1

Aspect	**C1 / Emotion-Aware Reranking**	**C2 / EC-SereRank**
Affective unit	Dominant emotional cluster	Emotional combination $EC(r)$
Representation	Single dominant affective profile	Multi-emotional signature
Emotional modeling	Main emotional tendency	Co-activated emotional patterns
Clustering method	K-means	K-medoids
Interpretation	Six emotional groups	Affective meta-groups
Recommendation signal	Emotion alignment	Emotional combination compatibility
Main value	Emotion-aware reranking and segmentation	Fine-grained affective compatibility

Based on this representation, EC-SereRank introduces an emotional combination compatibility component into the reranking process. This component measures the compatibility between the user and the item at the level of observed emotional configurations. Therefore, EC-SereRank captures finer affective relationships than a representation based only on individual emotions. This is particularly important for serendipity-oriented recommendation, where the objective is not only to recommend relevant items, but also to promote unexpected and emotionally meaningful discoveries.

Table VII summarizes the conceptual added value of C2 compared with C1. While C1 mainly provides an interpretable

TABLE VIII. EMPIRICAL GAINS OF EC-SERERANK OVER THE RELEVANCE-ONLY BASELINE IN C2

K	**Precision Gain**	**Recall Gain**	**MAP Gain**	**NDCG Gain**	**Serendipity Common Gain**
10	+8.60%	+8.60%	+13.97%	+12.14%	+12.03%
20	+11.78%	+11.78%	+14.40%	+13.37%	+7.06%
50	+14.43%	+14.42%	+14.61%	+14.27%	+6.01%

emotion-aware segmentation, C2 enriches the framework by introducing a fine-grained affective compatibility layer through EC-SereRank.

The empirical results further confirm the usefulness of the combinatorial signal inside the C2 framework. Compared with the relevance-only baseline, EC-SereRank improves the main ranking metrics and the proposed common serendipity score at different values of K. These gains show that emotional combinations provide a useful complementary signal when they are injected into the reranking function with a controlled weight.

As shown in Table VIII, EC-SereRank consistently improves the relevance-only baseline in Precision, Recall, MAP, NDCG, and Serendipity Common. At $K = 10$, the model improves Precision and Recall by 8.60%, MAP by 13.97%, NDCG by 12.14%, and the common serendipity metric by 12.03%. Similar improvements are observed at $K = 20$ and $K = 50$, which confirms that emotional combinations contribute positively to the recommendation process.

Nevertheless, C2 also has some limitations. First, its absolute Top-K performance remains lower than the best C1 model. This suggests that the combinatorial representation may lose part of the fine-grained continuous affective information captured by C1. Second, the variants that include novelty, unexpectedness, and meta-group information do not always improve the results compared with EC-SereRank. This indicates that additional C2 components may require more careful weighting. Third, the standard serendipity score slightly decreases compared with the relevance-only baseline, while the common serendipity score improves. This means that C2 is especially beneficial when serendipity is evaluated through a definition that includes affective compatibility.

Overall, C1 remains the strongest contribution from a purely quantitative recommendation perspective, while C2 provides a richer and more interpretable affective modeling strategy. The main value of EC-SereRank is not that it replaces C1, but that it introduces a more detailed emotional representation able to capture co-occurring affective signals and use them as complementary evidence for serendipity-oriented recommendation.

G. Discussion

Overall, the results support the main hypothesis of this work: affective information extracted from textual reviews can improve serendipity-oriented recommendation. C1 shows that calibrated BERT-based emotional vectors can be effectively exploited to build an emotion-aware reranking strategy. The results demonstrate that integrating emotional signals into the recommendation process improves both ranking quality and serendipity-oriented metrics compared with relevance and unexpectedness alone.

The ablation study confirms that relevance and unexpectedness are not sufficient to produce strong serendipitous recommendations. Adding unexpectedness alone leads to only a limited improvement, whereas the integration of affective information produces a more substantial gain. The best results are obtained when relevance, novelty, unexpectedness, and emotional compatibility are jointly considered. This confirms that emotion acts as a complementary signal that helps identify recommendations that are not only relevant, but also more meaningful from an affective and discovery-oriented perspective.

