Sentiment-Analysis and Content-Filtering: A Hybrid Approach for Recommending STRs

[redacted]
[redacted], Archana Baskaran, [redacted]

I. INTRODUCTION

The short-term rental (STR) market has grown significantly over the years, with many rental platforms offering a wide selection of properties. The process of selecting the ideal STR has become increasingly challenging due to the vast number of options available. A well-designed recommendation system can streamline this process by providing personalized suggestions that align with the needs and preferences of the traveler. Our recommendation system differs from traditional approaches by integrating both structured and unstructured data, leveraging sentiment analysis from reviews, and incorporating safety scores into the filtering process. By incorporating these innovative elements, our recommendation system will improve booking confidence, enhance user satisfaction, and deliver a personalized STR search experience for travelers.

II. PROBLEM DEFINITION

Current STR search methods often require extensive review reading to find unique preferences, as traditional systems rely on basic filters such as price, location, and amenities. Travelers seeking specific features such as community safety level and sentiment-based reviews struggle to find relevant listings that meet their needs. These limitations highlight the need for more personalized and efficient search methods that better address the unique preferences and requirements of travelers.

III. LITERATURE SURVEY

This section provides an overview of research on machine learning algorithms and sentiment analysis for STRs.

A. Machine Learning Algorithms

The transition to deep learning has enhanced STR recommendation systems, particularly for platforms like Airbnb. Initially, Gradient-Boosted Decision Trees (GBDT) improved booking conversions but plateaued, leading to the use of Deep Neural Networks (DNNs) and multi-task learning for ranking optimization [1]. Recurrent Neural Networks (RNNs) and listwise context encoding expanded search diversity [2], while embedding-based models improved real-time personalization using dynamic user and listings data [3]. LambdaRank neural networks and factorization machine-enhanced (FM) DNNs strengthened ranking accuracy by combining structured and unstructured data [1]; moreover, the recent Journey Ranker, a multi-task deep learning framework, has been used to balance guest and host preferences [4]. Unfortunately, deep learning models still face data sparsity issues, creating "cold-start problems" [3], while search diversity optimization prioritizes listing variety over relevance to a user's search [2].

Recommendation systems for STRs have transitioned from simple filters to complex multi-criteria and learning-based approaches. One approach integrated semantic clustering and sentiment analysis to align user preferences with attractions, improving the tourism recommendation system's accuracy [5]. Another method, example-critiquing, allowed users to iteratively refine rental searches based on their preferences [6]. Additionally, Graph Neural Networks (GNNs) used continuous training to enhance Airbnb recommendations by adapting to user behavior for improved flexibility and customization [7]. Alternatively, the Analytic Hierarchy Process (AHP) used user-defined weighted criteria for ideal selections [8]. Although these approaches improve personalization, they struggle to balance dynamic user preferences with diversity in the presented listings. To address this challenge, this project will implement features derived from sentiment analysis in combination with unique filter criteria in creating a Neural Network-based recommendation model.

B. Sentiment Analysis

Sentiment analysis improves recommendation systems by extracting user emotions to refine search suggestions and enhance personalization. While techniques like Structural Topic Modeling (STM) identified key Airbnb service attributes, they lacked integration into ranking models and failed to assess user sentiment [9]. Deep learning models such as RNNs, Convolutional Neural Networks (CNNs), and Long Short-Term Memory Networks (LSTMs) improved sentiment classification but faced issues with negation and sarcasm [10]. Meanwhile, context-aware sentiment analysis addressed data sparsity in collaborative filtering [11] and a game-theoretic approach optimized sentiment-driven recommendations [12].While these methods addressed key challenges, they faced limitations with real-time updates and adaptation. Similarly, semantic rule-based models captured implicit sentiment but struggled with adaptability [13]. In response, hybrid lexicon-based and deep learning models were introduced to enhance feature extraction but required extensive labeled data [14]. To address these drawbacks, we use context-aware sentiment analysis – a toolkit for Airbnb that proves user-generated reviews provide valuable insights into service quality [15]. This enhances recommendation accuracy by incorporating factors beyond noise levels and local atmosphere.

IV. PROPOSED METHOD

This project focuses on enhancing STR recommendation systems using sentiment analysis to create unique features. Common filters are bedrooms, price, kid-friendliness and noise level. We will add unique features: a safety score, a sentiment score, and a priority system that categorizes recommendations based on user preference.