C2 extends this contribution by introducing a richer representation of affective information. While C1 mainly relies on dominant emotional tendencies, C2 preserves emotional co-activation patterns through emotional combinations. This allows EC-SereRank to model fine-grained affective compatibility between users and items. The empirical gains obtained by EC-SereRank over the relevance-only baseline show that emotional combinations provide useful additional information within the C2 framework, especially when evaluated through the proposed common serendipity metric.

However, the results also show that C2 should not be interpreted as a complete replacement for C1. C1 remains stronger in terms of absolute Top-K performance, while C2 provides a more expressive and interpretable affective modeling layer. Therefore, the two contributions are complementary: C1 provides the strongest quantitative emotion-aware reranking model, whereas C2 introduces EC-SereRank as an advanced combinatorial extension that improves the granularity and explainability of affective recommendation.

These findings suggest that serendipity-oriented recommendation can benefit from moving beyond classical relevance-based ranking. By incorporating calibrated emotional signals and emotional combinations, the proposed framework offers a more nuanced way to connect users and items through affective compatibility. Nevertheless, the evaluation remains offline, and future work should validate whether these affective and serendipitous recommendations are also perceived as useful, surprising, and satisfying by real users.

Therefore, the proposed framework contributes in two complementary ways. First, it introduces an emotion-aware reranking strategy based on calibrated BERT outputs. Second, it proposes a combinatorial emotional representation that captures co-occurring affective signals and exploits them through EC-SereRank. Together, these two contributions show that textual emotions can be transformed into actionable recommendation signals for improving serendipity.

V. CONCLUSION AND FUTURE WORK

This study proposed an emotion-aware serendipity recommendation framework based on textual reviews. The main objective was to move beyond relevance-driven recommendation by exploiting affective information extracted from user-generated reviews. To this end, weak emotional annotation and BERT-based binary classifiers were used to detect six emotional dimensions: surprise, curiosity, trust, nostalgia, frustration, and enchantment. A threshold-based calibration step was then applied to obtain more stable and comparable emotional representations.

Two complementary contributions were introduced. C1 developed an emotion-aware reranking strategy based on calibrated emotional vectors, while C2 extended this representation through emotional combinations and EC-SereRank. The first contribution demonstrated the effectiveness of direct emotional integration in the reranking process, whereas the second provided a richer and more interpretable affective representation by preserving emotional co-activation patterns.

The experimental results confirm that affective information can improve serendipity-oriented recommendation. The ablation study showed that relevance and unexpectedness alone are not sufficient to achieve strong serendipitous recommendations. The best results were obtained when relevance, novelty, unexpectedness, and emotional compatibility were jointly considered. In addition, EC-SereRank showed that fine-grained emotional combinations can provide a useful complementary signal within the C2 framework.

Despite these promising results, some limitations remain. The evaluation was conducted in an offline setting and on a single dataset, which limits the ability to fully assess real user perception and cross-domain generalization. Moreover, the performance of the framework depends on the quality of the emotion classifiers and on the calibration thresholds.

Future work will focus on validating the proposed framework on additional datasets and domains, integrating user-centered or online evaluation, and improving the weighting strategy used in the reranking process. Another promising direction is to combine emotional combinations with explainable recommendation mechanisms in order to provide users with clearer reasons for serendipitous recommendations.

REFERENCES

[1] F. Ricci, L. Rokach, and B. Shapira, Eds., *Recommender Systems Handbook*, 2nd ed. New York, NY, USA: Springer, 2015.

[2] 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.

[3] D. Kotkov, S. Wang, and J. Veijalainen, "A survey of serendipity in recommender systems," *Knowledge-Based Systems*, vol. 111, pp. 180–192, 2016.

[4] R. J. Ziarani and R. Ravanmehr, "Serendipity in recommender systems: A systematic literature review," *Journal of Computer Science and Technology*, vol. 36, no. 2, pp. 375–396, 2021, doi: 10.1007/s11390-020-0135-9.

[5] M. Ge, C. Delgado-Battenfeld, and D. Jannach, "Beyond accuracy: Evaluating recommender systems by coverage and serendipity," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, 2010, pp. 257–260.

[6] M. Kaminskas and D. Bridge, "Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems," *ACM Transactions on Interactive Intelligent Systems*, vol. 7, no. 1, pp. 1–42, 2016.

[7] T. Murakami, K. Mori, and R. Orihara, "Metrics for evaluating the serendipity of recommendation lists," in *New Frontiers in Artificial Intelligence*, Lecture Notes in Computer Science, vol. 4914, Springer, 2008, pp. 40–46.

[8] P. Adamopoulos and A. Tuzhilin, "On unexpectedness in recommender systems: Or how to better expect the unexpected," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 4, pp. 1–32, 2014.

[9] T. Duricic, D. Kowald, E. Lacic, and E. Lex, "Beyond-accuracy: A review on diversity, serendipity and fairness in recommender systems based on graph neural networks," *Frontiers in Big Data*, vol. 6, p. 1251072, 2023, doi: 10.3389/fdata.2023.1251072.

[10] Y. Xu, Y. Yang, and E. Wang, "Generating self-serendipity preference in recommender systems for addressing cold start problems," *arXiv preprint arXiv:2204.12651*, 2022.

[11] R. Shrivastava, D. S. Sisodia, N. K. Nagwani, and U. Roy, "An optimized recommendation framework exploiting textual review based opinion mining for generating pleasantly surprising, novel yet relevant recommendations," *Pattern Recognition Letters*, vol. 159, pp. 91–99, 2022.

[12] L. Zheng, V. Noroozi, and P. S. Yu, "Joint deep modeling of users and items using reviews for recommendation," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 2017, pp. 425–434.

[13] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural attentional rating regression with review-level explanations," in *Proceedings of the 2018 World Wide Web Conference Companion*, 2018, pp. 1583–1592.

[14] C. N. Dang, M. N. Moreno-Garcia, and F. De la Prieta, "An approach to integrating sentiment analysis into recommender systems," *Sensors*, vol. 21, no. 16, p. 5666, 2021.

[15] S. M. Al-Ghuribi and S. A. M. Noah, "A comprehensive overview of recommender system and sentiment analysis," *arXiv preprint arXiv:2109.08794*, 2021.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

[17] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[18] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 4040–4054.

[19] T. Kaya and C. Kaleli, "A novel top-n recommendation method for multi-criteria collaborative filtering," *Expert Systems with Applications*, vol. 198, p. 116695, 2022, doi: 10.1016/j.eswa.2022.116695.

[20] W. Al Jurdi, J. Bou Abdo, J. Demerjian, and A. Makhoul, "Exploring the power of weak ties on serendipity in recommender systems," in *Complex Networks 2023*, Studies in Computational Intelligence, vol. 1144, Springer, 2024, pp. 205–216, doi: 10.1007/978-3-031-53503-1_17.

[21] X. Chen, Q. Liang, Y. Chen, P. Wang, H. Yu, and X. Luo, "Cognitive-based knowledge learning framework for recommendation," *Knowledge-Based Systems*, vol. 287, p. 111446, 2024, doi: 10.1016/j.knosys.2024.111446.

[22] D. Yu, T. Yu, D. Wang, and S. Wang, "Long tail service recommendation based on cross-view and contrastive learning," *Expert Systems with Applications*, vol. 238, p. 121957, 2024, doi: 10.1016/j.eswa.2023.121957.

[23] T. Yang, F. Yang, and J. Men, "Recommendation content matters! Exploring the impact of the recommendation content on consumer decisions from the means-end chain perspective," *International Journal of Information Management*, vol. 68, p. 102589, 2023, doi: 10.1016/j.ijinfomgt.2022.102589.