Three datasets for reviews, listings, and safety scores will be combined for several U.S. cities. Each city's data is sufficiently large – i.e., the NYC reviews dataset contains 950,000 rows, sourced from Inside Airbnb. The crime dataset is from Centers for Open Science. All datasets are sourced from publicly available information.

Fig. 1. Figure 1 shows the project workflow, described in detail through the remainder of the report.

A. Data Pre-processing

The primary dataset used for this project focuses on 33 major U.S. cities selected for their high STR activity and diverse safety profiles. The listings and review data from *Inside Airbnb* includes detailed data from more than 20,000 user reviews per city. The initial processing involved implementing text cleaning and conversions on the data. Specifically, the dollar signs were removed from listing prices, and the percent signs were removed from host response rate and host acceptance rate. The response time feature, which indicates host response time to a traveler with inquiries, was converted from text to a numerical value. For instance, if the response time for a particular listing was "within an hour," the new response time will be 1. Extra slashes, quotations, brackets, and digits were removed from features like amenities, host verification, and host since. Additionally, significant binary features were derived using Term Frequency-Inverse Document Frequency (TF-IDF) on the amenities, description, and reviews features. One-hot encoding was done on categorical features.

To assess safety, city-level crime data was sourced from the *FBI Crime Data Explorer*, which consists of both violent and property crime categories. Each type of crime was assigned a severity-based weight informed by the average incarceration rate. A weighted crime score was calculated based on the frequency and severity of the offenses in each area. Min-max was applied to rescale the values to a range of 0-100, translating the values into interpretable safety scores where 100 represents the safest city.

B. Sentiment Analysis on Reviews

Context-aware sentiment analysis was applied to the reviews data, which consists of listing id,

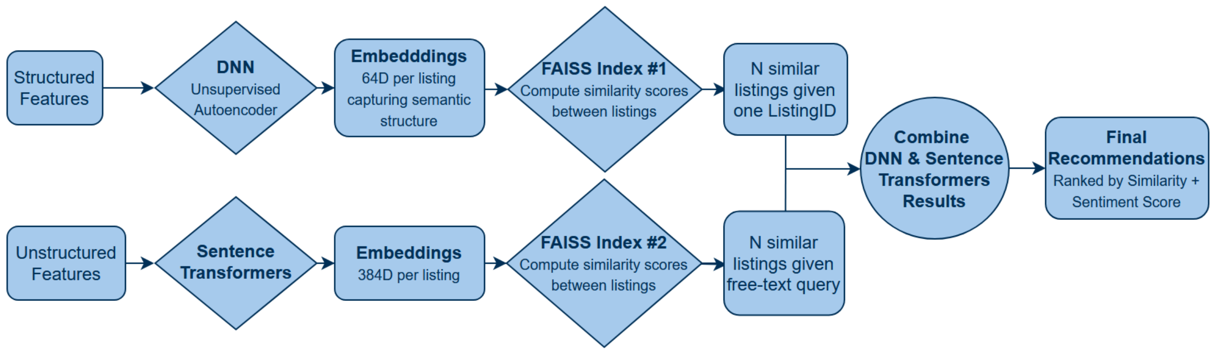

Fig. 2. As shown in the model diagram (Fig.2.), Structured and unstructured features are separately converted to embeddings. Then, cosine similarity is computed, N similar listings are returned, and the listings are combined and ranked, yielding the final recommendations.

user id, and comments. BERT and VADER sentiment scores are averaged to enhance sentiment predictions where the comments have mixed feelings, leading to different BERT and VADER positive/negative values.

C. Recommendation System

To ensure a user query matches what is reflected in a listing's features, two approaches were combined (Fig.2). The first approach takes structured features, such as price, bedrooms, and host acceptance rate, as input to a DNN. The second approach leverages Sentence Transformers with unstructured features, including description and amenities. Facebook AI Similarity Search (FAISS) was used separately on the embeddings from both approaches to compute cosine similarity between listings. N similar listings from both approaches are returned and combined to give the final recommendations.

Specifically, in the first approach an unsupervised autoencoder was used to reduce dimensionality and create lower-dimensional embeddings. The DNN activation function for embeddings prediction was regression and the loss function was MSE. For faster convergence, ADAM was used as the optimizer. Next, a FAISS index was built from the 64D per listing embeddings and similarity scores were calculated. From FAISS Index #1, users can query by listing ID to find N similar listings.

The second approach follows a similar structure, utilizing a Sentence Transformer pipeline to convert each listing's text into 384D embeddings. FAISS Index #2 is built on these embeddings and allows free-text queries to return the top N matching listings.

In this hybrid strategy, the user enters a query and listings are searched with the Sentence Transformer FAISS index. The N matching listing IDs are then searched in the DNN FAISS index. The final output is the union of these listings ordered by similarity and sentiment score.

D. User Interface

The user interface of this project incorporated a search feature where the user can freely input their desired STR. Using Sentence Transformers, we then provided suggested STRs based on recommendation score to allow users to visualize the relationship between their search criteria and available STR options.

Users will see the top three ranked listings based on their search query. They will then see additional results based on the similarity and sentiment scores.

V. Evaluation

The evaluation of the proposed recommendation system was conducted using a combination of quantitative performance metrics and qualitative user feedback. The assessment focused on three critical areas: the accuracy of sentiment analysis, the performance of the recommendation system, and the usability of the user interface. The testbed was setup using real-world data from publicly available sources to evaluate the performance of the recommendation system, as described in detail in section IV under Data Pre-processing.

The experiments were designed to answer the following questions:

- How effectively does sentiment analysis improve the relevance and personalization of STR recommendations?
- Are the recommendations displayed relevant to the search query?
- Does the addition of personalized safety scores and preference-based filters in the hybrid recommendation model give users more confidence in listing recommendations when choosing where to stay?
- Is the interactive interface intuitive and engaging for users exploring their preferences influenced by the personalized STR results?

A. *Sentiment Analysis*

Evaluating the performance of our context-aware sentiment analysis involves assessing how effectively the combined use of BERT and VADER distinguishes between positive and negative user reviews, particularly in cases where reviews contain mixed or unclear sentiment. Since the review data is unlabeled, performance of sentiment scores will be measured by viewing the correlation between the predicted sentiment scores and review scores grouped by the listing. Table 1 shows the results, indicating a positive correlation between review score ratings and sentiment scores; the sentiment analysis is effective.

	Correlation with Rating
Vader Compound	0.61
Bert Sentiment Score	0.67
Bert-Vader Sentiment Score	0.70

TABLE I

CORRELATION OF SENTIMENT SCORES AND REVIEW SCORES.

Additionally, the correlation between VADER and BERT sentiment scores was 0.70. The correlation measure close to 1 indicates that the two scores are moving in the same direction, which aligns with the expected outcome. This helps verify that the sentiment scores contribute meaningfully to the quality and relevance of the recommended listings.

B. *Recommendation System*

To evaluate the accuracy of the recommendation system, we first assessed how effectively the DNN generates listing embeddings. The DNN is trained on a combination of structured data from the listings dataset, sentiment scores extracted from user reviews, and neighborhood safety scores. As shown in Fig. 3, the steadily decreasing training and validation loss curves indicate that the DNN effectively learned patterns in the data while maintaining stable performance throughout training, suggesting low overfitting and reliable model performance. The inclusion of review-based sentiment and safety scores allows the system to account for additional listing qualities, resulting in more accurate and meaningful recommendations.

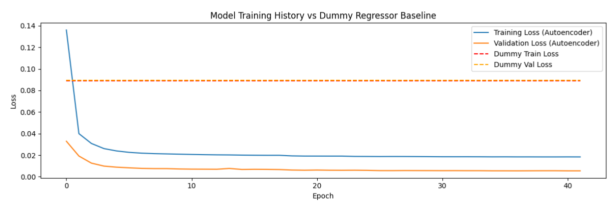

Fig. 3. The graph illustrates that the autoencoder achieves steadily decreasing loss and outperforms the dummy regressor baseline across over 42 epochs.

User testing was also used to evaluate the recommendation model. Because both unstructured and structured features were input to the recommendation system, listings shown should reflect the search query if it is performing well. A small-scale user study (*n=20*) involving classmates and friends was conducted to gather feedback. To assess the quality of recommendations, users classified the displayed listings as "relevant" or "irrelevant" to their search query.

17 out of 20 respondents found the displayed listings to be relevant (*85% recommendation accuracy*). In terms of amenities and qualities, such as "trails" and "quiet", the listings displayed were relevant. However, 3 respondents noted that including a specific location in the search did not guarantee listings to be in the location. Future work should consider adding a filter button so users can manually choose there location, and comparing Sentence Transformers with a keyword-based method.

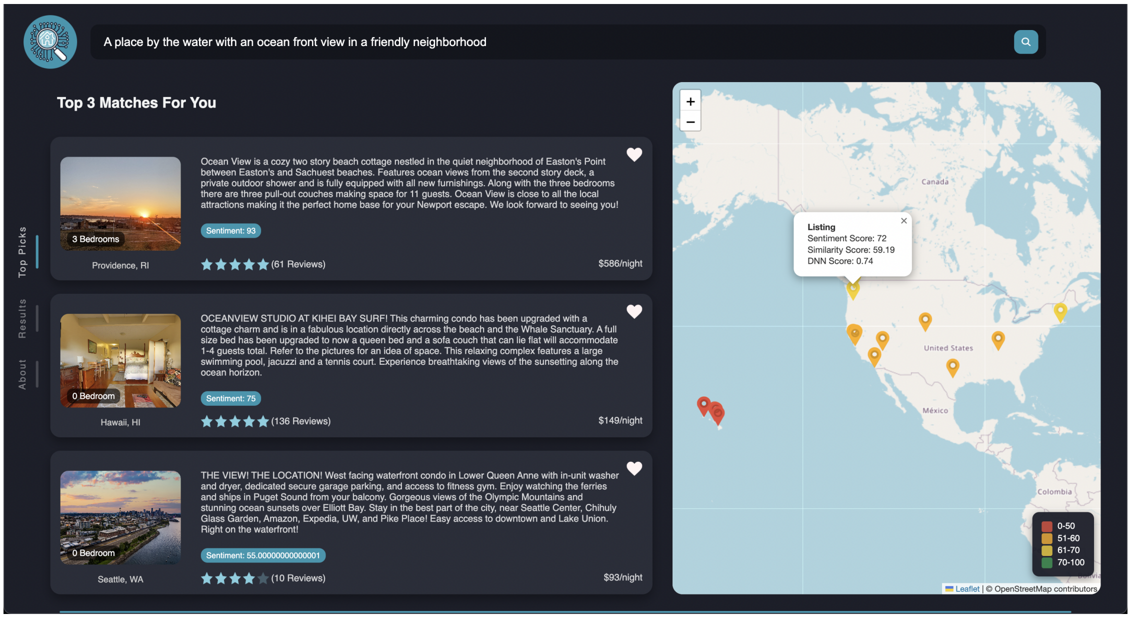

Fig. 4. The UI features an example of the top 3 listing recommendations tailored to the user's preferences. An interactive map on the right displays the geographic locations of the recommended listings, with the pin colors indicating the similarity to the user's search criteria.

C. *User Interface*

The web app design is intended to promote visual clarity by helping users explore their search criteria and visualize how different keywords influence the recommended STR options, as illustrated in the UI shown in Fig. 4. This will help determine whether the proposed structure increases user engagement while enabling faster, clearer, and more confident rental decisions.

Usability evaluation focused on determining how well the interface supports user-friendly interaction and provides a seamless experience. The same group evaluating the recommendation system provided feedback on the interface through a survey. Users rated their experience with the search bar, including sentiment score and similarity score, and wrote additional comments.

Users found the interface to be informative yet simple. Specifically, users appreciated the pins colored by similarity score (*4.5/5 average rating*). As well, the search bar was easy to use (*5/5 average rating*). Additionally, the listing information displayed provided a concise, informative summary – survey respondents were pleased with the sleek appearance (*4.75/5 average rating*). Overall, users were more confident in choosing where to stay (*4.75/5 average rating*). Based on the feedback received from user-testing, the UI successfully engages users and promotes visual clarity while exploring listings.

VI. CONCLUSIONS AND DISCUSSION

This project achieved the goal of improving STR search systems. The hypothesis was that incorporating unique features like sentiment, safety, and similarity scores would improve recommendations and user experience. The experiment results showed that users saw relevant listings, liked the ability to free-text search, and found the UI to be especially informative and intuitive. The outcome is significant because the recommendation system and interface empower users, giving them more personalization and location information in their search while maintaining an intuitive interface.

While the project was a success based on the experiment results, there are some limitations and potential future directions. One implication is that bias could be introduced through qualitative features, such as safety scores. However, it ultimately benefits users by eliminating the need for additional location research. More research should be done on the extent incorporating safety scores affects the booking rate of listings. Another potential improvement is replacing Sentence Transformers, which uses sentence embeddings, with a keyword-based method. As mentioned in section V under Evaluation, querying a location does not guarantee the displayed listings will be in that area. Finally, a larger user study should be conducted to gather feedback and ratings representative of the STR-user population.

Overall, the hybrid approach using Sentence Transformers and a DNN showed promising results. By integrating sentiment analysis from user reviews and neighborhood safety information, the hybrid recommendation model goes beyond traditional filters like price and amenities, delivering suggestions that are personalized and context-aware. **All team members have contributed a similar amount of effort.**

REFERENCES

[1] M. Haldar, M. Abdool, P. Ramanathan, T. Xu, S. Yang, H. Duan, Q. Zhang, N. Barrow-Williams, B. C. Turnbull, B. M. Collins, and T. Legrand, "Applying deep learning to airbnb search," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1927–1935. [Online]. Available: https://doi.org/10.1145/3292500.3330658

[2] M. Abdool, M. Haldar, P. Ramanathan, T. Sax, L. Zhang, A. Mansawala, S. Yang, and T. Legrand, "Managing diversity in airbnb search," *CoRR*, vol. abs/2004.02621, 2020. [Online]. Available: https://arxiv.org/abs/2004.02621

[3] M. Grbovic and H. Cheng, "Real-time personalization using embeddings for search ranking at airbnb," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 311–320. [Online]. Available: https://doi.org/10.1145/3219819.3219885

[4] C. H. Tan, A. Chan, M. Haldar, J. Tang, X. Liu, M. Abdool, H. Gao, L. He, and S. Katariya, "Optimizing airbnb search journey with multi-task learning," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 4872–4881. [Online]. Available: https://doi.org/10.1145/3580305.3599881

[5] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Expert Systems with Applications*, vol. 167, p. 114324, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417420310174

[6] G. N. Bernard Shibwabo, "A system for recommending rental properties," *Journal of Systems Integration*, vol. 8, 2017. [Online]. Available: https://www.researchgate.net/publication/322652906_A_System_for_Recommending_Rental_Properties

[7] Y. Chan, K.-W. Ng, S.-C. Haw, and N. Palanichamy, "Continuous training of recommendation system for airbnb listings using graph learning," *JOIV : International Journal on Informatics Visualization*, vol. 8, 03 2024.

[8] H.-Y. Jeong, "Multi criteria based personalized recommendation service using analytical hierarchy process for airbnb," *The Journal of Supercomputing*, vol. 77, p. 13224–13242, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11227-021-03812-6

[9] K. Ding, W. C. Choo, K. Y. Ng, and S. I. Ng, "Employing structural topic modelling to explore perceived service quality attributes in airbnb accommodation," *International Journal of Hospitality Management*, vol. 91, p. 102676, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278431920302280

[10] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics*, vol. 9, no. 3, p. 483, Mar. 2020. [Online]. Available: http://dx.doi.org/10.3390/electronics9030483

[11] N. A. Osman, S. A. Mohd Noah, M. Darwich, and M. Mohd, "Integrating contextual sentiment analysis in collaborative recommender systems," *PLOS ONE*, vol. 16, no. 3, pp. 1–21, 03 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0248695

[12] V. W. Zhang, G. Wang, N. I. Arshad, Q. Nguyen, M. Xia, and N. Innab, "Advancing consumer insights: efficient sentiment analysis-based recommendation system using multi-criteria decision making with a game theoretic approach," *Annals of Operations Research*, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:275310544

[13] O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A hybrid approach to sentiment analysis," in *2016 IEEE Congress on Evolutionary Computation (CEC)*, 2016, pp. 4950–4957.

[14] F. D. l. P. Cach N. Dang, María N. Moreno-García, "An approach to integrating sentiment analysis into recommender systems," *Sensors*, vol. 21, 2021.

[15] M. von Hoffen, M. Hagge, J. H. Betzing, and F. Chasin, "Leveraging social media to gain insights into service delivery: a study on airbnb," *Information Systems and e-Business Management*, vol. 16, pp. 247–269, 2018. [Online]. Available: https://doi.org/10.1007/s10257-017-0358-7